1 **Development and Evaluation of a Multiple-Locus Variable-Number Tandem-Repeats**

2 **Analysis Assay for Subtyping *Salmonella* Typhi Strains from sub-Saharan Africa**

3

4 Nomsa P. Tau[1,2], Anthony M. Smith[1,2], John R. Wain[3], Andrew Tarupiwa[4], Kalpy J.

5 Coulibaly[5], Karen H. Keddy[1,2] for GERMS-SA

6

7 Author affiliations:

8 [1]Centre for Enteric Diseases, National Institute for Communicable Diseases, a division of the

9 National Health Laboratory Service, Johannesburg, South Africa

10 [2]Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

11 [3]University of East Anglia, Norwich, United Kingdom

12 [4]National Microbiology Reference Laboratory, Harare, Zimbabwe

13 [5] Pasteur Institute, Ivory Coast

14

15 Address for correspondence: Nomsa Tau, Centre for Enteric Diseases, National Institute for

16 Communicable Diseases, Private Bag X4, Sandringham, 2131, South Africa. Tel: +27-11-

17 5550426, Fax: +27-11-5550433, Email: nomsat@nicd.ac.za

18

19 Institution where work was done: Centre for Enteric Diseases, National Institute for

20 Communicable Diseases, a division of the National Health Laboratory Service, Johannesburg,

21 South Africa

22

23 Key words: Typhoid fever, *Salmonella* Typhi, sub-Saharan Africa, MLVA, VNTR

24

25 **Word count**

26 Abstract: 214

27 Text: 3198

28 Number of references: 31

Abstract

Typhoid fever is a significant public health problem in sub-Saharan Africa (SSA). Molecular epidemiological investigations of the highly clonal *Salmonella enterica* subspecies *enterica* serovar Typhi (*Salmonella* Typhi) are important in outbreak detection and tracking disease transmission. In this study, we developed and evaluated a multiple-locus variable-number tandem-repeats (VNTR) analysis (MLVA) assay for characterization of *Salmonella* Typhi isolates from SSA. The MLVA assay was developed for use on capillary electrophoresis systems by testing a collection of 50 *Salmonella* Typhi isolates which showed diverse pulsed-field gel electrophoresis (PFGE) pulsotypes. Thirteen VNTR loci were evaluated and a MLVA assay consisting of five polymorphic loci was adopted. The MLVA assay characterized the isolates into 47 MLVA types while PFGE analysis of the same isolates revealed only 34 pulsotypes. The MLVA assay showed higher discriminatory ability (Simpson diversity index (*D*) 0.998 (95% confidence interval (CI) 0.995 – 1.000) compared to PFGE *D* 0.984 (95% CI 0.974 – 0.994). The Wallace coefficient suggested that MLVA assay predicted PFGE pulsotypes (67%) better than PFGE analysis predicted MLVA type (9%). The MLVA assay is a simple, rapid and more accessible alternative to molecular techniques such as PFGE and whole-genome sequencing. The MLVA assay described in this study is a highly discriminatory molecular epidemiological tool for the effective epidemiological analysis of *Salmonella* Typhi isolates from SSA.

Introduction

Typhoid fever, a systemic infection caused by *Salmonella enterica* subspecies *enterica* serovar Typhi (*Salmonella* Typhi) remains a major public health problem worldwide (1,2). Recent global estimates indicate that typhoid fever causes approximately 26.9 million illnesses annually (1). This disease is one of the most important causes of morbidity and mortality in Asia, Africa and Latin America where it is endemic. The spread of typhoid fever is aggravated by poor living conditions, substandard hygiene practises and unsafe drinking water (3-8). Although *Salmonella* Typhi has proven to be a significant public health problem in Africa (9), little is known about the continental level molecular epidemiology and strain relatedness. Highly discriminatory molecular sub-typing methods, which are accessible to strategic African laboratories, are essential to elucidate the epidemiology of *Salmonella* Typhi; which, in turn, would allow the implementation of appropriate control strategies in the sub-Saharan African region; for many of these countries, the true burden of typhoid fever is unknown (1,10).

Molecular subtyping techniques have transformed the ability to discriminate bacterial strains (11,12). These methods allow for examination of bacterial strain relatedness at the DNA level and as a result provide a powerful tool for surveillance and outbreak investigations (10-12). The increasing importance of *Salmonella* Typhi as well as the emergence and dissemination of the multidrug resistant *Salmonella* Typhi haplotype H58 across sub-Saharan Africa has made it imperative to develop new molecular subtyping methods that allow for sensitive strain discrimination (13-16). Current methodologies used for characterization of *Salmonella* Typhi isolates include multi-locus sequence-typing (MLST), pulsed-field gel electrophoresis (PFGE) and whole-genome sequencing [WGS] (16-18); none of these are available to the vast majority

of laboratories in the region, and only PFGE and WGS can discriminate within the haplotype H58 clone.

