MICROBIAL GENOMICS

RESEARCH ARTICLE

Bogema *et al.*, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000255

Whole-genome analysis of extraintestinal *Escherichia coli* sequence type 73 from a single hospital over a 2 year period identified different circulating clonal groups

D. R. Bogema,[1,2]† J. McKinnon,[2]† M. Liu,[2] N. Hitchick,[3] N. Miller,[3] C. Venturini,[4] J. Iredell,[4] A. E. Darling,[2] P. Roy Chowdury[2] and S. P. Djordjevic[2,*]

Abstract

Sequence type (ST)73 has emerged as one of the most frequently isolated extraintestinal pathogenic *Escherichia coli*. To examine the localized diversity of ST73 clonal groups, including their mobile genetic element profile, we sequenced the genomes of 16 multiple-drug resistant ST73 isolates from patients with urinary tract infection from a single hospital in Sydney, Australia, between 2009 and 2011. Genome sequences were used to generate a SNP-based phylogenetic tree to determine the relationship of these isolates in a global context with ST73 sequences (*n*=210) from public databases. There was no evidence of a dominant outbreak strain of ST73 in patients from this hospital, rather we identified at least eight separate groups, several of which reoccurred, over a 2 year period. The inferred phylogeny of all ST73 strains (*n*=226) including the ST73 clone D i2 reference genome shows high bootstrap support and clusters into four major groups that correlate with serotype. The Sydney ST73 strains carry a wide variety of virulence-associated genes, but the presence of *iss*, *pic* and several iron-acquisition operons was notable.

DATA SUMMARY

1. All sequencing reads and assemblies for isolates sequenced in this study have been submitted to the ENA Sequence Read Archive (SRA) and GenBank, respectively. GenBank, SRA accession numbers and URLs are included in Table S2, available in the online version of this article. which has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477461 (URL – https://doi.org/10.6084/m9.figshare.5477461).

2. Scripts used for the analysis of SNP phylogeny have been deposited in Github; (URL – https://github.com/bogemad/snp_phylogeny).

3. Fig. S1 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477449 (URL – https://doi.org/10.6084/m9.figshare.5477449).

4. Fig. S2 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477485 (URL – https://doi.org/10.6084/m9.figshare.5477485).

5. Table S1 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477464 (URL – https://doi.org/10.6084/m9.figshare.5477464).

6. Table S3 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477473 (URL – https://doi.org/10.6084/m9.figshare.5477473).

7. Table S4 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477476 (URL – https://doi.org/10.6084/m9.figshare.5477476).

8. Table S5 has been deposited in Figshare; DOI: 10.6084/m9.figshare.5477479 (URL – https://doi.org/10.6084/m9.figshare.5477479).

Received 14 October 2018; Accepted 22 January 2019; Published 27 February 2019

Author affiliations: [1]Elizabeth Macarthur Agricultural Institute, NSW Department of Primary Industries, Menangle, NSW 2568, Australia; [2]The ithree Institute, University of Technology Sydney, NSW 2007, Australia; [3]San Pathology, Sydney Adventist Hospital, Wahroonga, NSW 2076, Australia; [4]Centre for Infectious Diseases and Microbiology, Westmead Institute for Medical Research, The University of Sydney, Westmead, NSW 2145, Australia.

***Correspondence:** S. P. Djordjevic, steven.djordjevic@uts.edu.au

Keywords: uropathogenic *Escherichia coli*; class 1 integron.

Abbreviations: CRL, complex resistance locus; ExPEC, extraintestinal pathogenic *Escherichia coli*; MDR, multiple-drug resistant; S1-PFGE, S1 nuclease-PFGE; SRA, Sequence Read Archive; ST, sequence type; UPEC, uropathogenic *Escherichia coli*; VAG, virulence-associated gene.

†These authors contributed equally to this work.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. There are two supplementary figures and five supplementary tables.

000255 © 2020 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Extraintestinal pathogenic *Escherichia coli* (ExPEC) are phylogenetically diverse and comprise uropathogenic *E. coli* (UPEC), neonatal meningitis-causing *E. coli* (NMEC) and avian pathogenic *E. coli* (APEC). ExPEC account for~75–95 % of urinary tract infections. A proportion of these infections can spread from the urinary tract with invasion of epithelial cells in the bladder (cystitis) and kidney cells (pyelonephritis) and transmission to systemic circulation (blood sepsis), posing a serious threat to human health. ExPEC are enteric bacteria, but their capacity to capture a wide array of virulence-associated genes (VAGs) by lateral gene transfer has expanded the repertoire of niches they colonize. ExPEC may carry diverse and often redundant combinations of VAGs whose impact on human health remains ill-defined. Epidemiological studies indicate that a subset of pathogenic *E. coli* lineages, including sequence type (ST)73, ST131, ST405, ST393, ST69, ST95, ST10, ST38 and ST127 [1–5], are responsible for most ExPEC infections [3, 6–8]. Carriage of combinations of virulence genes enhances virulence [9]; however, carriage of antimicrobial-resistance genes, particularly those encoding extended-spectrum β-lactamases and fluoroquinolones, as well as an ability to cause opportunistic infections in vulnerable (elderly) hosts, may also contribute to virulence. It is notable that none of these hypotheses have been experimentally validated [3, 6].

ExPEC have become the leading cause of blood sepsis in Europe [10]. Notable in this regard is the alarming rise in the incidence of ST73, now one of the most frequently isolated UPEC globally and the leading cause of bacteraemia in the East Midlands region of the UK [1, 11–14]. ST73 belongs to Clermont phylogroup B2 and is known to display different serogroups, with serotype O6 predominating (ST73-O6-B2) [15, 16]. It has recently been suggested that the rise in the incidence of multiple-drug resistant (MDR) ST73 in the UK is not due to the emergence of a dominant clone, because they are genetically diverse and carry a different array of plasmids encoding resistance to multiple antimicrobials. Many recently described isolates of ST73 carry genes that encode extended-spectrum β-lactamases and resistance to antimicrobials used in veterinary medicine [14]. This seems to be a recent adaptation in this ST, as previously characterized ST73 isolates from cases of uncomplicated urinary tract infection sourced from Greece, Portugal, Sweden and the UK were susceptible to most clinically relevant antimicrobials and most (75 %) did not carry plasmids, classic vehicles of multiple-drug resistance [17]. These data combined with the most recent findings seem to suggest that the rise in the carriage of MDR plasmids in ST73 may be a recent concerning event [14, 18].

Here, we have characterized whole-genome sequences of 15 class 1 integrase (*intI1*)-containing ST73 strains from a hospital in Sydney. To determine whether these highly localized strains were from a limited number of clonal lineages, phylogenetic inferences were made by comparing SNP differences in core genomes shared by ST73 strains from our Sydney collection with those from six high-quality reference genome sequences and ST73 strains (n=204) from seven countries sourced from global sequence read archives. We also examined mobile and chromosomal genetic content within this localized isolate cohort to further examine their accessory genome diversity. We compiled the repertoire of antimicrobial genes and virulence genes, and mapped the class 1 integron structures carried by these isolates. As carriage of the class 1 integrase is considered a reliable proxy for multiple-drug resistance [19], we used S1 nuclease-PFGE (S1-PFGE), followed by Southern hybridization with an *intI1* probe, to examine plasmid content and carriage of the class 1 integrase on plasmids.

IMPACT STATEMENT

Sequence type (ST)73 is a major clonal lineage of extraintestinal pathogenic *Escherichia coli* (ExPEC) that causes urinary tract infections, often with uroseptic sequelae, but has not garnered substantial scientific interest as has the globally disseminated ST131. Isolation of multiple-antimicrobial-resistant variants of ExPEC ST73 has increased in frequency, but little is known about the carriage of class 1 integrons in this ST and the plasmids that are likely to mobilize them. This pilot study examines the ST73 isolates within a single hospital in Sydney, Australia, and provides, to the best of our knowledge, the first large-scale core-genome phylogenetic analysis of ST73 utilizing public sequence read datasets. We used this analysis to identify at least eight sub-groups of ST73 within this single hospital. Mobile genetic elements associated with antibiotic resistance were less diverse and only three class 1 integron structures were identified, all sharing the same basic structure, suggesting that the acquisition of drug resistance is a recent event. Genomic epidemiological studies are needed to further characterize established and emerging clonal populations of multiple-drug resistant ExPEC to identify sources and aid outbreak investigations.

METHODS

Isolate source and culture conditions

Clinical samples in this project were from a larger collection obtained from the Sydney Adventist Hospital from 2009 to 2011. Bacterial species were identified by the VITEK 2 (bioMérieux) system at the Sydney Adventist Hospital. For DNA extraction, strains were first grown on a lysogeny broth (LB) agar plate to isolate single colonies, of which one was used to inoculate 2 ml LB, followed by shaking for 16 h at 37 °C. Antibiotic-susceptibility testing for ampicillin, cefotaxime, chloramphenicol, streptomycin and sulfafurazole was performed via the Calibrated Dichotomous Sensitivity (CDS) method [20]. These antibiotics were

selected based on antibiotic-resistance gene content inferred from genome sequencing.

Nucleic acid purification and whole-genome sequencing

E. coli DNA was extracted using the Isolate II genomic DNA extraction kit (Bioline), according to the manufacturer's instructions. For each sample, 'tagmentation' of genomic DNA and PCR amplification of tagged DNA were performed in triplicate using the Nextera system (Illumina). Sequencing libraries were pooled, then cleaned and size selected using SPRI beads (Beckman Coulter). Normalization was guided by read counts obtained from a Nano flowcell run on a MiSeq instrument. An Agilent 2100 Bioanalyzer, with a High Sensitivity DNA kit, was used to quantitate the pooled library before loading onto an Illumina HiSeq. Paired-end 150 bp reads were generated using the HiSeq 2500 v4 system.

Genome assembly and gene presence

Genome assembly was achieved with raw reads using the A5-MiSeq pipeline [21] and checked for consistency by additional assembly with SPAdes 3.9.0 [22]. Antibiotic-resistance genes and VAGs were identified from assembled genomes using BLASTN and SRST2 [23]. Searches were performed against antibiotic-resistance genes sourced from the ARG-ANNOT v3 database and a panel of VAGs identified from the Virulence Factors Database (VFDB) and literature searches [24, 25]. Serotyping was performed *in silico* with SRST2 using EcOH sequences supplied with this package. Draft genome reads obtained from the Sequence Read Archive (SRA) were searched using SRST2 for a minimal set of marker genes derived from integron structures and commonly associated transposons characterized in the Sydney strains. Low-quality alignments based on SRST2 output were discounted (n=3).

