Gut Pathogens

GENOME ANNOUNCEMENT

Open Access

Genome sequencing and analysis of *Salmonella enterica* serovar Typhi strain CR0063 representing a carrier individual during an outbreak of typhoid fever in Kelantan, Malaysia

Ramani Baddam[1], Narender Kumar[1], Sabiha Shaik[1], Tiruvayipati Suma[1,2], Soo Tein Ngoi[2,3], Kwai-Lin Thong[2,3] and Niyaz Ahmed[1,2*]

Abstract

Salmonella Typhi is a human restricted pathogen with a significant number of individuals as asymptomatic carriers of the bacterium. *Salmonella* infection can be effectively controlled if a reliable method for identification of these carriers is developed. In this context, the availability of whole genomes of carrier strains through high- throughput sequencing and further downstream analysis by comparative genomics approaches is very promising. Herein we describe the genome sequence of a *Salmonella* Typhi isolate representing an asymptomatic carrier individual during a prolonged outbreak of typhoid fever in Kelantan, Malaysia. Putative genomic coordinates relevant in pathogenesis and persistence of this carrier strain are identified and discussed.

Background

Salmonella enterica serovar Typhi, the aetiologic agent of typhoid fever is still posing a major health problem for the developing world, as about 16 million new cases are reported each year [1]. *S.* Typhi causes systemic infections (typhoid fever) as well as chronic infections (asymptomatic carriers) in humans, the latter serve as the source of infection [2]. The transmission of *S.* Typhi is primarily through faecal-oral route and a significant number of infected individuals become chronic asymptomatic carriers and keep shedding *S.* Typhi in faeces for decades [3]. This results in endemicity of *S.* Typhi in regions of the world with underdeveloped sanitation and community hygiene [4].

Carrier identification becomes extremely important as some of the ancestral haplotypes were observed in recent isolates suggesting their persistence in these asymptomatic carriers [5]. Traditional methods such as

culturing of bacteria from faecal samples are not fool proof as the carriers shed bacteria intermittently. Serological tests to detect specific antibodies such as anti-H and anti-O are unable to differentiate between carriers and individuals who have recovered from the infection [6]. Especially, in areas endemic for *S.* Typhi, due to high background levels of these antibodies, serological tests cannot be adopted for the identification of a carrier [7]. Thus, there is an urgent need for inexpensive and efficient detection methods for the establishment of carrier state, perhaps based on genomic markers.

The genetic typing tools such as PFGE, AFLP, ribotyping etc. can resolve limited genetic variation occurring within specific sites, and therefore are incapable of differentiating highly clonal strains such as outbreak related strains from the ones not associated with the outbreak (carrier isolates) [8-10]. High-throughput sequencing technologies have already been employed as a high resolution molecular epidemiologic tool to discern microevolution of highly related strains [11].

In this study, we attempted to determine if whole genome sequencing of *S.* Typhi isolated from a carrier individual can provide insights related to persistence and or adaptation mechanisms. We describe the genome

* Correspondence: ahmed.nizi@gmail.com
[1]Pathogen Biology Laboratory, Department of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad, Hyderabad, India
[2]Institute of Biological sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia
Full list of author information is available at the end of the article

sequence of a *Salmonella enterica* serovar Typhi strain (ST CR0063) isolated from a carrier individual during a prolonged outbreak of typhoid fever in Kelantan, Malaysia.