PFGE is widely used for subtyping of *Salmonella* Typhi isolates (10,19). This involves macro-restriction analysis of bacterial chromosomal DNA and discrimination of isolates based on the resulting banding patterns. The major drawbacks of PFGE are that it is time consuming which can delay foodborne outbreak identification, and that subjective interpretation of PFGE patterns still remains even with the availability of standardized PulseNet protocols (10). Furthermore, the discriminatory capacity of this methodology in strain typing *Salmonella* Typhi isolates is not absolute (10,20). *Salmonella* Typhi isolates are prone to chromosomal rearrangements. This involves the movement or reshuffling of DNA throughout the genome from one location to another. The rearrangement of the chromosomal gene order may result in altered of PFGE patterns (20-22). The chromosomal rearrangement that occurs within *Salmonella* Typhi isolates has resulted in multiple PFGE patterns being identified from a single outbreak strain (20).

In recent years, several approaches have been made in an effort to improve molecular sub-typing of *Salmonella* Typhi using MLVA (17,23-25). The availability of two fully sequenced *Salmonella* Typhi genomes (CT18 and Ty2) instigated the determination of VNTRs in the *Salmonella* Typhi genome, from which MLVA assays have been analysed. Even so, the homogenous nature of *Salmonella* Typhi has made the development of a highly-discriminatory MLVA typing assay a cumbersome process (17,23-25), hence till today, a standardized set of VNTR loci for the typing of homologous *Salmonella* Typhi strains has not been established.

94

95 In this study, we evaluated thirteen previously reported (17,23-25) VNTR markers for

96 epidemiological investigation of *Salmonella* Typhi strains from sub-Saharan Africa. Our aim

97 was to develop a MLVA assay that targeted five VNTR markers and suitable for use on a

98 capillary electrophoresis system.

99

100 **Methods**

101 <u>**Bacterial isolates:**</u> The Centre for Enteric Diseases (CED) of the National Institute for

102 Communicable Diseases, a division of the National Health Laboratory Services, serves as a

103 reference centre for human enteric pathogens. *Salmonella* Typhi isolates are submitted to the

104 CED from across the country for national surveillance through the Group for Enteric,

105 Respiratory and Meningeal Disease Surveillance network in South Africa (GERMS-SA). A

106 total of 50 *Salmonella* Typhi isolates from the CED database were used to evaluate the MLVA

107 assay. These isolates were selected to give a good representation of diverse collection dates, of

108 diverse PFGE pulsotypes in the CED database and a good representation of geographic areas in

109 South Africa. Isolates from known outbreaks as well as those collected in other African

110 countries, including Zimbabwe and Ivory Coast.

111

112 <u>**Pulsed-field gel electrophoresis**</u>

113 PFGE was performed as part of routine surveillance using the standardized PulseNet protocol

114 for *Salmonella*, *Escherichia* coli and *Shigella sonnei* incorporating XbaI restriction enzyme for

115 genomic digestion (26). PFGE patterns were analysed and compared using the BioNumerics

116 software (Applied Maths, Belgium) with dendrograms of the patterns created using the

117 unweighted pair group method with arithmetic averages (UPGMA), and with analysis of

118 banding patterns incorporating the Dice-coefficient at an optimization setting of 1.5% and a

119 position tolerance setting of 1.5%.

120

121 **Crude DNA preparations from bacteria – to be used as template for PCR**

122 Crude DNA was prepared by inoculating half a loopful of bacterial culture in autoclaved Tris-

123 EDTA buffer (pH 8.0) and incubated for 25 minutes at 95°C. The supernatant containing DNA

124 was used as PCR template**.**

125

126 **MLVA**

127 **Selection of VNTR loci and PCR primers:** Thirteen previously reported polymorphic VNTR

128 were selected and used as markers to explore their potential in determining the strain

129 relatedness of *Salmonella* Typhi isolates from sub-Saharan Africa (17,23-25). MLVA

130 incorporating previously described primers was performed.

131

132 **Screening for length polymorphism of VNTR using simplex PCR:** The evaluation of the

133 presence of allelic variation at each VNTR locus in the 50 *Salmonella* Typhi strain panel began

134 with simplex PCRs. All PCR assays were performed using the Qiagen multiplex PCR Kit.