Archived sequence read selection

All additional ST73 sequences not generated by this study were obtained from complete whole-genome assemblies (n=6) [26–30] and public sequence read archives (NCBI/EMBL/DDBJ). Raw Illumina reads sourced from ST73 isolates (n=284) were considered for SNP-based phylogenetic analysis, including strains sequenced in this study from the Sydney Adventist Hospital (n=16), isolates with host, source and isolation location meta-data identified from the EnteroBase database (n=246; http://enterobase.warwick.ac.uk/; accessed 5/12/2016) and a previous ST73-focused study from the UK (n=22) [11]. Samples were excluded if the ST could not be confirmed as ST73 using SRST2 (n=4). Further samples were excluded (n=30) if isolate status could not be confirmed by BioProject meta-data or where the description of methods could not be identified by an associated publication [31–38]. Samples were additionally excluded (n=30) if they produced low reference genome coverage (>90 %) in whole-genome alignments. Additional sample filtering of ST73 reads is described in Table S1.

S1-PFGE analysis

The complement of large (>20 kb) plasmids in each bacterial isolate was determined by S1 nuclease (Promega) digestion and PFGE, as described previously [39, 40]. Southern blot hybridization was used to determine the genomic location of the *intI1* gene. PCR amplicons for *intI1* were obtained using published primers (int1F and int1R [41]) and labelled using a PCR DIG probe synthesis kit (Roche). DNA was transferred from the S1-PFGE gel to a nylon membrane (GE Health) using a VacuAid vacuum transfer apparatus (HybAid) and hybridization was performed using a DIG filter hybridization system (Roche), following the manufacturer's instructions. Images were acquired on a ChemiDoc MP System (Bio-Rad Laboratories).

SNP-based phylogenetic analyses

Our initial attempts to examine ST73 phylogeny with our 16 genome sequences and 6 complete whole-genome sequences (CFT073, ABU83972, ATCC25922, Nissle 1917, clone D i2 and clone D i14) using marker gene approaches [42] provided limited resolving power (Fig. S1). Consequently, to more appropriately examine these ST73 strains, we employed SNP-based phylogenetic methods.

For SNP-based phylogenetic trees, core genome alignments were generated with Snippy v3.1.0 (https://github.com/tseemann/snippy) using default options. Briefly, reads were mapped using BWA MEM v0.7 to an ST73 reference genome. Raw alignments were processed with samtools v1.3.1 and variants called using freebayes v1.1.0. SNP-derived genomes were reconstructed using vcftools v0.1.14, with low-coverage (<10×) and degenerate reference positions filtered. Recombinant regions were removed using Gubbins v2.20 (option -i 10) [43], yielding aligned, SNP-derived, recombination-filtered core genomes. From this alignment, core genome phylogenetic trees were inferred by maximum-likelihood using RAxML v8.2.9 [44]. Branch support was estimated by bootstrap analysis employing 100 replicate trees. Trees were rooted using the ultrametric tree method included with RAxML. Scripts used for the analysis of SNP phylogeny can be found online (at https://github.com/bogemad/snp_phylogeny).

Initially, we performed this analysis using only isolates sequenced in this study and a high-quality published ST73 reference (Fig. 1). To identify the most suitable reference genome for this purpose, we aligned reads individually to the six complete reference genomes (above) and examined reference sequence quality, core alignment lengths and final tree support values. Using this methodology, the most suitable reference was identified as clone D i2, which with archived public sequence reads generated a core genome of 3 818 344 bp, representing 75.8 % of the ST73 clone D i2 sequence.

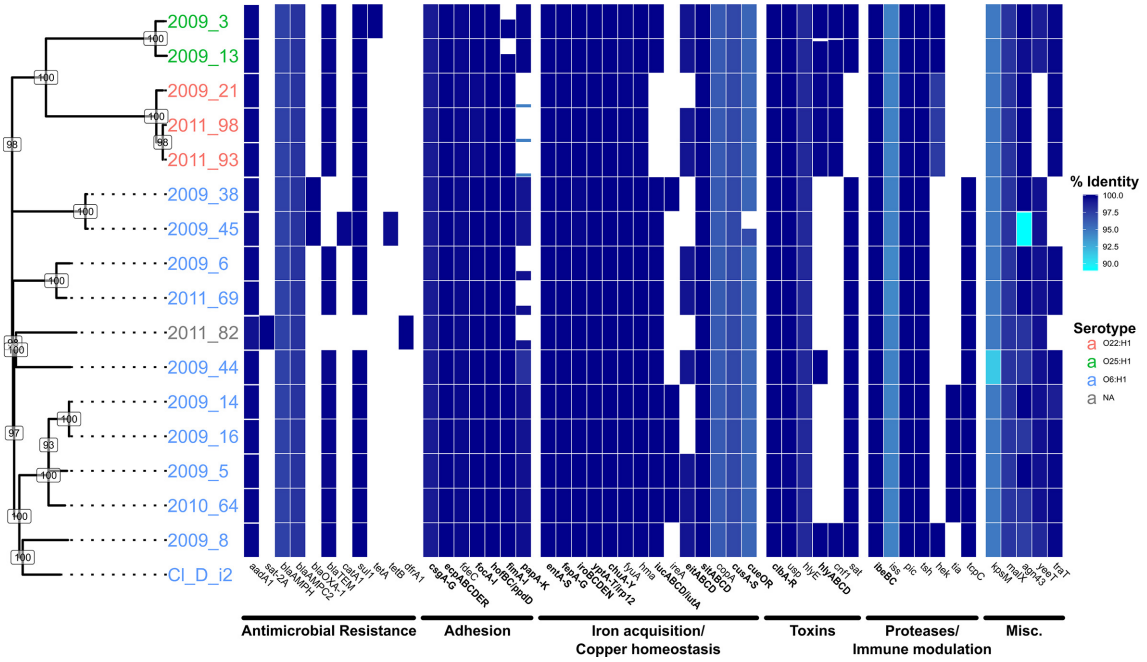

Fig. 1. A SNP-derived phylogenetic tree of the Sydney ST73 strains sequenced in this study compared with antimicrobial-resistance and virulence profiles. Bootstrap values based on 100 replicate trees are shown at labelled nodes. Isolate serotypes as determined by *in silico* serotyping are shown by coloured tip labels. Antibiotic resistance and virulence gene/gene-family presence (blue) or absence (white) is shown by the linked bar graph/heatmap. Percentage identity of BLAST matches is indicated by heatmap shade, with darker shades representing higher identity. BLAST match coverage is represented by tile height with solid tiles representing 100 % coverage. For gene families (*x*-axis; bold), tile height represents total BLAST match coverage of all gene family members and shows completeness of the gene family. Where single genes are indicated (*x*-axis; plain text), bar height represents BLAST match coverage of the gene.

RESULTS

Assembly information and statistics

The genome sequences of 16 ST73 strains from the Sydney Adventist Hospital were determined here. The whole-genome shotgun project has been deposited at GenBank/ENA/DDBJ and the SRA. Assembly statistics, as well as accession numbers, number of sequencing reads and the amount of sequencing data used to generate assemblies can be found in Table S2.

Public read high-throughput sequencing analysis

The Sydney isolates separated into several groups that closely aligned with *E. coli* serotype, including O22:H1, O25:H1 and O6:H1. However, most of the Sydney isolates clustered within the larger O6:H1 group. To further interrogate observed diversity within the O6:H1 group of isolates and to place Sydney strains within a broader global context, we expanded the SNP-based phylogenetic analysis to include additional ST73 sequence reads obtained from public sequence read archives (NCBI/EMBL/DDBJ) and the six complete ExPEC genomes with the 16 Sydney genomes. The inferred phylogeny of the 226 strains (Figs 2 and S2 for all strain labels and branch support values) shows strong major branch support. Analysis of the SNP-derived phylogenetic tree shows correlation with observed serotypes O6:H1, O25:H1 and O22:H1, with most strains observed within the O6:H1 cluster. Strain 2011_82 could not be assigned an O-type from *in* silico serotyping, but was identified as H1 and clustered most closely with O6:H1 isolates (Fig. 2).

Examination of ST73 phylogenetic structure reveals four significant clades (Figs 2 and S2): one exclusively associated with serotype O6:H1 (group A), one exclusively associated with serotype O2:H1 (group B), one primarily serotype O6:H1 with a subclade of O18:H1 (group C), and finally a polyphyletic group consisting of O2:H1, O6:H1, O22:H1, O25:H1 and O120:H31 serotypes (group D). Strains sequenced from Sydney separated into eight distinct groups correlated with serotype (Fig. 2 – red labels). Isolates with O22:H1 and O25:H1 serotypes formed their own clusters, while serotype O6:H1 separated into five distinct clusters (O6-1 : 2009_38/45; O6-2 : 2009_6/2011_69; O6-3 : 2009_44; O6-4 : 2009_5/14/16/2010_64; O6-5 : 2009_8). Strain 2011_82 clustered with other O6:H1 strains and formed the final group (Ox).