Results and discussion

Genome statistics

The size of the draft genome of *Salmonella* Typhi (ST CR0063) is 4,585,851 bp with a coding percentage of 86.1%. The G + C content of this strain is about 51.71%. The total number of CDS determined are 4946 with an average length of gene about 798 nucleotides. The genome of ST CR0063 revealed 77 tRNA and 22 rRNA genes. The subsystems distribution of basic metabolic machinery of this strain is represented in Figure 1. The assembled draft genome shows high degree of similarity and shared core genome regions with *Salmonella* Typhi ST BL196 [12], the one identified as associated with a typhoid outbreak in Kelantan during the same period (Figure 2).
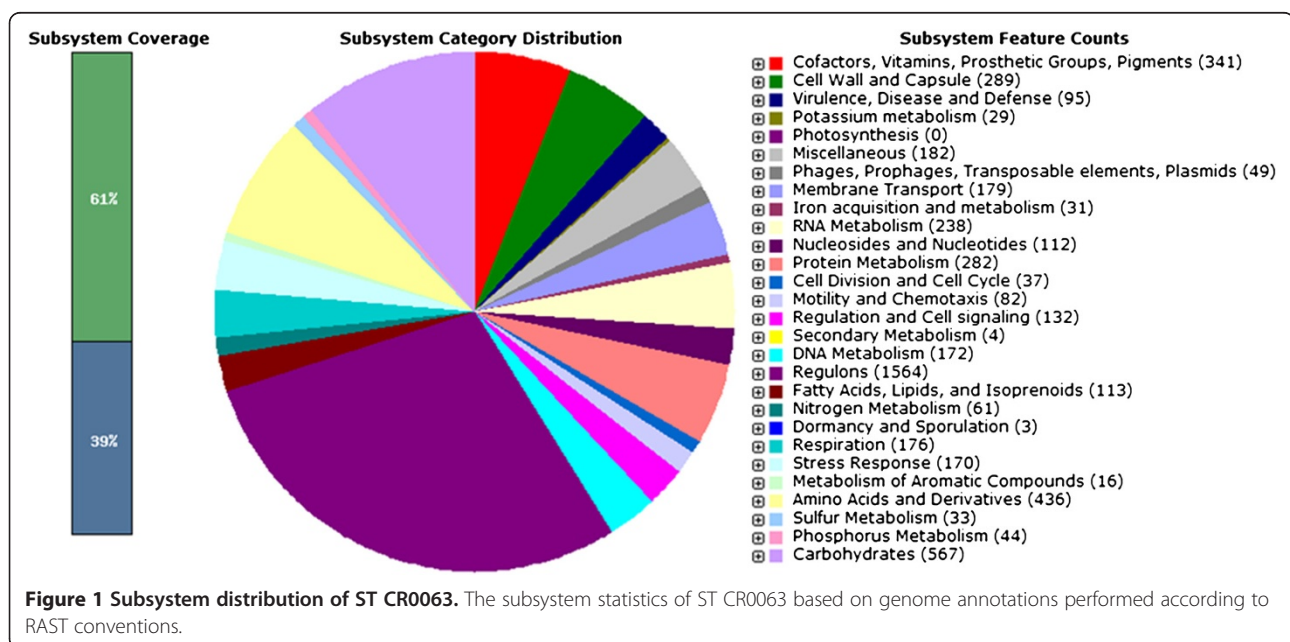
Virulence factors

The gene *shdA*, a key factor predicted to be involved in persistence of the bacterium in the intestines [14] by binding to its extracellular matrix, was identified and annotated. This gene, by mimicking the host heparin, is able to bind to the extracellular matrix proteins, fibronectin and collagen, and probably plays an important role in carriers by contributing to prolonged faecal shedding [15]. The *fim* gene cluster [16] of chaperone –usher family involved in adhesion to non-phagocytic cells was detected along with its negative regulator fimW. Type IV pili and *agf* operon [17,18] encoding curli fimbriae