135 Each 25µl reaction contained 12.5µl of the Qiagen master mix, 2.5µl Qiagen Q-solution, 1 µM

136 each of the forward (labelled with a 6-FAM fluorophore) and reverse primers. For PCR

137 amplification, the initial denaturation at 95°C for 15 min, was followed by 35 cycles of a three

138 step cycle protocol: 94°C for 60 sec, 55°C for 90 sec and 72°C for 90 sec, and a final extension

139   at 72ºC for 10 min. PCR amplicons were diluted 2:70 in sterile distilled water and 2µl aliquots

140   of the dilutions were mixed with 0.2µl GeneScan 600 LIZ® size standard v2.0 (Applied

141   Biosystems) and 11µl Hi-Di formamide solution (Applied Biosystems). The samples were

142   evaluated by capillary electrophoresis on the Applied Biosystems 3500 genetic analyser

143   (Applied Biosystems) and fragments sizes were analysed using the Gene-Mapper Software

144   (Applied Biosystems). DNA fragments were automatically allocated to length bins and alleles

145   were assigned based on the bin fragment sizes. The VNTR allele numbers were entered into the

146   BioNumerics 6.5 software as character values and a dendrogram was constructed using

147   categorical coefficient with a 1.5 tolerance and UPGMA.

148

149   **VNTR measure of diversity:** In order to determine the measure of diversity and the degree of

150   polymorphism at each VNTR locus, the Simpson's index of diversity (DI) and 95% confidence

151   intervals (CIs) were calculated using an online tool available at the Public Health England

152   (PHE) website (http://www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl). The Simpson's

153   diversity index was calculated using previously described methods (27) and the Wallace

154   coefficient was determined in order to assess the congruence between the MLVA assay and

155   PFGE analysis via an online tool (https://www.comparingpartitions.info/).

156

157   **Nucleotide sequencing:** Nucleotide sequencing was performed in order to determine the size

158   of the flanking region at each of the five VNTR loci and to also confirm that the variations in

159   the length of amplicons were a result of variation in copy number. For each VNTR locus, PCR

160   amplicons of ten isolates representing different product sizes served as template DNA in a PCR

161   cycle sequencing reaction using the BigDye Terminator Cycle Sequencing Kit (Applied

162   Biosystems, Foster City, CA, USA) and an Applied Biosystems 3500 genetic analyser. DNA

163   sequences were analysed using DNASTAR Lasergene (version 8.0) software (DNASTAR Inc.,

164   Madison, WI, USA).

165

166   **Multiplex PCR assay:** The five VNTR loci that had the highest degree of polymorphism were

167   selected for the development of the multiplex PCR assay. Each 25µl reaction contained 12.5µl

168   of the Qiagen master mix, 2.5µl Qiagen Q-solution, 1µl crude DNA and primer concentrations

169   as shown on table 2. Forward primers for TR1, TR2, Sal02, Sal20 and TR4699 were labelled

170   with PET, 6-FAM, 6-FAM, VIC and NED respectively. PCR amplicons were diluted as

171   described above and samples were evaluated by capillary electrophoresis on the Applied

172   Biosystems 3500 genetic analyser (Applied Biosystems) and fragments sizes were analysed

173   using the Gene-Mapper Software (Applied Biosystems).

174

175   **Results**

176   **Evaluation of VNTR loci polymorphism**

177   We evaluated a total of 13 VNTR loci using a panel of 50 *Salmonella* Typhi isolates from SSA

178   previously shown by PFGE to be genetically diverse. The repeat sequences at each tandem

179   repeat locus varied from three to 26 bp (Table 1). The primer sets for all thirteen VNTR loci

180   were able to produce PCR amplicons in all 50 isolates at an annealing temperature of 55°C.

181   Capillary electrophoresis was employed for the determination of VNTR PCR amplicons sizes.

182   Only seven of the thirteen VNTR loci showed allelic diversity in the 50 isolates.

183

184   Of the 13 potential VNTR loci evaluated, eight VNTR loci (TR1, TR2, TR3, TR4500, TR4699,

185   Sal02, Sal16 and Sal20) demonstrated the ability to discriminate *Salmonella* Typhi strains.

186   Simpson's diversity indices for these loci ranged from 0.607 to 0.940. Of these, VNTR locus

187   TR3 contained long repeats units (26 bp repeat length) (23) was excluded from the study as it

188   and preference given to short repeats which are more variable (28). For the remaining five, two

189   VNTR loci (TR5 and Sal06) demonstrated decreased ability to discriminate *Salmonella* Typhi

190   strains with lower diversity indices of 0.039 and 0.339 respectively, while the other three

191   VNTR loci (Sal15, Sal10 and TR4) were homogenous and showed no variation amongst the

192   strains (only 1 allele present) with diversity index values of 0.00 (Table 1).