Virulence profiles of Sydney strains

We identified differences in virulence gene profiles of *E. coli* strains examined in this study (Fig. 1). Significantly, we found that virulence gene profiles were largely consistent in strains from the same phylogenetic groups O25, O22, O6-1 – O6-5, Ox (Fig. 1, see Tables S3 and S4 for

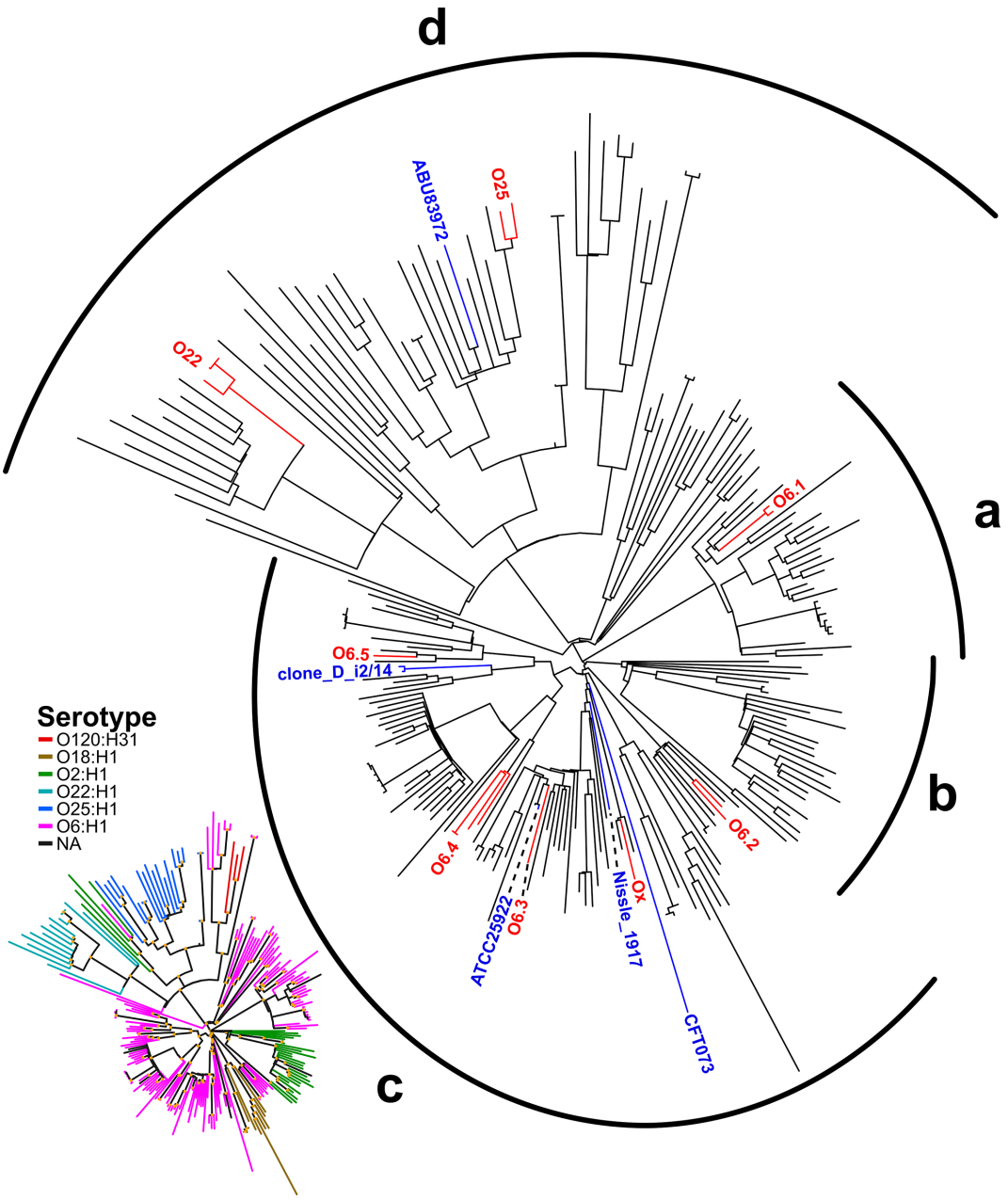

Fig. 2. SNP-based maximum-likelihood phylogram of 226 ST73 strains. A more detailed tree with branch support values and tip labels can be found in Fig. S2. ST73 isolates separate into four distinct groups, labelled A–D, which correlate well with *in silico* serotyping (inset). ST73 isolates sequenced in this study cluster into eight distinct groups, shown in red, high-quality complete ST73 genomes are shown in blue. Trees were reconstructed using 18 426 SNPs identified by read mapping to the clone D i2 reference sequence, reduced from 27 568 SNPs by filtering of recombination regions.

greater detail). For adhesion-related genes, critical components of P fimbriae (*papACDEFGHJK*) were absent in six strains (Fig. 1). Additionally, in strains of group O25, genes encoding the type I fimbriae major subunit (*fimA*), periplasmic chaperone (*fimC*), regulatory subunit (*fimE*) and the fimbriae-associated *fimI* were missing in BLASTN searches. Other genes encoding F1C fimbriae, curli fibres, type IV pili, *E. coli* common pili and the *fdeC* adhesion genes were present in all strains. The importance of FdeC as a putative virulence factor is underpinned by the observation that it is: (i) a broadly conserved *E. coli* adhesin whose expression is upregulated on the surface of UPEC when it contacts host cells; and (ii) a major target during humoral immune responses that significantly reduced kidney colonization in mice challenged transurethrally with UPEC strain 536 [45].

Table 1. Phenotypic antimicrobial resistance

Patient	AMP	AMC	LEX	CIP	GEN	NIT	NOR	TMP
2009_3	R	R	S	S	S	S	S	S
2009_5	R	R	S	S	S	S	S	S
2009_6	R	R	S	S	S	S	–	S
2009_8	R	S	S	S	S	S	–	S
2009_13	R	S	S	S	S	S	S	S
2009_14/16	R	S	S	S	S	S	–	S
2009_21	R	S	S	S	S	S	–	S
2009_38	R	S	S	S	S	S	S	S
2009_44	R	–	S	S	S	S	S	S
2009_45	R	S	S	S	S	S	S	S
2010_64	R	S	S	S	S	S	–	S
2011_69	R	S	S	S	S	R	S	S
2011_82	S	–	S	S	S	S	S	R
2011_93/98	R	S	S	S	S	S	–	S

AMC, amoxicillin-clavulanic acid; AMP, ampicillin; CIP, ciprofloxacin; GEN, gentamicin; LEX, cephalexin; NIT, nitrofurantoin; NOR, norfloxacin; R, resistant; S, susceptible; TMP, trimethoprim.

Iron acquisition is critical for the growth of ExPEC in low-iron environments *in vivo*, and it is not uncommon to identify genes linked to siderophore production and processing in UPEC. Complete enterobactin, salmochelin and yersiniabactin gene clusters were identified in all ST73 strains, while aerobactin genes were identified in all strains except those belonging to group O22. Genes for haem uptake, including the *chu* operon and *hma* gene, were present in all strains, as were those related to iron uptake such as the *sit* ABC transporter operon and ferric I *Yersinia* uptake (*fyuA*) gene. In contrast, the putative iron-uptake gene cluster *eitABCD* and adhesion/iron-uptake gene *ireA* were only identified in a subset of strains. In addition to iron uptake, genes encoding copper resistance have also been linked to virulence [46] and antimicrobial resistance [47]. The *cus* system, encoding a four-component copper efflux pump, was present and complete in all strains. However, in strain 2009_45, *cueR*, an important regulator controlling copper detoxification and efflux *copA* and *cueO* genes, was not located in all searches.

Larger differences were observed in the presence of toxin genes. Strains from group O25 and strain 2009_8 contained the highest number of toxin genes, including cytotoxic necrotizing factor 1 (*cnf1*), the haemolysin (*hly*ABCD) cluster, haemolysin E (*hlyE*) and secreted autotransporter toxin (*sat*). Genes that have been previously shown to promote propagation of *E. coli* in blood, such as proteases *pic* and *tsh* and the increased serum survival (*iss*) gene, were present in all strains, as well as the cellular invasion promoting *ibe* gene cluster. Closely related *hek* and *tia* genes, associated with epithelial cell invasion in neonatal meningitis-causing and enterotoxigenic *E. coli*, respectively, are both found in separate strains. Furthermore, *tcpC* associated with immune modulation via inhibition of Toll/IL-1 receptor signalling was only found in groups O6-1-5 and Ox.

Antibiotic resistance

All *intI1*-positive isolates were tested for resistance to ampicillin, cefotaxime, chloramphenicol, streptomycin, sulfafurazole and trimethoprim using the CDS method. Strain 2011_82 did not have a class 1 integrase gene and was not tested. All strains were resistant to ampicillin, streptomycin and sulfafurazole (Table 1). Genes encoding resistance to these antibiotics were all accounted for in the genome sequence data by the class 1 integron-associated genes *aadA1* and *sul1*, as well as one of three *bla* gene variants (Table S5). Only strain 2009_45 was resistant to the third-generation cephalosporin cefotaxime, likely due to the presence of the $bla_{\text{OXA-1}}$ gene. However, this resistance was not observed in strain 2009_38, which contained an almost identical antimicrobial-resistance region, suggesting this gene is not expressed in this strain. Interestingly, both of these strains also showed phenotypic resistance to chloramphenicol despite only 2009_45 containing a complete copy of the *catA1* gene. The full repertoire of antibiotic-resistance genes found in the 16 Sydney ST73 strains is presented in Fig. 1.

Structure of class 1 integrons in ST73 strains from Sydney

All locally sourced strains in this study, excepting 2011_82, were positive for a complete copy of the sulfonamide-resistance gene *sul1*, a structural marker of the 3′-conserved segment (3′-CS) of class 1 integrons. Similarly, all strains contained the aminoglycoside-resistance gene cassette *aadA1*. Strain 2011_82 was found to contain only a class 2 integron carrying the standard *dfrA1-sat2-aadA1* cassette array, resulting in trimethoprim resistance as tested by the hospital upon initial isolation.

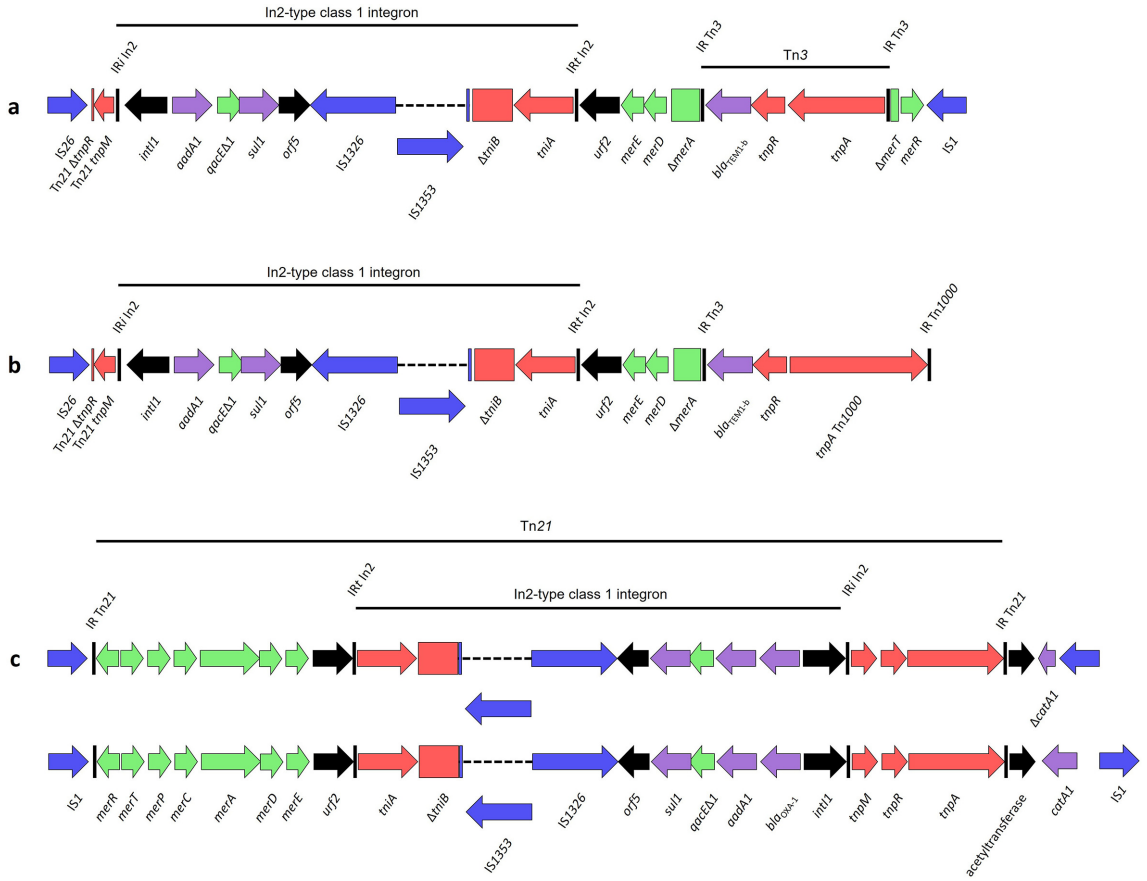

Fig. 3. Schematic representations of integron structures found within this collection.

