A Genotyping Scheme for *Salmonella Enterica* Serovar Typhi, the Cause of Typhoid Fever

Vanessa K. Wong[1,2] and Elizabeth J. Klemm[1], Stephen Baker[3,4,5], Derek Pickard[1], Andrew J. Page[1], Nicholas A. Feasey[6], Florian Marks[7], Samuel Kariuki[1,8], Stephen Obaro[9,10,11], Gordon Dougan[1] and Kathryn E. Holt[12], International Typhoid Consortium[†]

[1]*The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom;* [2]*Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom;* [3]*The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam;* [4]*Centre for Tropical Medicine & Global Health, Nuffield Department of Clinical Medicine, Oxford University, Oxford, United Kingdom;* [5]*Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom;* [6]*Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, United Kingdom;* [7]*International Vaccine Institute, Department of Epidemiology, Kwanak, Seoul, Republic of Korea;* [8]*Kenya Medical Research Institute (KEMRI), Nairobi, Kenya;* [9]*Division of Pediatric Infectious Diseases, University of Nebraska Medical Center, Omaha, NE;* [10]*University of Abuja Teaching Hospital, Gwagwalada, FCT, Nigeria;* [11]*Bingham University, Karu, Nigeria;* [12]*Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Australia*

† *Full list of members of the International Consortium can be found at* coalitionagainsttyphoid.org/itc

Background: Typhoid caused by *Salmonella enterica subsp. enterica* serovar Typhi (*S.* Typhi) is a global health problem. *S.* Typhi exhibits minimal genetic variation, which makes discrimination of isolates in epidemiological studies challenging. Public health laboratories have utilised methods such as pulsed-field gel electrophoresis, which provide limited discrimination and phylogenetic information. *S.* Typhi classification has been based on genotyping of 88 single nucleotide polymorphisms (SNPs) allowing subdivision of the population into 85 haplotypes. Subsequently, whole genome sequencing (WGS) has been used to identify more SNPs within the *S.* Typhi genome providing greater resolution to the scheme.

Methods: We utilized data from WGS of 1,831 *S.* Typhi isolates sourced from over 60 countries to generate a robust genotyping scheme that provides high-resolution phylogenetic information. We explored the utility of the genotyping framework to predict the geographical source of 99 travel-associated *S.* Typhi in the United Kingdom.

Results: We identified a set of 68 SNPs that can be used genotype *S.* Typhi into the four primary clusters, 16 clades and 49 subclades. For each of these groups, we identified one SNP to be used for genotyping. For the travel-associated isolates, prediction of geographical origin based on the closest strain of known location in the global framework would have yielded the correct region of origin in

all cases, and the correct country of origin in 71% of cases (95% confidence interval of 66%-76%).

Conclusion: Our extended genotyping scheme gives greater discriminatory power and improved phylogenetic information than the existing scheme. Analysis of novel *S*. Typhi isolates to the global population framework is predictive of geographic origin at the regional level and has potential to predict origins to the country of origin level. This approach forms a robust framework for tracking typhoid in the field and will enable enhanced surveillance of this important disease.