193

MLVA validation by nucleotide sequencing

195   As the diversity at each VNTR locus is a function of both the number of alleles as well as their

196   distribution frequency within a population, high diversity indices indicated more variable

197   VNTR loci. The seven most variable VNTR loci with high Simpson's diversity indices were

198   selected for nucleotide sequencing. VNTR loci with 100% conserved flanking sequences were

199   selected for inclusion in our MLVA assay. Nucleotide sequencing of the amplified PCR

200   products showed that polymorphism in five VNTR loci including, TR1, TR2, Sal02, Sal20 and

201   TR4699 was due to variation in the number of tandem repeat units. The flanking sequences in

202   all the five VNTR loci were conserved (Table 2). Nucleotide sequencing revealed that the

203   flanking sequences at VNTR locus Sal16 were not conserved. Aside from the six base pair

204   tandem repeat sequence 'ACCATG' at VNTR locus Sal16, an additional twelve base pair

205   repeat sequence 'ACCACCATCACG' was identified. Therefore, polymorphism at this VNTR

206   locus was due to variation at both the tandem repeat regions. Nucleotide sequencing also

207   revealed that the flanking sequences at VNTR locus TR4500 were not conserved. A seven base

208   'TTGCCAC' insertion sequence was identified in a number of *Salmonella* Typhi isolates that

209   were subjected to nucleotide sequencing. The correct number of repeat units for each strain

210   could not be correctly determined as not all strains harboured the seven base pair sequence.

211   Redesigning primers would have not resolved the problem as the insertion sequence is located

212   adjacent to the tandem repeat region. For this reasons, VNTR locus Sal16 and VNTR locus

213   TR4500 were excluded from the MLVA assay.

214

215   **Optimisation of the MLVA multiplex PCR assay consisting of five VNTR loci**

216   We developed a MLVA assay consisting of five highly polymorphic VNTR loci including

217   VNTR locus TR1, TR2, Sal02, Sal20 and TR4699 (Figure 1). Three of these VNTR loci

218   (TR4699, Sal02 and Sal20) were located in gene regions and the other two VNTR loci (TR1

219   and TR2) intergenic (Table 2). The five polymorphic VNTR loci were pooled into one PCR

220   reaction for the development of the MLVA multiplex PCR. Capillary electrophoresis was used

221   for the analysis of the MLVA assay in order to correctly determine the VNTR loci amplicons

222   sizes. This required labelling of forward VNTR primers with fluorescent dyes in order to be

223   able to differentiate each of the VNTR loci PCR amplicons. The forward primers for VNTR

224   loci TR1, TR2, Sal02, Sal20 and TR4699 were labelled with PET, 6-FAM, 6-FAM, VIC and

225   NED fluorophores respectively.

226

227   **MLVA analysis of *Salmonella* Typhi isolates**

228   We assessed the discriminatory power of the developed MLVA assay using a panel of 50

229   *Salmonella* Typhi isolates from SSA previously characterized PFGE. The *Salmonella* Typhi

230   strain panel used in this study included diverse collection of isolates that were

231   epidemiologically unrelated as well as closely related isolates. PFGE analysis of these isolates

232   showed 34 unique pulsotypes (figure 2). In contrast, the MLVA assay presented in this study

233   identified 47 MLVA profiles (figure 3). This MLVA assay was able to distinguish between

234   epidemiologically related and unrelated isolates, in particular, clusters (figure 3) of isolates

235   from recent outbreaks in South Africa namely, Delmas, Mpumalanga outbreak in 2005 and the

236   Pretoria, Gauteng outbreak in 2010, were identified. Our MLVA assay was able to distinguish

237   isolates from Ivory Coast, South Africa and Zimbabwe

238

239   **Discriminatory power of MLVA assay**

240   The discriminatory power of the MLVA assay was calculated using the Simpson's DI applied

241   to the *Salmonella* Typhi strain panel. The DI does not only depend on the number of alleles

242   present at each locus but it also takes in to consideration the equitability with which the alleles

243   are distributed at each locus (27,29). VNTR loci with a Simpson's DI value closer to 1 are

244   better markers to differentiate the strains for epidemiological purposes. Of the 50 *Salmonella*

245   Typhi isolates that were analysed, 48 MLVA types were identified. For the MLVA assay,

246   Simpson's DI was calculated at 0.998 (95% CI 0.995 – 1.000). This was a high value compared

247   to the Simpson's DI for PFGE analysis of the same isolates which was calculated at 0.984

248   (95% CI 0.974 – 0.994). With non-overlapping CIs, the difference in Simpson's DI for MLVA

249   and PFGE were statistically significant ($P = 0.010$).