There were three class 1 integron-containing resistance regions represented within our collection (Fig. 3), all containing the same base structures with minor variations. The first structure was identified in 11 out of 15 class 1 integron-containing isolates (Fig. 3a). It consisted of an In-2 type class 1 integron with and *aadA1* gene cassette housed within an incomplete Tn*21* transposon, matching (99 % sequence identity) the sequence in the R100 plasmid identified in Japan in the 1950s from *Shigella flexneri* (accession no. NC_002134.1) [48]. However, our structure bears an IS*26*-mediated partial deletion of the Tn*21 tnpR* gene, which is a signature that has been reported previously twice within a UPEC strain from Australia, and in association with a different class 1 integron structure [49]. A Tn*3* transposon has inserted within the *mer* module of Tn*21* with partial deletion of *merA* and *merT*, and complete deletion of *merC* and *merP*. The transposon is abutted downstream of *merR* by an inward facing IS*1* insertion element. One strain, 2009–64, housed this exact structure apart from the Tn*21 tnpM*, which appears to have been lost due to an IS*26*-mediated deletion event.

The complex resistance locus (CRL) shown in Fig. 3(b) was identified in isolates 2009–6 and 2011–69, and shares homology with the structure in Fig. 3(a). It bore identical IS*26* and Tn*3* insertion points, with the only major difference being a crossover event where the standard Tn*3 tnpA* gene and terminal inverted repeat have been replaced by that of Tn*1000*, a transposon originally identified in a cosmid clone of a human DNA sequence in 1995 [50]. This signature was recently identified in the sequence of an unannotated plasmid of a *Salmonella enterica* serovar Typhi strain sequenced as part of a larger study of Typhi from typhoid-endemic regions of Asia and Africa (accession no. LT904889.1). This is, therefore, the first report of this hybrid transposon and its presence in an *E. coli* isolated in Australia, to the best of our knowledge. Due to the nature of Illumina sequence technology, we have no confirmed sequence information downstream of Tn*3*/Tn*1000*.

Structure 3 (Fig. 3c) shares homology with the previously discussed structures. However, this CRL, present in strains 2009–38 and 2009–45, has a $bla_{\text{OXA-1}}$ gene cassette within the integron cassette array in addition to *aadA1*. Here, the Tn*21* transposon housing the class 1 integron is complete, with both the initial and terminal inverted repeats intact, and has an inward facing IS*1* flanking its *mer* end. There are two variants of this CRL in our collection, one of which

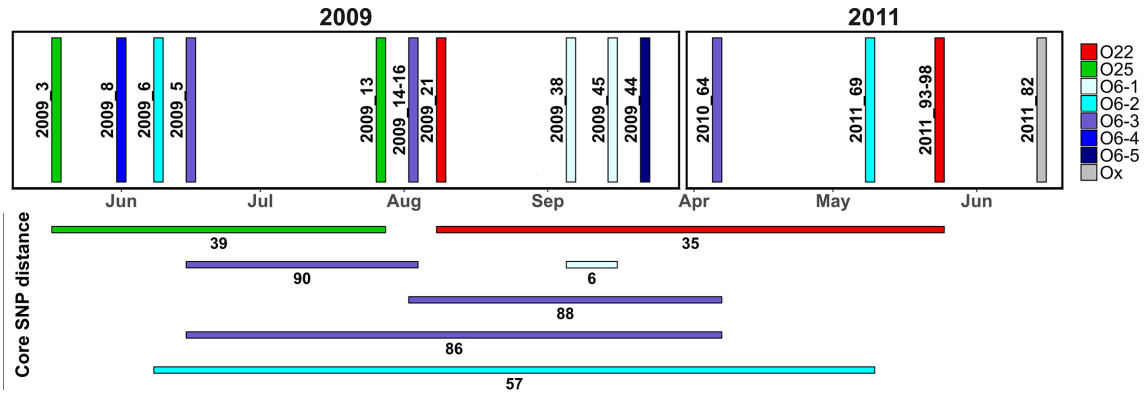

Fig. 4. Epidemic curve for ST73 isolates sequenced in this study. Isolates are coloured to match groups identified by comparison to isolates from the SRA. Core SNP distances between samples of the same group are shown with coloured horizontal lines.

contains a complete *catA1* gene downstream of Tn*21* followed by a second IS*1* element in the same orientation as the first, with the intergenic ORF identified as an acetyltransferase. This appears to be an established insertion event, with numerous reports in GenBank. In the second variant, the terminal IS*1* is inverted with consequent deletion of 476 bp of the *catA1* gene, forming a signature unique to isolates 2009–38 and 2009–45. This integron has been reported in its entirety in *Shigella dysenteriae* 1 plasmid p3099-85 (KT754164.1), *Salmonella enterica* serovar Typhimurium plasmid pUO-StVR2 (AM991977.1) and *Salmonella enterica* serovar Typhimurium strain T000240 (AP011957.1) Less than 10 SNPs were identified in comparative BLASTN alignments spanning the integron.

Sixty genomes from the SRA cohort returned adequate alignments to integron marker genes, with 25 of these appearing to possibly have only the base class 1 integron with an *aadA1* cassette, but no indication of a bordering Tn*21* transposon. Eight contained an *intI1* gene but no *aadA1*, suggesting the likely presence of a class 1 integron with a different cassette array. Sixteen contained *aadA1*, but no class 1 integrase; this could indicate a deletion event or more likely the presence of *aadA1* in a class 2 integron, though the *aadA1* gene can also exist independent of integron context. Five genomes contained an unidentifiable integron structure, possibly variants of those described in the Sydney collection, although it is impossible to say this definitively from read alignments against the abridged gene database used here.

Only three genomes, HVH_93_4–5851025, MOD1-EC6690 and MOD1-EC6783, contained all marker genes necessary to potentially contain integron C (Fig. 3). However, within the SRA cohort, the presence of integrons A and B could not be confirmed.

All 16 strains of ST73 that were sourced from Sydney were shown to carry one or more plasmids (up to five) that ranged in size from 15 to 180 kb. Only one plasmid in each strain hybridized with the *intI1* probe (data not shown). The sizes varied greatly between 80 kb and >200 kb.

DISCUSSION

This study forms a part of wider global efforts to further understand the structure of disease-causing ST73 clones. Whole-genome sequencing and maximum-likelihood phylogenetic analyses of these clones is providing important information on the community structure of ExPEC. Here, we examined 16 ST73 isolates sourced from a single hospital and used sequence data sourced from the SRA to place these isolates into a broader global context and aid in identifying clonal lineages. Phylogenetic trees from this combined dataset, when overlaid with geographical and temporal data sourced from EnteroBase (data not shown), indicate that ST73 is globally disseminated in a manner similar to ST131 [51, 52], which is currently the most studied pandemic ExPEC lineage due to the frequency of CTX-M gene carriage. However, while ST131 tends to be relatively conserved in terms of core genome, ST73 appears more variable. Analysis of locally sequenced strains and comparison to globally sourced reads from public databases can provide context that can allow the identification of outbreak clusters with more confidence than using total SNP counts alone, and may help elucidate key outbreak groups and improve public health control of disease. This is valuable as the identification of clonal groups associated with outbreaks within larger bacterial populations remains a challenge.

Characterization of molecular signatures can also assist in the identification of outbreaks as their transfer requires physical proximity of cells. CRL including integrons and transposons are common sites of genetic rearrangement and frequently carry unique molecular signatures due to insertion elements such as IS*26* [53–55]. While all class 1 integrons in the Sydney collection are not necessarily novel, there are IS-mediated signature deletions that do not appear to have been widely reported based on the current literature, such as that of the *catA1* gene. This suggests that these are

local integron variants, an idea consistent with the lack of these structures in the global SRA cohort. The major representative class 1 integron described here has been reported in its entirety once within an Australian *E. coli* O2:K1:H7 ST95 strain isolated from a bloodstream infection in 2010 (K. G. K. Goh *et al*, unpublished data; GenBank accession no. CP021289.1). This integron also shares an IS*26*-mediated deletion of the Tn*21* resolvase gene *tnpR* with plasmid pUO-SeVR1 from a Spanish *Salmonella enterica* serovar Enteritidis strain sourced from a child with gastroenteritis [49, 56]. This is significant as this precise signature is likely the product of a single event. As such, a lateral transfer event is a likely explanation for the occurrence of this signature in disparate and geographically separate strains, followed by changes in class 1 integron cassette content. Based on the plasmid typing and PFGE data, it is likely that transfer of these integrons is being facilitated by IncF plasmids similar to pUO-SeVR1, as this is the major plasmid incompatibility type within our ST73 collection and our S1-PFGE data confirm that the class 1 integrons described here are plasmid-borne. Plasmids appear to increasingly play an important role in the mobilization of drug-resistance genes in ExPEC ST73, and their characterization relies heavily on the use of whole-genome sequencing (ideally long-read) and read-mapping technologies such as those described here.