which aid in attachment of the bacterium to intestinal villi and also with each other, were found in the genome. These adherence factors determine the sites of bacterial colonisation and thereby adaptation and pathogenicity of a particular strain [19,20].
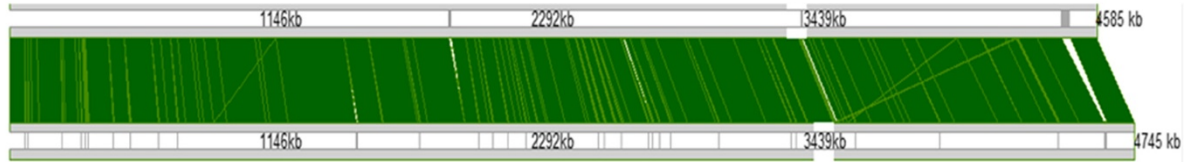
The *S.* Typhi strain ST CR0063 genome also revealed *viaA* and *viaB* loci, the prime regulators of Vi antigen expression. The *viaB* locus contains all genes for the biosynthesis (*tviA-E*) and export (*vexA-E*) of the Vi antigen, a well-known virulence factor [21,22]. The *mgtC* gene involved in Magnesium uptake and ferric uptake regulators (fur) [23] were also identified in ST CR0063. The PhoPQ regulon [24], which induces cytokine secretion and cationic antimicrobial peptide resistance, was also found to be conserved in our carrier strain. The RpoS sigma factor needed to cope up with external stress and nutrient depletion conditions [25] was also identified and annotated. The co-ordinates of these virulence factors in the genome of ST CR0063 are depicted in Figure 3.

Phages and pathogenicity islands (PAIs)

The phages gifsy-1 and fels-2 [27] together with many phage proteins and a few hypothetical proteins were identified in the genome of ST CR0063 by various algorithms (See Methods for details). It is expected that these phages are acquired by horizontal gene transfer (HGT) events as they were embedded in some of the genomic islands recognized. The phage encoding SopE effector protein of SPI-1 (Salmonella Pathogenicity Island) was present in ST CR0063 as recognized in other Typhi genomes [28,29].

Figure 1 Subsystem distribution of ST CR0063. The subsystem statistics of ST CR0063 based on genome annotations performed according to RAST conventions.

Figure 2 Comparison of *Salmonella* Typhi strains ST CR0063 and ST BL196. Comparison of whole genome sequences of *S. Typhi* strains using MG-CAT – one strain was isolated from a carrier individual (ST CR0063) and another from an infected individual (ST BL196) during a prolonged outbreak of Typhoid fever in Kelantan [13].

More than 15 PAIs that encode clusters of virulence associated genes have been identified across various serovars of *Salmonella enterica*. Ten pathogenicity islands have been identified by us in ST CR0063 and as expected [30], they were characterised by different G + C content and bounded by t-RNA genes. The SPI-1 type III secretion system (TTSS) structural genes spaM-NOPQRS and invABCEFGH and their regulatory proteins HilA, HilC, HilD [31] were also identified and annotated. The SPI-1 secreted effector proteins SopE,
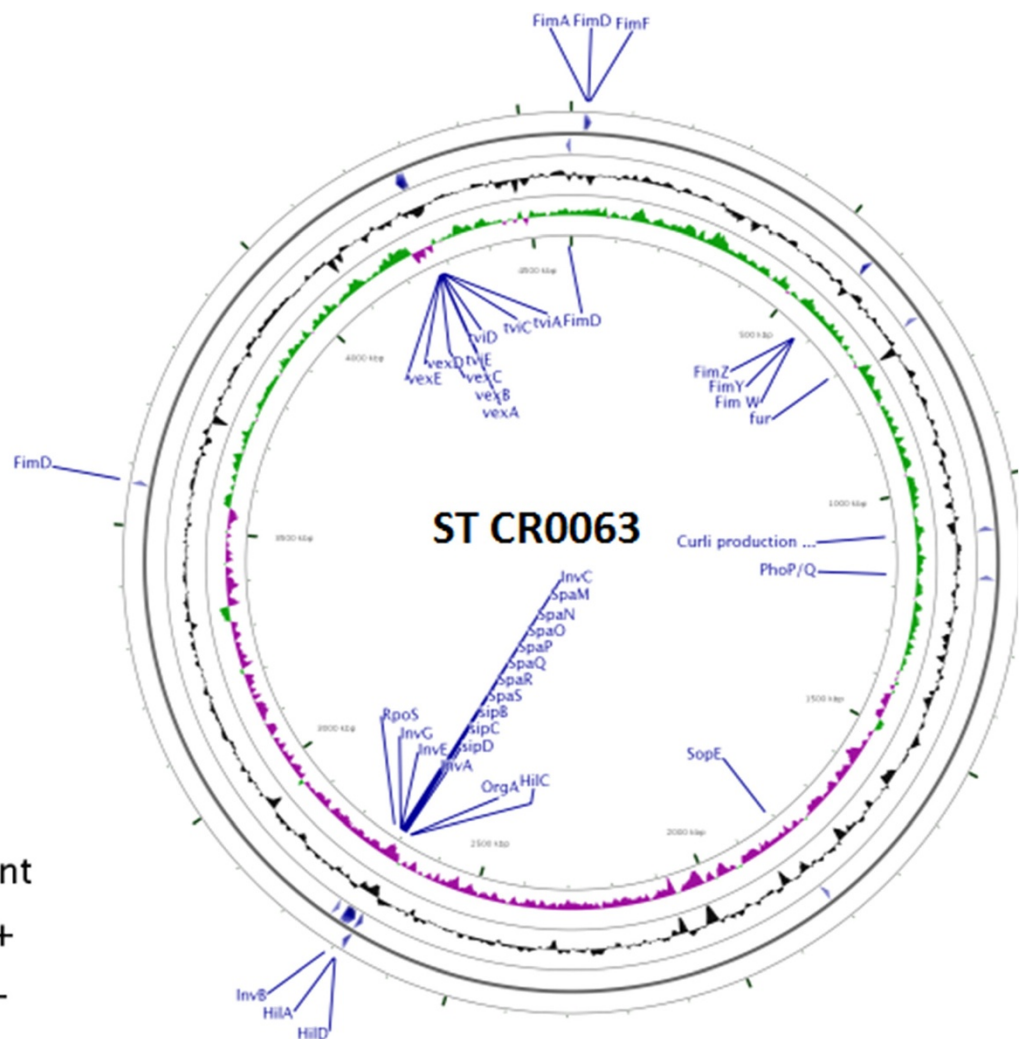