250

251   The Wallace coefficient was calculated in order to assess the congruence between the MLVA

252   assay and PFGE analysis. This coefficient indicates the probability that 2 isolates that cluster

253   together by one subtyping method could also be clustered together using another subtyping

254   method (30). The Wallace coefficient between MLVA assay and PFGE pulsotypes was 67%.

255   In contrast, the probability that 2 isolates belonging to the same MLVA type could have the

256   same pulsotypes was 9%. MLVA assay showed more discriminatory power than PFGE

257   analysis.

258

Discussion

Epidemiological investigations are important for the control of the dissemination of typhoid fever. The ability to study the incidence and spread of *Salmonella* Typhi in SSA relies on the selection and use of suitable and rapid molecular methods that are accessible to the many laboratories in this region. *Salmonella* Typhi is highly homogenous; and in order to study the dissemination of this pathogen in the SSA, highly discriminatory molecular methods are required for characterisation of these isolates. PFGE has been widely used for subtyping *Salmonella* Typhi isolates; however, the suboptimal discriminatory power of this molecular method coupled with the high intensive labour involved makes PFGE unsuitable. WGS has become the gold standard for molecular subtyping of human pathogens; however, the cost involved in using this molecular method for surveillance in the developing countries of the SSA region makes WGS unobtainable, at present. MLVA analysis supersedes PFGE with the high discrimination of *Salmonella* Typhi isolates, yet is more cost effective than WGS.

In this study, we evaluated VNTR loci and used them as molecular markers to discriminate amongst *Salmonella* Typhi isolates from SSA. VNTR loci with shorter repeats sequences ($\leq$ 8bp in length) were included for selection, as these show more variation in copy number. In addition, VNTR loci and primers that harboured perfect homogenous repeat sequences, had 100% conserved flanking sequences and had no insertions or deletions in repeats sequences were selected. Some VNTR loci were not selected for inclusion in the MLVA assay due to inconsistencies observed in the flanking sequences. The Simpson DI values calculated at each locus indicated that the selected loci are of highly polymorphic nature and have greater discriminatory power sufficient to differentiate epidemiologically unrelated strains. A multiplex PCR assay containing primers targeting five highly polymorphic VNTR loci was

developed and proved to be rapid and highly discriminatory in characterizing *Salmonella* Typhi isolates from SSA.

Capillary electrophoresis was employed in order to correctly determine the VNTR loci amplicons sizes. Contrary to the previously used gel electrophoresis method, capillary electrophoresis provides accurate determination of VNTR fragment sizes. The use of capillary electrophoresis allows for separation of DNA at a resolution of one base pair; therefore, nucleotide variations at VNTR loci can be detected. Furthermore, due to the fluorescent labels used in the labelling of primers for the various VNTR loci, capillary electrophoresis makes it possible to identify individual VNTR loci.

Discriminatory power is commonly used to assess the ability of a subtyping method to distinguish epidemiologically unrelated strains. The Simpson's DI is the most commonly used mathematical measure of the probability that two epidemiologically unrelated strains will be characterized, correctly, as different. The calculated Simpson's DI indicated that the MLVA assay developed in this study exhibited higher discriminatory power than PFGE (27,29). Epidemiologic concordance was calculated in order to evaluate the utility of the MLVA assay. Epidemiological concordance is defined as the ability of a subtyping method to correctly identify epidemiologically related isolates as genetically related, distinguish between outbreak isolates from sporadic isolates and also distinguish between outbreaks due to an uncommon source (31). MLVA provided higher discrimination of epidemiologically related isolates than PFGE and was able to distinguish outbreak isolates from background sporadic isolates.

305   Dendrogram analysis was performed with VNTR loci allele numbers entered as character

306   values on BioNumerics 6.5 software using categorical coefficient with a 1.5 tolerance and

307   UPGMA to illustrate the clustering of the MLVA profiles. The categorical coefficient provides

308   a measure of the number of locus variants between two MLVA types. The MLVA assay

309   exhibited higher discrimination of the isolates as 47 MLVA types were identified as compared

310   to 34 pulsotypes observed from PFGE analysis.

311

Conclusion

313   Typhoid fever surveillance is of crucial importance in the SSA region were there's a high

314   burden of disease. MLVA has a great ability to differentiate homogenous *Salmonella* Typhi

315   strains from sub-Saharan Africa. Being PCR based, MLVA can be used in many laboratories in

316   the developing countries which may not have access to WGS.