Whole-genome sequencing allows for the analysis of gene presence/absence in clinical isolates, which will provide data on the importance of virulence genes in pathogenesis. The virulence profiles of strains sequenced in this study are consistent with other examinations of virulence in ST73 and in ExPEC more broadly. Genes encoding P fimbrial adhesins, the aerobactin siderophore *(iuc/iut*), and toxins haemolysin A and cytotoxic necrotizing factor 1 are not universally identified in worldwide ExPEC populations sourced from humans and animals [1, 57–59]. In previous work on ST73 isolates sourced from the UK, the prevalence of these genes/gene families was also non-universal; however, *hlyA* and *cnf1* showed a substantially higher prevalence in ST73 compared with ExPEC-associated ST10, ST69 and ST95 [1]. In isolates sourced from Sydney, a relatively clear association could be identified between phylogenetic groups and virulence profiles. Further *in silico* categorization of virulence profiles using global ST73 reads would provide insight into virulence patterns/groups within ST73 and ExPEC, which could potentially lead to improved response, prevention and treatment of ExPEC-linked disease.

In endemic pathogens like *E. coli*, genetic comparisons of clonal group and mobile genetic element diversity can be difficult to perform with localized populations, as high numbers of closely related isolates are required for robust SNP-phylogenetic analysis and this may require the long-term collection of bacterial isolates to isolate a sufficient number of representatives. Here, we used Illumina sequencing combined with SNP-phylogenetic methods to identify at least eight distinct clonal lineages in a pilot sample of 16 ST73 isolates collected from a single hospital, indicating the wealth of diversity within the ST73 population sourced from highly localized sampling over an extended period (Fig. 4). Contrastingly, the diversity of mobile elements within this cohort is much less profound. Only three resistance-containing class 1 integron structures were identified, all were linked to plasmids, and all showed high structural similarity. Our study is an example of how genome sequencing can provide a depth of information not available with previous molecular epidemiology methodologies, which is useful in the determination of outbreak groups among ST73.

Funding information
This work was supported by the Australian Research Council, linkage grant LP150100912. This project was partly funded by the Australian Centre for Genomic Epidemiological Microbiology (Ausgem), a collaborative partnership between the NSW Department of Primary Industries and the University of Technology Sydney. J. M. is a recipient of Australian Government Research Training Program Scholarships.

Acknowledgements
We acknowledge the efforts of staff from the Sydney Adventist Hospital for providing the Sydney ST73 strains and associated meta-data for this study.

Author contributions
D. R. B. performed the phylogenetic analysis, contributed to gene presence analyses, prepared most of the figures, tables and supplementary data, and drafted multiple iterations of the manuscript. J. M. isolated genomic DNA, contributed to gene presence analyses, characterized structures of class 1 integrons of Sydney strains, as well as generating figures and contributing to writing the manuscript. M. L. constructed libraries and sequenced the Sydney ST73 strains. C. V. and J. I. performed the S1-PGFE and Southern blot analyses, and edited final drafts of the manuscript. A. E. D. assisted with data analysis. P. R. C. initiated the collaboration with the Sydney Adventist Hospital, screened isolates to create the strain collection included in this study and helped J. M. with analysis of the data presented. S. P. D. initiated and coordinated the project, and drafted iterations of the final manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Data Bibliography
1. Public Health England. NCBI Sequence Read Archive. SRR3578961 (Sample name: 129511; BioSample: SAMN05170964). (2016).

2. Public Health England. NCBI Sequence Read Archive. SRR3578647 (Sample name: 132052; BioSample: SAMN05170806). (2016).

3. Public Health England. NCBI Sequence Read Archive. SRR3578295 (Sample name: 142256; BioSample: SAMN05170584). (2016).

4. Public Health England. NCBI Sequence Read Archive. SRR3578905 (Sample name: 142261; BioSample: SAMN05170869). (2016).

5. Public Health England. NCBI Sequence Read Archive. SRR4787808 (Sample name: 147737; BioSample: SAMN05965817). (2016).

6. Public Health England. NCBI Sequence Read Archive. SRR3578984 (Sample name: 164650; BioSample: SAMN05171017). (2016).

7. Public Health England. NCBI Sequence Read Archive. SRR3578606 (Sample name: 164662; BioSample: SAMN05170699). (2016).

8.Public Health England. NCBI Sequence Read Archive. SRR3578926 (Sample name: 177456; BioSample: SAMN05170891). (2016).

9.Public Health England. NCBI Sequence Read Archive. SRR3578811 (Sample name: 178725; BioSample: SAMN05170867). (2016).

10. Public Health England. NCBI Sequence Read Archive. SRR5006246 (Sample name: 186548; BioSample: SAMN06006300). (2016).

11. Public Health England. NCBI Sequence Read Archive. SRR3578795 (Sample name: 187570; BioSample: SAMN05170851). (2016).

12. Public Health England. NCBI Sequence Read Archive. SRR3579386 (Sample name: 194182; BioSample: SAMN05171056). (2016).

13. Public Health England. NCBI Sequence Read Archive. SRR3578800 (Sample name: 195747; BioSample: SAMN05170856). (2016).

14. Public Health England. NCBI Sequence Read Archive. SRR3581418 (Sample name: 195748; BioSample: SAMN05171876). (2016).

15. Public Health England. NCBI Sequence Read Archive. SRR3574289 (Sample name: 208732; BioSample: SAMN05163766). (2016).

16. Public Health England. NCBI Sequence Read Archive. SRR3574256 (Sample name: 208733; BioSample: SAMN05163733). (2016).

17. Public Health England. NCBI Sequence Read Archive. SRR3574282 (Sample name: 208779; BioSample: SAMN05163759). (2016).

18. Public Health England. NCBI Sequence Read Archive. SRR3574318 (Sample name: 209934; BioSample: SAMN05163795). (2016).

19. Public Health England. NCBI Sequence Read Archive. SRR3574227 (Sample name: 209958; BioSample: SAMN05163704). (2016).

20. Public Health England. NCBI Sequence Read Archive. SRR5017311 (Sample name: 222576; BioSample: SAMN06014912). (2016).

21. Public Health England. NCBI Sequence Read Archive. SRR3574247 (Sample name: 241766; BioSample: SAMN05163724). (2016).

22. Public Health England. NCBI Sequence Read Archive. SRR4897304 (Sample name: 251826; BioSample: SAMN05980507). (2016).

23. Public Health England. NCBI Sequence Read Archive. SRR4897271 (Sample name: 279352; BioSample: SAMN05980484). (2016).

24. Public Health England. NCBI Sequence Read Archive. SRR4787388 (Sample name: 279505; BioSample: SAMN05965667). (2016).

25. Public Health England. NCBI Sequence Read Archive. SRR4788298 (Sample name: 279506; BioSample: SAMN05966834). (2016).

26. Public Health England. NCBI Sequence Read Archive. SRR4897089 (Sample name: 280635; BioSample: SAMN05980256). (2016).

27. Arizona State Public Health Laboratory. NCBI Sequence Read Archive. SRR2890033 (Sample name: AZ_TG78624; BioSample: SAMN04125201). (2015).

28. Arizona State Public Health Laboratory. NCBI Sequence Read Archive. SRR2890035 (Sample name: AZ_TG78632; BioSample: SAMN04125203). (2015).

29. Arizona State Public Health Laboratory. NCBI Sequence Read Archive. SRR2890041 (Sample name: AZ_TG78656; BioSample: SAMN04125209). (2015).

30. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314217 (Sample name: blood-08-0215; BioSample: SAMN02801814). (2014).

31. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314218 (Sample name: blood-08-0379; BioSample: SAMN02801815). (2014).

32. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314283 (Sample name: blood-10-1105; BioSample: SAMN02801880). (2014).

33. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314284 (Sample name: blood-10-1126; BioSample: SAMN02801881). (2014).

34. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314291 (Sample name: blood-10-1386; BioSample: SAMN02801888). (2014).

35. Weimer, B. NCBI Sequence Read Archive. SRR1122534 (Sample name: ESC0187; BioSample: SAMN02368188). (2015).

36. Weimer, B. NCBI Sequence Read Archive. SRR1122531 (Sample name: ESC0196; BioSample: SAMN02368193). (2015).

37. Weimer, B. NCBI Sequence Read Archive. SRR1840546 (Sample name: ESC0215; BioSample: SAMN02368210). (2016).

38. Weimer, B. NCBI Sequence Read Archive. SRR1122520 (Sample name: ESC0216; BioSample: SAMN02368211). (2015).

39. Weimer, B. NCBI Sequence Read Archive. SRR1840548 (Sample name: ESC0221; BioSample: SAMN02368213). (2016).

40. Charlesworth, J. NCBI Sequence Read Archive. SRR3051035 (Sample name: HICF103; BioSample: SAMN04357686). (2016).

41. Charlesworth, J. NCBI Sequence Read Archive. SRR3050918 (Sample name: HICF121; BioSample: SAMN04357569). (2016).

42. Charlesworth, J. NCBI Sequence Read Archive. SRR3050920 (Sample name: HICF123; BioSample: SAMN04357571). (2016).

43. Charlesworth, J. NCBI Sequence Read Archive. SRR3050923 (Sample name: HICF128; BioSample: SAMN04357574). (2016).

44. Charlesworth, J. NCBI Sequence Read Archive. SRR3050949 (Sample name: HICF164; BioSample: SAMN04357600). (2016).

45. Charlesworth, J. NCBI Sequence Read Archive. SRR3050950 (Sample name: HICF165; BioSample: SAMN04357601). (2016).

46. Charlesworth, J. NCBI Sequence Read Archive. SRR3050960 (Sample name: HICF184; BioSample: SAMN04357611). (2016).

47. Charlesworth, J. NCBI Sequence Read Archive. SRR3050968 (Sample name: HICF196; BioSample: SAMN04357619). (2016).

48. Charlesworth, J. NCBI Sequence Read Archive. SRR3050972 (Sample name: HICF200; BioSample: SAMN04357623). (2016).

49. Charlesworth, J. NCBI Sequence Read Archive. SRR3050975 (Sample name: HICF204; BioSample: SAMN04357626). (2016).

50. Charlesworth, J. NCBI Sequence Read Archive. SRR3050977 (Sample name: HICF207; BioSample: SAMN04357628). (2016).

51. Charlesworth, J. NCBI Sequence Read Archive. SRR3050980 (Sample name: HICF210; BioSample: SAMN04357631). (2016).

52. Charlesworth, J. NCBI Sequence Read Archive. SRR3050984 (Sample name: HICF216; BioSample: SAMN04357635). (2016).

53. Charlesworth, J. NCBI Sequence Read Archive. SRR3050993 (Sample name: HICF227; BioSample: SAMN04357644). (2016).

54. Charlesworth, J. NCBI Sequence Read Archive. SRR3050860 (Sample name: HICF23; BioSample: SAMN04357511). (2016).

55. Charlesworth, J. NCBI Sequence Read Archive. SRR3050999 (Sample name: HICF237; BioSample: SAMN04357650). (2016).

56. Charlesworth, J. NCBI Sequence Read Archive. SRR3051022 (Sample name: HICF28; BioSample: SAMN04357673). (2016).

57. Charlesworth, J. NCBI Sequence Read Archive. SRR3050871 (Sample name: HICF41; BioSample: SAMN04357522). (2016).

58. Charlesworth, J. NCBI Sequence Read Archive. SRR3050878 (Sample name: HICF48; BioSample: SAMN04357529). (2016).

59. Charlesworth, J. NCBI Sequence Read Archive. SRR3051028 (Sample name: HICF59; BioSample: SAMN04357679). (2016).

60. Charlesworth, J. NCBI Sequence Read Archive. SRR3050891 (Sample name: HICF66; BioSample: SAMN04357542). (2016).

61. Charlesworth, J. NCBI Sequence Read Archive. SRR3050851 (Sample name: HICF9; BioSample: SAMN04357510). (2016).

62. Charlesworth, J. NCBI Sequence Read Archive. SRR3050904 (Sample name: HICF90; BioSample: SAMN04357555). (2016).

63. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800527, SRR783923, SRR783922, SRR800547 [Sample name: HVH 103 (4-5904188); BioSample: SAMN01885751]. (2013).

64. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B.

NCBI Sequence Read Archive. SRR783978, SRR783979, SRR800524, SRR800573 [Sample name: HVH 109 (4-6977162); BioSample: SAMN01885757]. (2013).

65. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR770082, SRR958202, SRR958203, SRR784020 [Sample name: HVH 114 (4-7037740); BioSample: SAMN01885762]. (2013).

66. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958215, SRR770085, SRR784023, SRR958214 [Sample name: HVH 116 (4-6879942); BioSample: SAMN01885765]. (2013).

67. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958218, SRR958217, SRR784024, SRR770086 [Sample name: HVH 117 (4-6857191); BioSample: SAMN01885766]. (2013).

68. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958223, SRR958222, SRR784027, SRR770089 [Sample name: HVH 120 (4-6978681); BioSample: SAMN01885769]. (2013).

69. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784129, SRR800591, SRR784130, SRR800539 [Sample name: HVH 13 (4-7634056); BioSample: SAMN01885678]. (2013).

70. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR770097, SRR958246, SRR958247, SRR784103 [Sample name: HVH 132 (4-6876862); BioSample: SAMN01885777]. (2013).

71. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958262, SRR770105, SRR784174, SRR958263 [Sample name: HVH 142 (4-5627451); BioSample: SAMN01885787]. (2013).

72. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR770107, SRR958267, SRR958266, SRR784176 [Sample name: HVH 144 (4-4451937); BioSample: SAMN01885789]. (2013).

73. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR770121, SRR958296, SRR958295, SRR784243 [Sample name: HVH 156 (4-3206505); BioSample: SAMN01885801]. (2013).

74. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784244, SRR958297, SRR958298, SRR770123 [Sample name: HVH 157 (4-3406229); BioSample: SAMN01885802]. (2013).

75. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784265, SRR958301, SRR958302, SRR770125 [Sample name: HVH 159 (4-5818141); BioSample: SAMN01885804]. (2013).

76. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784313, SRR800553, SRR800463, SRR784314 [Sample name: HVH 16 (4-7649002); BioSample: SAMN01885679]. (2013).

77. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958303, SRR958304, SRR770126, SRR784286 [Sample name: HVH 160 (4-5695937); BioSample: SAMN01885805]. (2013).

78. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR770128, SRR958305, SRR958306, SRR784287 [Sample name: HVH 161 (4-3119890); BioSample: SAMN01885806]. (2013).

79. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR770133, SRR958313, SRR958314, SRR784316 [Sample name: HVH 169 (4-1075578); BioSample: SAMN01885811]. (2013).

80. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J.,

McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784396, SRR958320, SRR770136, SRR958319 [Sample name: HVH 172 (4-3248542); BioSample: SAMN01885814]. (2013).

81. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800420, SRR784980, SRR784979, SRR800461 [Sample name: HVH 2 (4-6943160); BioSample: SAMN01885669]. (2013).

82. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800575, SRR800491, SRR784652, SRR784653 [Sample name: HVH 20 (4-5865042); BioSample: SAMN01885683]. (2013).

83. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784678, SRR958401, SRR958400, SRR770170 [Sample name: HVH 207 (4-3113221); BioSample: SAMN01885846]. (2013).

84. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958435, SRR784707, SRR784706, SRR958436 [Sample name: HVH 212 (3-9305343); BioSample: SAMN01885851]. (2013).

85. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784759, SRR958443, SRR958444, SRR784758 (Sample name: HVH 216 (4-3042952); BioSample: SAMN01885855). (2013).

86. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784763, SRR958448, SRR958447, SRR784762 [Sample name: HVH 218 (4-4500903); BioSample: SAMN01885857]. (2013).

87. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958450, SRR784780, SRR784778, SRR958449 [Sample name: HVH 220 (4-5876842); BioSample: SAMN01885858]. (2013).

88. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR784927, SRR958459, SRR958460, SRR784928 [Sample name: HVH 227 (4-2277670); BioSample: SAMN01885863]. (2013).

89. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800538, SRR784985, SRR800588, SRR784986 [Sample name: HVH 27 (4-7449267); BioSample: SAMN01885690]. (2013).

90. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800439, SRR800426, SRR785017, SRR785018 [Sample name: HVH 31 (4-2602156); BioSample: SAMN01885694]. (2013).

91. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785074, SRR800438, SRR800533, SRR785075 [Sample name: HVH 37 (4-2773848); BioSample: SAMN01885699]. (2013).

92. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800499, SRR785079, SRR785078, SRR800431 [Sample name: HVH 39 (4-2679949); BioSample: SAMN01885701]. (2013).

93. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785080, SRR785081, SRR800471, SRR800498 [Sample name: HVH 40 (4-1219782); BioSample: SAMN01885702]. (2013).

94. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800560, SRR785301, SRR785300, SRR800479 [Sample name: HVH 58 (4-2839709); BioSample: SAMN01885716]. (2013).

95. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800601, SRR785327, SRR785326, SRR800484 [Sample name: HVH 61 (4-2736020); BioSample: SAMN01885718]. (2013).

96. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A.

M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800475, SRR785356, SRR800545, SRR785357 [Sample name: HVH 68 (4-0888028); BioSample: SAMN01885721]. (2013).

97. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800598, SRR785411, SRR785410, SRR800496 [Sample name: HVH 7 (4-7315031); BioSample: SAMN01885674]. (2013).

98. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800506, SRR785409, SRR800467, SRR785408 [Sample name: HVH 74 (4-1034782); BioSample: SAMN01885725]. (2013).

99. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785438, SRR800572, SRR785437, SRR800423 [Sample name: HVH 77 (4-2605759); BioSample: SAMN01885727]. (2013).

100. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785469, SRR800583, SRR800548, SRR1010959, SRR1010958, SRR785470 [Sample name: HVH 83 (4-2051087); BioSample: SAMN01885732]. (2013).

101. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785498, SRR785497, SRR800536, SRR800424 [Sample name: HVH 86 (4-7026218); BioSample: SAMN01885735]. (2013).

102. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800537, SRR785526, SRR785525, SRR800515 [Sample name: HVH 89 (4-5885604); BioSample: SAMN01885738]. (2013).

103. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785533, SRR800430, SRR800582, SRR785532 [Sample name: HVH 92 (4-5930790); BioSample: SAMN01885741]. (2013).

104. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785552, SRR785551, SRR800453, SRR800534 [Sample name: HVH 93 (4-5851025); BioSample: SAMN01885742]. (2013).

105. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR800454, SRR785581, SRR800472, SRR785583 [Sample name: HVH 95 (4-6074464); BioSample: SAMN01885743]. (2013).

106. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785596, SRR785597, SRR800594, SRR800460 [Sample name: HVH 96 (4-5934869); BioSample: SAMN01885744]. (2013).

107. Hamilton, S. NCBI Sequence Read Archive. SRR785703, SRR958505, SRR785702, SRR958506 [Sample name: KOEGE 43 (105a); BioSample: SAMN01885872]. (2013).

108. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3951664, SRR4101513, SRR3466132, SRR3957118, SRR3466142 (Sample name: MOD1-EC5001; BioSample: SAMN04279329). (2016).

109. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3993808 (Sample name: MOD1-EC5192; BioSample: SAMN04279531). (2016).

110. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3987994 (Sample name: MOD1-EC5696; BioSample: SAMN05452879). (2016).

111. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3993878 (Sample name: MOD1-EC5875; BioSample: SAMN05468036). (2016).

112. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3987714 (Sample name: MOD1-EC5956; BioSample: SAMN05439405). (2016).

113. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3989640 (Sample name: MOD1-EC6122; BioSample: SAMN05439519). (2016).

114. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4260307 (Sample name: MOD1-EC662; BioSample: SAMN05591566). (2016).

115. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3987641 (Sample name: MOD1-EC6690; BioSample: SAMN04992520). (2016).

116. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4261194 (Sample name: MOD1-EC673; BioSample: SAMN05591556). (2016).

117. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI

Sequence Read Archive. SRR3974429 (Sample name: MOD1-EC6744; BioSample: SAMN04992575). (2016).

118. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3974861 (Sample name: MOD1-EC6783; BioSample: SAMN04992131). (2016).

119. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4340471 (Sample name: MOD1-EC679; BioSample: SAMN05591550). (2016).

120. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4261569 (Sample name: MOD1-EC681; BioSample: SAMN05591548). (2016).

121. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4261646 (Sample name: MOD1-EC682; BioSample: SAMN05591547). (2016).

122. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4262058 (Sample name: MOD1-EC691; BioSample: SAMN05591538). (2016).

123. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4262121 (Sample name: MOD1-EC692; BioSample: SAMN05591537). (2016).

124. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4262526 (Sample name: MOD1-EC695; BioSample: SAMN05591534). (2016).

125. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4262643 (Sample name: MOD1-EC697; BioSample: SAMN05591532). (2016).

126. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4262710 (Sample name: MOD1-EC699; BioSample: SAMN05591530). (2016).

127. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4262859 (Sample name: MOD1-EC705; BioSample: SAMN05591524). (2016).

128. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4263453 (Sample name: MOD1-EC713; BioSample: SAMN05591516). (2016).

129. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4263643 (Sample name: MOD1-EC714; BioSample: SAMN05591650). (2016).

130. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4263647 (Sample name: MOD1-EC719; BioSample: SAMN05591660). (2016).

131. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR4263652 (Sample name: MOD1-EC724; BioSample: SAMN05591655). (2016).

132. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3951485, SRR4098824 (Sample name: MOD1-ECOR23; BioSample: SAMN04158360). (2016).

133. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3989533 (Sample name: MOD1-ECOR51; BioSample: SAMN05439312). (2016).

134. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3989534 (Sample name: MOD1-ECOR54; BioSample: SAMN05439311). (2016).

135. Gangiredla, J., Barnaba, T., Gebru, S., Mammel, M.K., Lacher, D., Tartera, C., Patel, I.R., Leonard, S., Lampel, K.A. and Elkins, C.A. NCBI Sequence Read Archive. SRR3989515 (Sample name: MOD1-ECOR57; BioSample: SAMN05452816). (2016).

136. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966595 (Sample name: PRJEB9931-B10; BioSample: SAMEA3488053). (2015).

137. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966606 (Sample name: PRJEB9931-B134; BioSample: SAMEA3488064). (2015).

138. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966596 (Sample name: PRJEB9931-B14; BioSample: SAMEA3488054). (2015).

139. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966597 (Sample name: PRJEB9931-B18; BioSample: SAMEA3488055). (2015).

140. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966598 (Sample name: PRJEB9931-B29; BioSample: SAMEA3488056). (2015).

141. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966599 (Sample name: PRJEB9931-B36; BioSample: SAMEA3488057). (2015).

142. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966600 (Sample name: PRJEB9931-B40; BioSample: SAMEA3488058). (2015).

143. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966601 (Sample name: PRJEB9931-B72; BioSample: SAMEA3488059). (2015).

144. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966602 (Sample name: PRJEB9931-B73; BioSample: SAMEA3488060). (2015).

145. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966603 (Sample name: PRJEB9931-B84; BioSample: SAMEA3488061). (2015).

146. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966607 (Sample name: PRJEB9931-U1; BioSample: SAMEA3488065). (2015).

147. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966610 (Sample name: PRJEB9931-U24; BioSample: SAMEA3488068). (2015).

148. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966611 (Sample name: PRJEB9931-U30; BioSample: SAMEA3488069). (2015).

149. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966612 (Sample name: PRJEB9931-U36; BioSample: SAMEA3488070). (2015).

150. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966613 (Sample name: PRJEB9931-U42; BioSample: SAMEA3488071). (2015).

151. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966614 (Sample name: PRJEB9931-U48; BioSample: SAMEA3488072). (2015).

152. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966615 (Sample name: PRJEB9931-U50; BioSample: SAMEA3488073). (2015).

153. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966608 (Sample name: PRJEB9931-U7; BioSample: SAMEA3488066). (2015).

154. Alhashash, F., Wang, X., Paszkiewicz, K., Diggle, M., Zong, Z. and McNally, A. EBI Sequence Read Archive. ERR966616 (Sample name: PRJEB9931-U76; BioSample: SAMEA3488074). (2015).

155. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive.

ERR1618956 (Sample name: SCK22-23; BioSample: SAMEA4429064). (2016).

156. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1617767 (Sample name: SCK27-56; BioSample: SAMEA4427832). (2016).

157. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1617912 (Sample name: SCK27-57; BioSample: SAMEA4427977). (2016).

158. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1619000 (Sample name: SCK28-81; BioSample: SAMEA4429108). (2016).

159. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1619301 (Sample name: SCK52-18; BioSample: SAMEA4429543). (2016).

160. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1617950 (Sample name: SCK62-49; BioSample: SAMEA4428015). (2016).

161. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1618164 (Sample name: SCP10-30; BioSample: SAMEA4428288). (2016).

162. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1618165 (Sample name: SCP10-31; BioSample: SAMEA4428289). (2016).

163. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1618166 (Sample name: SCP10-32; BioSample: SAMEA4428290). (2016).

164. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1618167 (Sample name: SCP10-35; BioSample: SAMEA4428291). (2016).

165. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1618190 (Sample name: SCP11-56; BioSample: SAMEA4428314). (2016).

166. Kluytmans-van den Bergh, M.F., Rossen, J.W., Bruijning-Verhagen, P.C., Bonten, M.J., Friedrich, A.W., Vandenbroucke-Grauls, C.M., Willems, R.J. and Kluytmans, J.A. EBI Sequence Read Archive. ERR1619410 (Sample name: SCP23-30; BioSample: SAMEA4429645). (2016).

167. Coil, D. NCBI Sequence Read Archive. SRR1824515 (Sample name: UCD_JA17; BioSample: SAMN02650861). (2015).

168. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785837, SRR958544, SRR785836, SRR958545 (Sample name: UMEA 3014-1; BioSample: SAMN01885885). (2013).

169. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR785945, SRR958571, SRR785946, SRR958570 (Sample name: UMEA 3121-1; BioSample: SAMN01885898). (2013).

170. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786045, SRR958596, SRR958597, SRR786046 (Sample name: UMEA 3159-1; BioSample: SAMN01885909). (2013).

171. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958607, SRR786055, SRR958606, SRR786056 (Sample name: UMEA 3172-1; BioSample: SAMN01885914). (2013).

172. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786080, SRR958609, SRR958608, SRR786079 (Sample name: UMEA 3173-1; BioSample: SAMN01885915). (2013).

173. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786106, SRR786105, SRR958612, SRR958613 (Sample name: UMEA 3175-1; BioSample: SAMN01885917). (2013).

174. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958617, SRR786110, SRR958616, SRR786109 (Sample name: UMEA 3178-1; BioSample: SAMN01885919). (2013).

175. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958637, SRR958638, SRR786214, SRR786213 (Sample name: UMEA 3208-1; BioSample: SAMN01885929). (2013).

176. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958656, SRR786246, SRR958657,

SRR786247 (Sample name: UMEA 3220-1; BioSample: SAMN01885934). (2013).

177. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958665, SRR786251, SRR786250, SRR958664 (Sample name: UMEA 3222-1; BioSample: SAMN01885936). (2013).

178. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786269, SRR958713, SRR958712, SRR786270 (Sample name: UMEA 3233-1; BioSample: SAMN01885938). (2013).

179. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958719, SRR800752, SRR800749, SRR958718 (Sample name: UMEA 3252-1; BioSample: SAMN01885941). (2013).

180. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786299, SRR958723, SRR958722, SRR786300 (Sample name: UMEA 3264-1; BioSample: SAMN01885943). (2013).

181. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786301, SRR958725, SRR958724, SRR800417 (Sample name: UMEA 3268-1; BioSample: SAMN01885944). (2013).

182. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786323, SRR958756, SRR958755, SRR800502 (Sample name: UMEA 3337-1; BioSample: SAMN01885956). (2013).

183. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786327, SRR958760, SRR958759, SRR786326 (Sample name: UMEA 3342-1; BioSample: SAMN01885958). (2013).

184. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786336, SRR958780, SRR786337, SRR958779 (Sample name: UMEA 3490-1; BioSample: SAMN01885963). (2013).

185. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958792, SRR958791, SRR786347, SRR786348 (Sample name: UMEA 3652-1; BioSample: SAMN01885969). (2013).

186. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR958818, SRR958817, SRR786362, SRR786361 (Sample name: UMEA 3694-1; BioSample: SAMN01885976). (2013).

187. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786367, SRR958833, SRR958832, SRR786368 (Sample name: UMEA 3705-1; BioSample: SAMN01885979). (2013).

188. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786370, SRR786369, SRR958835, SRR958834 (Sample name: UMEA 3707-1; BioSample: SAMN01885980). (2013).

189. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786376, SRR786375, SRR958843, SRR958842 (Sample name: UMEA 3821-1; BioSample: SAMN01885983). (2013).

190. Feldgarden, M., Frimodt-Moller, N., Leihof, R.F., Rasmussen, L., Young, S.K., Zeng, Q., Gargeya, S., Abouelleil, A., Alvarado, L., Berlin, A. M., Chapman, S.B., Gainer-Dewar, J., Goldberg, J., Gnerre, S., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T., Priest, M., Roberts, A., Saif, S., Shea, T., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. NCBI Sequence Read Archive. SRR786401, SRR958980, SRR958981, SRR786400 (Sample name: UMEA 4207-1; BioSample: SAMN01885991). (2013).

191. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314322 (Sample name: upec-109; BioSample: SAMN02801919). (2014).

192. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314402 (Sample name: upec-181; BioSample: SAMN02801999). (2014).

193. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314415 (Sample name: upec-193; BioSample: SAMN02802012). (2014).

194. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314457 (Sample name: upec-230; BioSample: SAMN02802054). (2014).

195. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314459 (Sample name: upec-232; BioSample: SAMN02802056). (2014).

196. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314464 (Sample name: upec-237; BioSample: SAMN02802061). (2014).

197. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314480 (Sample name: upec-251; BioSample: SAMN02802077). (2014).

198. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314491 (Sample name: upec-261; BioSample: SAMN02802088). (2014).

199. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314518 (Sample name: upec-287; BioSample: SAMN02802115). (2014).

200. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314519 (Sample name: upec-288; BioSample: SAMN02802116). (2014).

201. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314520 (Sample name: upec-289; BioSample: SAMN02802117). (2014).

202. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314532 (Sample name: upec-39; BioSample: SAMN02802129). (2014).

203. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314583 (Sample name: upec-85; BioSample: SAMN02802180). (2014).

204. Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T. and Shendure, J. NCBI Sequence Read Archive. SRR1314588 (Sample name: upec-9; BioSample: SAMN02802185). (2014).