Figure 3 Circular Genome view of ST CR063. Positions of some of the major virulence factors and their regulators identified in ST CR0063 marked in the circular genome generated using CGview [26].

SopE2, SipA, SipB, SipC and SptP required for endothelial uptake and invasion [32] are also present. The genes SpiC, SseF, SseG, SifA, SifB secreted by SPI-2 TTSS and that are needed for survival in macrophages and colonisation of host organs [33] were also recognised in the present genome. The known regulators of SPI-2, OmpR-EnvZ and PhoP-PhoQ [34] were present. SPI-3, identified by us, contained magnesium transport genes *mgtC* and *marT* which are required for survival in macrophages [35]. Type I secretion system and its associated proteins encoded by SPI-4, and that are involved in the invasion of the intestinal epithelium [36], were also located in the present genome. The SPI-1 effector proteins SopB and PipB associated with enteritis and coded by SPI-5 [37] were also detected and annotated. The chaperone-usher fimbrial operons carried by SPI-6, SPI-10 and bacteriocin immunity proteins carried by SPI-8 [38] were identified. The SPI-7 and SPI-9 were identified in the ST CR0063 genome and were found to encode *viaB* locus, type IV pili formation proteins and TISS [38,39].

Conclusions and prospective

The genomic blueprint of *Salmonella* Typhi isolate ST CR0063 was elucidated in this study. The genome sequence information presented herein may be harnessed to guide comparative genomics and identification of novel and specific diagnostic markers. However, further studies involving large scale genome sequencing of the strains from several of the endemic countries and especially those from carrier individuals of different socio-economical settings is needed to develop a reliable approach to decipher the characteristics of a carrier state. Also, it will be required to determine the true extent of the diversity of carrier strains as juxtaposed to their acutely pathogenic forms in terms of 1) gene gain/loss during colonization and adaptation; 2) dynamics of virulence acquisition/attenuation; 3) possible genomic rearrangements; and 4) the relative preponderance of carrier and virulent strains circulating in different endemic regions of the world. Finally, an in-depth analysis of the host-pathogen interactions and their influence on gut microbiota can only explain the adaptation and persistence mechanisms of the (asymptomatic) carrier strains.