317

318   Our study describes a MLVA assay consisting of 5 VNTR markers for the rapid analysis of

319   *Salmonella* Typhi strains from SSA which can assist epidemiological investigations of strain

320   relatedness and detection of typhoid fever outbreaks in SSA. This MLVA assay represents a

321   high-throughput typing method that is rapid and discriminatory tool that may be a viable

322   alternative to PFGE for subtyping *Salmonella* Typhi isolates from SSA. To the best of our

323   knowledge, capillary electrophoresis equipment is available in four other laboratories in the

324   SSA region to which the MLVA assay can be implemented. These include Kenya, Gambia,

325   Ivory Coast and Uganda. The implementation of MLVA analysis in these laboratories could be

326   helpful in monitoring the spread of typhoid fever across the continent and highlight the role of

327   specific clones in disease causation. The MLVA assay could also assist in identifying the role

328 of imported cases of typhoid fever into South Africa and their contribution to the burden of

329 disease. This will promote effective and appropriate disease intervention strategies, including

330 prevention and treatment.

331

343   Reference List

344

345 1. **Buckle, G. C., C. L. F. Walker, and R. E. Black**. 2012. Typhoid fever and paratyphoid fever:

346   Systematic review to estimate global morbidity and mortality for 2010. Journal of global health

347   **2**.

348 2. **Connor, B. A. and E. Schwartz**. 2005. Typhoid and paratyphoid fever in travellers. The Lancet

349   infectious diseases **5**:623-628.

350 3. **Feasey, N. A., B. N. Archer, R. S. Heyderman, A. Sooka, B. Dennis, M. A. Gordon, and K. H.**

351   **Keddy**. 2010. Typhoid fever and invasive nontyphoid salmonellosis, Malawi and South Africa.

352   Emerging infectious diseases **16**:1448.

353 4. **Kanungo, S., S. Dutta, and D. Sur**. 2008. Epidemiology of typhoid and paratyphoid fever in India.

354   The Journal of Infection in Developing Countries **2**:454-460.

355 5. **Mweu, E. and M. English**. 2008. Typhoid fever in children in Africa. Tropical Medicine &

356   International Health **13**:532-540.

357 6. **Beyene, G., D. Asrat, Y. Mengistu, A. Aseffa, and J. Wain**. 2008. Typhoid fever in Ethiopia. The

358   Journal of Infection in Developing Countries **2**:448-453.

359 7. **Kariuki, S.** 2008. Typhoid fever in sub-Saharan Africa: challenges of diagnosis and management

360   of infections. The Journal of Infection in Developing Countries **2**:443-447.

361 8. **Crump, J. A. and E. D. Mintz**. 2010. Global trends in typhoid and paratyphoid fever. Clinical

362   Infectious Diseases **50**:241-246.

363 9. **Wain, J., R. S. Hendriksen, M. L. Mikoleit, K. H. Keddy, and R. L. Ochiai**. 2014. Typhoid fever.

364   The Lancet.

365   10. **Wattiau, P., C. Boland, and S. Bertrand**. 2011. Methodologies for Salmonella enterica subsp.

366   enterica subtyping: gold standards and alternatives. Applied and environmental microbiology

367   **77**:7877-7885.

368   11. **Sabat, A. J., A. Budimir, D. Nashev, R. Sa-Leao, J. M. van Dijl, F. Laurent, H. Grundmann, A. W.**

369   **Friedrich, and ESCMID Study Group of Epidemiological Markers (ESGEM)**. 2013. Overview of

370   molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill

371   **18**:20380.

372   12. **Baker, S., W. P. Hanage, and K. E. Holt**. 2010. Navigating the future of bacterial molecular

373   epidemiology. Current opinion in microbiology **13**:640-645.

374   13. **Wain, J., R. S. Hendriksen, M. L. Mikoleit, K. H. Keddy, and R. L. Ochiai**. Typhoid fever. The

375   Lancet.

376   14. **Lunguya, O., V. Lejon, M. F. Phoba, S. Bertrand, R. Vanhoof, J. Verhaegen, A. M. Smith, K. H.**

377   **Keddy, J. J. Muyembe-Tamfum, and J. Jacobs**. 2012. Salmonella typhi in the democratic republic

378   of the congo: fluoroquinolone decreased susceptibility on the rise. PLoS neglected tropical

379   diseases **6**:e1921.