References

1. **Gibreel TM, Dodgson AR, Cheesbrough J, Fox AJ, Bolton FJ *et al.*** Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. *J Antimicrob Chemother* 2012;67:346–356.
2. **Kim S, Sung JY, Cho HH, Kwon KC, Koo SH.** Characteristics of the molecular epidemiology of CTX-M-producing *Escherichia coli* isolated from a tertiary hospital in Daejeon, Korea. *J Microbiol Biotechnol* 2016;26:1643–1649.
3. **Manges AR, Johnson JR.** Reservoirs of extraintestinal pathogenic *Escherichia coli*. *Microbiol Spectr* 2015;3:UTI-0006-2012.
4. **Olesen B, Scheutz F, Menard M, Skov MN, Kolmos HJ *et al.*** Three-decade epidemiological analysis of *Escherichia coli* O15:K52:H1. *J Clin Microbiol* 2009;47:1857–1862.
5. **Yun KW, Kim DS, Kim W, Lim IS.** Molecular typing of uropathogenic *Escherichia coli* isolated from Korean children with urinary tract infection. *Korean J Pediatr* 2015;58:20–27.
6. **Johnson JR, Thuras P, Johnston BD, Weissman SJ, Limaye AP *et al.*** The pandemic H30 subclone of *Escherichia coli* sequence type 131 is associated with persistent infections and adverse outcomes independent from its multidrug resistance and associations with compromised hosts. *Clin Infect Dis* 2016;62:1529–1536.
7. **Pitout JD.** Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. *Front Microbiol* 2012;3:9.
8. **Wirth T, Falush D, Lan R, Colles F, Mensa P *et al.*** Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 2006;60:1136–1151.
9. **Phan MD, Peters KM, Sarkar S, Lukowski SW, Allsopp LP *et al.*** The serum resistome of a globally disseminated multidrug resistant uropathogenic *Escherichia coli* clone. *PLoS Genet* 2013;9:e1003834.
10. **de Kraker ME, Jarlier V, Monen JC, Heuer OE, van de Sande N *et al.*** The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. *Clin Microbiol Infect* 2013;19:860–868.
11. **Alhashash F, Wang X, Paszkiewicz K, Diggle M, Zong Z *et al.*** Increase in bacteraemia cases in the East Midlands region of the UK due to MDR *Escherichia coli* ST73: high levels of genomic and plasmid diversity in causative isolates. *J Antimicrob Chemother* 2016;71:339–343.
12. **de Souza da-Silva AP, de Sousa VS, Martins N, da Silva Dias RC, Bonelli RR *et al.*** *Escherichia coli* sequence type 73 as a cause of community acquired urinary tract infection in men and women in Rio de Janeiro, Brazil. *Diagn Microbiol Infect Dis* 2017;88:69–74.
13. **Miajlovic H, Aogáin MM, Collins CJ, Rogers TR, Smith SG.** Characterization of *Escherichia coli* bloodstream isolates associated with mortality. *J Med Microbiol* 2016;65:71–79.
14. **Yahiaoui M, Robin F, Bakour R, Hamidi M, Bonnet R *et al.*** Antibiotic resistance, virulence, and genetic background of community-acquired uropathogenic *Escherichia coli* from Algeria. *Microb Drug Resist* 2015;21:516–526.
15. **Clermont O, Gordon D, Denamur E.** Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology* 2015;161:980–988.
16. **Riley LW.** Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin Microbiol Infect* 2014;20:380–390.
17. **Bengtsson S, Naseer U, Sundsfjord A, Kahlmeter G, Sundqvist M.** Sequence types and plasmid carriage of uropathogenic *Escherichia coli* devoid of phenotypically detectable resistance. *J Antimicrob Chemother* 2012;67:69–73.
18. **Kahlmeter G, Åhman J, Matuschek E.** Antimicrobial resistance of *Escherichia coli* causing uncomplicated urinary tract infections: a European update for 2014 and comparison with 2000 and 2008. *Infect Dis Ther* 2015;4:417–423.
19. **Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM *et al.*** Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* 2015;9:1269–1279.
20. **Bell SM, Pham JN, Rafferty DL, Allerton JK.** *Antibiotic Susceptibility Testing by the CDS Method: A Manual for Medical and Veterinary Laboratories*, 8th edn, vol. 2016. Kogarah: South Eastern Area Laboratory Services; 2016.
21. **Coil D, Jospin G, Darling AE.** A5-MiSeq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015;31:587–589.
22. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al.*** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
23. **Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ *et al.*** SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014;6:90.
24. **Chen L, Yang J, Yu J, Yao Z, Sun L *et al.*** VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 2005;33:D325–D328.
25. **Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M *et al.*** ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2014;58:212–220.

26. Minogue TD, Daligault HA, Davenport KW, Bishop-Lilly KA, Broomall SM *et al.* Complete genome assembly of *Escherichia coli* ATCC 25922, a serotype O6 reference strain. *Genome Announc* 2014;2:e00969-14.
27. Reeves PR, Liu B, Zhou Z, Li D, Guo D *et al.* Rates of mutation and host transmission for an *Escherichia coli* clone over 3 years. *PLoS One* 2011;6:e26907.
28. Reister M, Hoffmeier K, Krezdorn N, Rotter B, Liang C *et al.* Complete genome sequence of the gram-negative probiotic *Escherichia coli* strain Nissle 1917. *J Biotechnol* 2014;187:106–107.
29. Zdziarski J, Brzuszkiewicz E, Wullt B, Liesegang H, Biran D *et al.* Host imprints on bacterial genomes-rapid, divergent evolution in individual patients. *PLoS Pathog* 2010;6:e1001078.
30. Welch RA, Burland V, Plunkett G, Redford P, Roesch P *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* 2002;99:17020–17024.
31. Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT *et al.* Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. *Clin Infect Dis* 2015;61:305–312.
32. Dunitz MI, Coil DA, Jospin G, Eisen JA, Adams JY. Draft genome sequences of *Escherichia coli* strains isolated from septic patients. *Genome Announc* 2014;2:e01278-14.
33. Earle SG, Wu CH, Charlesworth J, Stoesser N, Gordon NC *et al.* Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat Microbiol* 2016;1:16041.
34. Kluytmans-van den Bergh MF, Rossen JW, Bruijning-Verhagen PC, Bonten MJ, Friedrich AW *et al.* Whole-genome multilocus sequence typing of extended-spectrum-beta-lactamase-producing enterobacteriaceae. *J Clin Microbiol* 2016;54:2919–2927.
35. Pesesky MW, Hussain T, Wallace M, Wang B, Andleeb S *et al.* KPC and NDM-1 genes in related *Enterobacteriaceae* strains and plasmids from Pakistan and the United States. *Emerg Infect Dis* 2015; 21:1034–1037.
36. Roach DJ, Burton JN, Lee C, Stackhouse B, Butler-Wu SM *et al.* A year of infection in the intensive care unit: prospective whole genome sequencing of bacterial clinical isolates reveals cryptic transmissions and novel microbiota. *PLoS Genet* 2015;11: e1005413.
37. Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B *et al.* Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Res* 2015;25:119–128.
38. Toro M, Cao G, Rump L, Nagaraja TG, Meng J *et al.* Genome sequences of 64 non-O157:H7 Shiga toxin-producing *Escherichia coli* strains. *Genome Announc* 2015;3:e01067-15.
39. Barton BM, Harding GP, Zuccarelli AJ. A general method for detecting and sizing large plasmids. *Anal Biochem* 1995;226:235–240.
40. Partridge SR, Zong Z, Iredell JR. Recombination in IS26 and Tn2 in the evolution of multiresistance regions carrying blaCTX-M-15 on conjugative IncF plasmids from *Escherichia coli*. *Antimicrob Agents Chemother* 2011;55:4971–4978.
41. Koeleman JG, Stoof J, van der Bijl MW, Vandenbroucke-Grauls CM, Savelkoul PH. Identification of epidemic strains of *Acinetobacter baumannii* by integrase gene PCR. *J Clin Microbiol* 2001;39: 8–13.
42. Darling AE, Jospin G, Lowe E, Matsen FA, Bik HM *et al.* PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2014; 2:e243.
43. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* 2011;331:430–434.
44. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30: 1312–1313.
45. Nesta B, Spraggon G, Alteri C, Moriel DG, Rosini R *et al.* FdeC, a novel broadly conserved *Escherichia coli* adhesin eliciting protection against urinary tract infections. *MBio* 2012;3:e00010-12.
46. Ladomersky E, Petris MJ. Copper tolerance and virulence in bacteria. *Metallomics* 2015;7:957–964.
47. Baker-Austin C, Wright MS, Stepanauskas R, Mcarthur JV. Co-selection of antibiotic and metal resistance. *Trends Microbiol* 2006; 14:176–182.
48. Nakaya R, Nakamura A, Murata Y. Resistance transfer agents in Shigella. *Biochem Biophys Res Commun* 1960;3:654–659.
49. García V, García P, Rodríguez I, Rodicio R, Rodicio MR. The role of IS26 in evolution of a derivative of the virulence plasmid of *Salmonella enterica* serovar Enteritidis which confers multiple drug resistance. *Infect Genet Evol* 2016;45:246–249.
50. Broom JE, Hill DF, Hughes G, Jones WA, McNaughton JC *et al.* Sequence of a transposon identified as Tn1000 (gamma delta). *DNA Seq* 1995;5:185–189.
51. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M *et al.* Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci USA* 2014;111:5694–5699.
52. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B *et al.* The epidemic of extended-spectrum-β-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio* 2013;4:e00377-13.
53. Dawes FE, Kuzevski A, Bettelheim KA, Hornitzky MA, Djordjevic SP *et al.* Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in *Escherichia coli* of human and animal origin. *PLoS One* 2010;5:e12754.
54. Roy Chowdhury P, Charles IG, Djordjevic SP. A role for Tn6029 in the evolution of the complex antibiotic resistance gene loci in genomic island 3 in enteroaggregative hemorrhagic *Escherichia coli* O104:H4. *PLoS One* 2015;10:e0115781.
55. Mckinnon J, Roy Chowdhury P, Djordjevic SP. Genomic analysis of multidrug-resistant *Escherichia coli* ST58 causing urosepsis. *Int J Antimicrob Agents* 2018;52:430–435.
56. Rodríguez I, Rodicio MR, Herrera-León S, Echeita A, Mendoza MC. Class 1 integrons in multidrug-resistant non-typhoidal *Salmonella enterica* isolated in Spain between 2002 and 2004. *Int J Antimicrob Agents* 2008;32:158–164.
57. Cunha MPV, Saidenberg AB, Moreno AM, Ferreira AJP, Vieira MAM *et al.* Pandemic extra-intestinal pathogenic *Escherichia coli* (ExPEC) clonal group O6-B2-ST73 as a cause of avian colibacillosis in Brazil. *PLoS One* 2017;12:e0178970.
58. Liu X, Thungrat K, Boothe DM. Multilocus sequence typing and virulence profiles in uropathogenic *Escherichia coli* isolated from cats in the United States. *PLoS One* 2015;10:e0143335.
59. Yun KW, Kim HY, Park HK, Kim W, Lim IS. Virulence factors of uropathogenic *Escherichia coli* of urinary tract infections and asymptomatic bacteriuria in children. *J Microbiol Immunol Infect* 2014;47:455–461.