Methods
Genome sequencing
DNA was isolated from the stool sample of an asymptomatic carrier individual from Kelantan, Malaysia in 2007 during a prolonged outbreak. The draft genome sequence of this strain (STCR0063) was determined by Illumina Genome Analyzer (GAIIx, pipe- line ver l.6). The 100 bp paired-end sequencing was done with an insert size of 300 bp. About 67X genome coverage was achieved and 1.9 gigabytes of data were obtained.

Assembly and annotation
The sequence data were assembled *de-novo* in the same way as described previously [40-45] into 538 contigs using Velvet [46] at optimal hash length 39. SSPACE [47] was used for scaffolding the pre-assembled contigs using paired-end data. The gaps within these scaffolds were filled using Gapfiller by aligning the reads against already generated Scaffolds by SSPACE [48].

A reference guided assembly was generated by aligning reads to *Salmonella* Typhi str. CT18 [GenBank: AL513382.1] using bwa tools [49]. This reference guided assembly was used to re-order the scaffolds generated in de-novo way. In-house written Perl scripts were used for this re-ordering process and to finalize the gaps. The de novo and reference guided approaches were used to finalize the consensus draft genome. The reference guided assembly and reordered scaffolds were loaded on to Tablet – NGS data visualisation tool, to visualise the repeats, insertions and deletions [50].

The final draft nucleotide sequence after manual curation was annotated in our laboratory using RAST [51] and ISGA pipeline [52]. The genome statistics were gleaned using Artemis [53]. The data were further validated using gene prediction tools such as Glimmer [54] and EasyGene [55]. The RNAmmer [56] and tRNAscan-SE [57] were used to identify rRNA and tRNA respectively.

Phages and PAIs
Prophages and putative phage like elements in the genome were identified using PhiSpy [58] and Prophage Finder [59]. The putative HGT events were determined using Alien Hunter tool [60]. An integrated interface Island Viewer was used to predict putative genomic islands within the genome [61].

Sequence data access
The Salmonella enterica subsp. enterica serovar Typhi str. CR0063 whole genome shotgun (WGS) project has been submitted to the GenBank and has the project accession AKIC00000000. The project version entailing draft assembly described herein has the accession number AKIC01000000, and consists of sequences AKIC01000001-AKIC01000538.

Authors' contributions
NA designed the study, interpreted the results and edited the manuscript. RB and NK managed Illumina sequencing, made the assemblies, analyzed the genome, and performed annotations. SS and TS provided computational tools and contributed to automation of the analysis process. KT provided

inputs related to the outbreak and the strain features, characterized the strain and maintained it in pure cultures. STN contributed to microbiology of the strain and prepared high molecular weight DNA for genome sequencing. All the authors read and approved the manuscript prior to submission.

Author details
[1]Pathogen Biology Laboratory, Department of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad, Hyderabad, India. [2]Institute of Biological sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia. [3]Laboratory of Biomedical Science and Molecular Microbiology, UMBIO Research Cluster, University of Malaya, Kuala Lumpur, Malaysia.