380   15. **Kariuki, S., G. Revathi, J. Kiiru, D. M. Mengo, J. Mwituria, J. Muyodi, A. Munyalo, Y. Y. Teo, K. E.**

381   **Holt, and R. A. Kingsley**. 2010. Typhoid in Kenya is associated with a dominant multidrug-

382   resistant Salmonella enterica serovar Typhi haplotype that is also widespread in Southeast Asia.

383   Journal of clinical microbiology **48**:2171-2176.

384   16. **Wong, V. K., S. Baker, D. J. Pickard, J. Parkhill, A. J. Page, N. A. Feasey, R. A. Kingsley, N. R.**

385   **Thomson, J. A. Keane, and F. X. Weill**. 2015. Phylogeographical analysis of the dominant

386   multidrug-resistant H58 clade of Salmonella Typhi identifies inter-and intracontinental

387   transmission events. Nature genetics **47**:632-639.

388   17. **Tien, Y. Y., H. Ushijima, M. Mizuguchi, S. Y. Liang, and C. S. Chiou**. 2012. Use of multilocus

389   variable-number tandem repeat analysis in molecular subtyping of Salmonella enterica serovar

390   Typhi isolates. Journal of medical microbiology **61**:223-232.

391   18. **Kidgell, C., U. Reichard, J. Wain, B. Linz, M. Torpdahl, G. Dougan, and M. Achtman**. 2002.

392   Salmonella typhi, the causative agent of typhoid fever, is approximately 50,000 years old.

393   Infection, Genetics and Evolution **2**:39-45.

394   19. **Thong, K. L., Y. M. Cheong, S. Puthucheary, C. L. Koh, and T. Pang**. 1994. Epidemiologic analysis

395   of sporadic Salmonella typhi isolates and those from outbreaks by pulsed-field gel

396   electrophoresis. Journal of clinical microbiology **32**:1135-1141.

397   20. **Echeita, M. A. and M. A. Usera**. 1998. Chromosomal rearrangements in Salmonella enterica

398   serotype typhi affecting molecular typing in outbreak investigations. Journal of clinical

399   microbiology **36**:2123-2126.

400   21. **Liu, S. L. and K. E. Sanderson**. 1996. Highly plastic chromosomal organization in Salmonella

401   typhi. Proceedings of the National Academy of Sciences **93**:10303-10308.

402   22. **Matthews, T. D., W. Rabsch, and S. Maloy**. 2011. Chromosomal rearrangements in Salmonella

403   enterica serovar Typhi strains isolated from asymptomatic human carriers. MBio **2**:e00060-11.

404   23. **Liu, Y., M. A. Lee, E. E. Ooi, Y. Mavis, A. L. Tan, and H. H. Quek**. 2003. Molecular typing of

405   Salmonella enterica serovar Typhi isolates from various countries in Asia by a multiplex PCR

406   assay on variable-number tandem repeats. Journal of clinical microbiology **41**:4388-4394.

407   24. **Octavia, S. and R. Lan**. 2009. Multiple-locus variable-number tandem-repeat analysis of

408   Salmonella enterica serovar Typhi. Journal of clinical microbiology **47**:2369-2376.

409   25. **Ramisse, V., P. Houssu, E. Hernandez, F. Denoeud, V. r. Hilaire, O. Lisanti, F. Ramisse, J. D.**

410   **Cavallo, and G. Vergnaud**. 2004. Variable number of tandem repeats in Salmonella enterica

411   subsp. enterica for typing purposes. Journal of clinical microbiology **42**:5722-5730.

412   26. **Ribot, E. M., M. A. Fair, R. Gautom, D. N. Cameron, S. B. Hunter, B. Swaminathan, and T. J.**

413   **Barrett**. 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of

414   Escherichia coli O157: H7, Salmonella, and Shigella for PulseNet. Foodbourne Pathogens &

415   Disease **3**:59-67.

416   27. **Hunter, P. R. and M. A. Gaston**. 1988. Numerical index of the discriminatory ability of typing

417   systems: an application of Simpson's index of diversity. Journal of clinical microbiology **26**:2465-

418   2466.

419   28. **Nadon, C. A., E. Trees, L. K. Ng, E. M++ller Nielsen, A. Reimer, N. Maxwell, K. A. Kubota, and P.**

420   **Gerner-Smidt**. 2013. Development and application of MLVA methods as a tool for inter-

421   laboratory surveillance. Euro Surveill **18**:20565.

422   29. **Simpson, E. H.** 1949. Measurement of diversity. Nature.

423   30. **Wallace, D. L.** 1983. A method for comparing two hierarchical clustering: Comment. Journal of

424   the American Statistical Association **78**:569-576.

425   31. **Boxrud, D., K. Pederson-Gulrud, J. Wotton, C. Medus, E. Lyszkowicz, J. Besser, and J. M.**

426   **Bartkus**. 2007. Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-

427   field gel electrophoresis, and phage typing for subtype analysis of Salmonella enterica serotype

428   Enteritidis. Journal of clinical microbiology **45**:536-543.

429

430

431

432 **<u>Figure Legends</u>**

433 Figure 1. Electropherogram showing all five VNTR loci incorporated in a multiplex PCR for

434 MLVA assay of *Salmonella* Typhi isolates

435 Figure 2. Dendrogram analysis of PFGE patterns of *Salmonella* Typhi isolates from sub-

436 Saharan Africa

437 Figure 3. Dendrogram analysis of MLVA profiles of *Salmonella* Typhi isolates from sub-

438 Saharan Africa

439

440

441 **Table 1. Simpson's diversity indices and primers selected for the amplification of 13 polymorphic _Salmonella_ Typhi VNTR**

VNTR Locus	Number of alleles	Diversity Index	Confidence Interval	VNTR Primer sequences	VNTR Primer references
TR2	26	0.940	0.916 - 0.964	Forward: CCCTGTTTTTCGTGCTGATACG Reverse: CAGAGGATATCGCAACAATCGG	(23)
TR4699	22	0.921	0.892 - 0.950	Forward: CGGGCAATTCGAGATAGGTA Reverse: TAGTTTTGGACCTGCCCATT	(24)
Sal02	16	0.916	0.896 - 0.936	Forward: CGATAGACAGCACCAGCAGA Reverse: TCGCCAATACCATGAGTACG	(17)
TR1	11	0.868	0.836 - 0.900	Forward: GCCAACGATCGCTACTTTTT Reverse: CAAGAAGTGCGCATACTACACC	(24,25)
Sal16	10	0.839	0.805 - 0.874	Forward: TGCAGTTAATTTCTGCGATCA Reverse: CCTTCCGGATGTATGTGACC	(17)
Sal20	6	0.730	0.657 - 0.804	Forward: CAGCCGACACAACTTAACGA Reverse: ACTGTACCGTGCGCGTTT	(24)
TR3	6	0.684	0.605 - 0.763	Forward: CGAAGGCGGAAAAAACGTCCTG Reverse: TGCGATTGGTGTCGTTTCTACC	(23)
TR4500	3	0.607	0.530 - 0.684	Forward: CGTTGCTGCTCCGAAAT Reverse: GCGGTGAAGTGGAAAAAG	(24)
Sal06	3	0.339	0.185 - 0.494	Forward: TTGGTCGCGGAACTATAACTG Reverse: CTTCGTCTGATTGCCACTCC	(25)
TR5	2	0.039	0.000 - 0.114	Forward: TGAAAACCGGCTCGTAGCAGTG Reverse: CATACGGTTACTGCGGATTGG	(23)
Sal15	1	0.000	0.000 - 0.132	Forward: GTGACCGGTTGAGTTTGCAT Reverse: GGCAGGTTGTACCAGTTCGT	(25)
Sal10	1	0.000	0.000 - 0.132	Forward: AAGCGACGTTCTTCTGCAAC Reverse: TGGAATATGATGGCATGACG	(25)
TR4	1	0.000	0.000 - 0.132	Forward: AAAAGCCCGTCTAGTCTTGCAG Reverse: ATCCTTCGGTATCGGGGTATCC	(23)

442

443 **Table 2. Features of seven highly polymorphic _Salmonella_ Typhi VNTR loci**

VNTR Locus	Gene	Product	Repeat sequence	Unit length	Forward offset (bp)	Reverse offset (bp)	MLVA multiplex PCR primer concentrations	Reference(s)
TR1	-	Intergenic region between yedD and yedE	AGAAGAA	7	39	116	1.2 μM	(23,24)
TR2	-	Intergenic region between arcD and yffB	CCAGTTCC	8	191	105	1.2 μM	(23,24)
TR4699	sefC	Outer membrane fimbral usher protein	TGTTGG	6	38	137	0.8 μM	(24)
Sal02	citT	Citrate carrier	TACCAG	6	136	59	1 μM	(24,25)
Sal16	-	Intergenic region between STY3169 (pseudogene) and STY3172	ACCATG	6	90	91, 97, 103, 109, 115, 127	-	(24)
Sal20	ftsN	Cell division protein	CAG	3	83	80	0.5μM	(24,25)
TR4500	STY4635	Hypothetical protein	GGACTC	6	76	195, 202	-	(24)

444

445