References
1. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ: Typhoid fever. *N Engl J Med* 2002, **347**:1770–1782.
2. Boyle EC, Bishop JL, Grassl GA, Finlay BB: Salmonella: from pathogenesis to therapeutics. *J Bacteriol* 2007, **189**:1489–1495.
3. Gonzalez-Escobedo G, Marshall JM, Gunn JS: Chronic and acute infection of the gall bladder by Salmonella Typhi: understanding the carrier state. *Nat Rev Microbiol* 2011, **9**:9–14.
4. Gopinath S, Carden S, Monack D: Shedding light on Salmonella carriers. *Trends Microbiol* 2012, **20**:320–327.
5. Roumagnac P, Weill FX, Dolecek C, Baker S, Brisse S, Chinh NT, Le TA, Acosta CJ, Farrar J, Dougan G, Achtman M: Evolutionary history of Salmonella Typhi. *Science* 2006, **314**:1301–1304.
6. Olopoenia LA, King AL: Widal agglutination test – 100 years later: still plagued by controversy. *Postgrad Med J* 2000, **76**:80–84.
7. Gupta A, My Thanh NT, Olsen SJ, Sivapalasingam S, My Trinh TT, Phuong Lan NT, Hoekstra RM, Bibb W, Minh NT, Danh TP, Cam PD, Mintz ED: Evaluation of community-based serologic screening for identification of chronic Salmonella Typhi carriers in Vietnam. *Int J Infect* 2006, **10**:309–314.
8. Nair S, Schreiber E, Thong KL, Pang T, Altwegg M: Genotypic characterization of Salmonella typhi by amplified fragment length polymorphism fingerprinting provides increased discrimination as compared to pulsed-field gel electrophoresis and ribotyping. *J Microbiol Methods* 2000, **41**:35–43.
9. Thong KL, Puthucheary S, Yassin RM, Sudarmono P, Padmidewi M, Soewandojo E, Handojo I, Sarasombath S, Pang T: Analysis of *Salmonella* typhi isolates from southeast Asia by pulsed-field gel electrophoresis. *J Clin Micro boil* 1995, **33**:1938–1941.
10. Baddam R, Thong KL, Avasthi TS, Shaik S, Yap KP, Teh CS, Chai LC, Kumar N, Ahmed N: Whole-genome sequences and comparative genomics of *Salmonella enterica* serovar Typhi isolates from patients with fatal and nonfatal typhoid fever in Papua New Guinea. *J Bacteriol* 2012, **194**:5122–5123.
11. Aziz RK, Nizet V: Pathogen microevolution in high resolution. *Sci Transl Med* 2010, **2**(16):16–4.
12. Baddam R, Kumar N, Thong KL, Ngoi ST, Teh CS, Yap KP, Chai LC, Avasthi TS, Ahmed N: Genetic fine structure of a Salmonella enterica serovar Typhi strain associated with the 2005 outbreak of typhoid fever in Kelantan, Malaysia. *J Bacteriol* 2012, **194**:3565–3566.
13. Treangen T, Messeguer X: M-GCAT: Interactively and efficiently constructing large-scale multiple genome comparison frameworks in closely related species. *BMC Bioinformatics* 2006, **7**:433.
14. Kingsley RA, Santos RL, Keestra AM, Adams LG, Bäumler AJ: Salmonella enterica serotype Typhimurium ShdA is an outer membrane fibronectin-binding protein that is expressed in the intestine. *Mol Microbiol* 2002, **43**:895–905.
15. Kingsley RA, van Amsterdam K, Kramer N, Bäumler AJ: The *shdA* gene is restricted to serotypes of Salmonella enterica subspecies I and contributes to efficient and prolonged fecal shedding. *Infect Immun* 2000, **68**:2720–2727.
16. Muscas P, Rossolini GM, Chiesurin A, Santucci A, Satta G: Purification and characterization of type 1 fimbriae of Salmonella typhi. *Microbiol Immunol* 1994, **38**:353–358.
17. Craig L, Pique ME, Tainer JA: Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* 2004, **2**:363–378.
18. Collinson SK, Clouthier SC, Doran JL, Banser PA, Kay WW: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae. *J Bacteriol* 1996, **178**:662–667.
19. Duncan MJ, Mann EL, Cohen MS, Ofek I, Sharon N, Abraham SN: The distinct binding specificities exhibited by enterobacterial Type 1 fimbriae are determined by their fimbrial shafts. *J Biol Chem* 2005, **280**:37707–37716.
20. Guo A, Cao S, Tu L, Chen P, Zhang C, Jia A, Yang W, Liu Z, Chen H, Schifferli DM: FimH alleles direct preferential binding of Salmonella to distinct mammalian cells or to avian cells. *Microbiology* 2009, **155**:1623–1633.
21. Virlogeux I, Waxin H, Ecobichon C, Popoff MY: Role of the viaB locus in synthesis, transport and expression of Salmonella typhi Vi antigen. *Microbiology* 1995, **141**:3039–3047.
22. Robbins JD, Robbins JB: Re examination of the protective role of the capsular polysaccharide (Vi antigen) of Salmonella typhi. *J Infect Dis* 1984, **150**:436–449.
23. Moncrief MB, Maguire ME: Magnesium transport in prokaryotes. *J Biol Inorg Chem* 1999, **4**:523–527.
24. Guo L, Lim KB, Gunn JS, Bainbridge B, Darveau RP, Hackett M, Miller SI: Regulation of lipid A modifications by Salmonella typhimurium virulence genes phoP-phoQ. *Science* 1997, **276**:250–253.
25. Chen CY, Eckmann L, Libby SJ, Fang FC, Okamoto S, Kagnoff MF, Fierer J, Guiney DG: Expression of Salmonella typhimurium rpoS and rpoS-dependent genes in the intracellular environment of eukaryotic cells. *Infect Immun* 1996, **64**:4739–4743.
26. Stothard P, Wishart DS: Circular genome visualization and exploration using CGView. *Bioinformatics* 2005, **21**:537–539.
27. Stanley TL, Ellermeier CD, Slauch JM: Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects Salmonella enterica serovar typhimurium survival in Peyer's patches. *J Bacteriol* 2000, **182**:4406–4413.
28. Mirold S, Rabsch W, Rohde M, Stender S, Tschape H, Russmann H, Igwe E, Hardt WD: Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic Salmonella typhimurium strain. *Proc Natl Acad Sci U S A* 1999, **96**:9845–9850.
29. Hardt WD, Urlaub H, Galan JE: A substrate of the centisome 63 type III protein secretion system of Salmonella typhimurium is encoded by acryptic bacteriophage. *Proc Natl Acad Sci U S A* 1998, **95**:2574–2579.
30. Marcus SL, Brumell JH, Pfeifer CG, Finlay BB: Salmonella pathogenicity islands: big virulence in small packages. *Microbes Infect* 2000, **2**:145–156.
31. Lee VT, Schneewind O: Type III secretion machines and the pathogenesis of enteric infections caused by Yersinia and Salmonella spp. *Immunol Rev* 1999, **168**:241–255.
32. McGhie EJ, Brawn LC, Hume PJ, Humphreys D, Koronakis V: Salmonella takes control: effector driven manipulation of the host. *Curr Opin Microbiol* 2009, **12**:117–124.
33. Waterman SR, Holden DW: Functions and effectors of the Salmonella pathogenicity island 2 type III secretion system. *Cell Microbiol* 2003, **5**:501–511.
34. Garmendia J, Beuzón CR, Ruiz-Albert J, Holden DW: The roles of SsrA-SsrB and OmpR-EnvZ in the regulation of genes encoding the Salmonella typhimurium SPI-2 type III secretion system. *Microbiology* 2003, **149**:2385–2396.
35. Blanc-Potard AB, Solomon F, Kayser J, Groisman EA: The SPI-3 Pathogenicity Island of Salmonella enterica. *J Bacteriol* 1999, **181**:998–1004.
36. Gerlach RG, Claudio N, Rohde M, Jackel D, Wagner C, Hensel M: Cooperation of Salmonella pathogenicity islands 1 and 4 is required to breach epithelial barriers. *Cell Microbiol* 2008, **10**:2364–2376.
37. Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, Galyov EE: Identification of a pathogenicity island required for Salmonella enteropathogenicity. *Mol Microbiol* 1998, **29**:883–891.
38. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar TyphiCT18. *Nature* 2001, **413**:848–852.
39. Pickard D, Wain J, Baker S, Line A, Chohan S, Fookes M, Barron A, Gaora PO, Chabalgoity JA, Thanky N, Scholes C, Thomson N, Quail M, Parkhill J,

Dougan G: Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding Salmonella enterica pathogenicity island SPI-7. *J Bacteriol* 2003, **185**:5055–5065.

40. Avasthi TS, Devi SH, Taylor TD, Kumar N, Baddam R, Kondo S, Suzuki Y, Lamouliatte H, Mégraud F, Ahmed N: **Genomes of two chronological isolates (Helicobacter pylori 2017 and 2018) of the West African Helicobacter pylori strain 908 obtained from a single patient.** *J Bacteriol* 2011, **193**:3385–3386.

41. Avasthi TS, Kumar N, Baddam R, Hussain A, Nandanwar N, Jadhav S, Ahmed N: **Genome of multidrug-resistant uropathogenic Escherichia coli strain NA114 from India.** *J Bacteriol* 2011, **193**:4272–4273.

42. Devi SH, Taylor TD, Avasthi TS, Kondo S, Suzuki Y, Megraud F, Ahmed N: **Genome of Helicobacter pylori strain 908.** *J Bacteriol* 2010, **192**:6488–6489.

43. Siddavattam D, Karegoudar TB, Mudde SK, Kumar N, Baddam R, Avasthi TS, Ahmed N: **Genome of a novel isolate of Paracoccus denitrificans capable of degrading N, N-dimethylformamide.** *J Bacteriol* 2011, **193**:5598–5599.

44. Yap KP, Teh CS, Baddam R, Chai LC, Kumar N, Avasthi TS, Ahmed N, Thong KL: **Insights from the genome sequence of a Salmonella enterica serovar Typhi strain associated with a sporadic case of typhoid fever in Malaysia.** *J Bacteriol* 2012, **194**:5124–5125.

45. Yap KP, Gan HM, Teh CS, Baddam R, Chai LC, Kumar N, Tiruvayipati SA, Ahmed N, Thong KL: **Genome sequence and comparative pathogenomics analysis of a Salmonella enterica serovar Typhi strain associated with a typhoid carrier in Malaysia.** *J Bacteriol* 2012, **194**:5970–5971.

46. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**:821–829.

47. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: **Scaffolding pre-assembled contigs using SSPACE.** *Bioinformatics* 2011, **27**:578–579.

48. Nadalin F, Vezzi F, Policriti A: **GapFiller: a de novo assembly approach to fill the gap within paired reads.** *BMC Bioinformatics* 2012, **13**(14):8. doi:10.1186/1471-2105-13-S14-S8.

49. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler Transform.** *Bioinformatics* 2009, **25**:1754–1760.

50. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D: **Tablet–next generation sequence assembly visualization.** *Bioinformatics* 2010, **26**:401–402.

51. Aziz RK, Devoid S, Disz T, Edwards RA, Henry CS, *et al*: **SEED Servers: high-performance access to the SEED genomes, annotations, and metabolic models.** *PLoS ONE* 2012, **7**(10):48053. doi:10.1371/journal.pone.0048053.

52. Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q: **An Ergatis-based prokaryotic genome annotation web server.** *Bioinformatics* 2010, **26**:1122–1124.

53. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16**:944–945.

54. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL: **Improved microbial gene identification with GLIMMER.** *Nucleic Acids Res* 1999, **27**:4636–4641.

55. Larsen TS, Krogh A: **EasyGene—a prokaryotic gene finder that ranks ORFs by statistical significance.** *BMC Bioinformatics* 2003, **4**:21.

56. Lagesen K, Hallin PF, Rødland E, Stærfeldt HH, Ussery DW RT: **RNammer: consistent annotation of rRNA genes in genomic sequences.** *Nucleic Acids Res* 2007, **35**:3100–3108.

57. Schattner P, Brooks AN, Lowe TM: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 2005, **33**:W686–W689.

58. Akhter S, Aziz RK, Edwards RA: **PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies.** *Nucleic Acids Res* 2012, **40**:e126.

59. Bose M, Barber RD: **Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences.** *In Silico Biol* 2006, **6**:223–227.

60. Vernikos GS, Parkhill J: **Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands.** *Bioinformatics* 2006, **22**:2196–2203.

61. Langille MG, Brinkman FS: **IslandViewer: an integrated interface for computational identification and visualization of genomic islands.** *Bioinformatics* 2009, **25**:664–665.