Whole-Genome Sequencing of the *Mycobacterium tuberculosis* Manila Sublineage Results in Less Clustering and Better Resolution than Mycobacterial Interspersed Repetitive-Unit–Variable-Number Tandem-Repeat (MIRU-VNTR) Typing and Spoligotyping

F. B. Jamieson,^{a,b} S. Teatero,^a J. L. Guthrie,^a A. Neemuchwala,^a N. Fittipaldi,^{a,b} C. Mehaffy^{a,b*}

Public Health Ontario, Toronto, Ontario, Canada^a; Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada^b

***Mycobacterium tuberculosis* isolates of the Manila sublineage are genetically homogeneous. In this study, we used whole-genome sequencing (WGS) to type a collection of 36 *M. tuberculosis* isolates of the Manila family. WGS enabled the subtyping of these 36 isolates into at least 10 distinct clusters. Our results indicate that WGS is a powerful approach to determining the relatedness of Manila family *M. tuberculosis* isolates.**

Molecular typing techniques help identify linked tuberculosis (TB) cases and provide information that can be used to implement control measures to prevent further TB transmission. Current genotyping techniques, such as mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) typing and spoligotyping, are easy to execute, fast, and, in general, very useful typing methods (1–4). However, for some *M. tuberculosis* sublineages or families, even 24-locus MIRU-VNTR (MIRU24) typing and spoligotyping combined result in high clustering rates of isolates from otherwise unrelated cases.

One of these sublineages is the Manila family, which has been shown to have very low genetic variability, resulting in large genotyping clusters (4, 5). Isolates from this family are found throughout the Pacific basin, including the Philippines, as well as areas of Asia and the western United States (4, 5). In Ontario, Canada, the Manila family is one of the most commonly observed genotypes, representing 13% of all culture-positive cases diagnosed in Public Health Ontario laboratories.

The Manila family is a large clonal group for which classical genotyping techniques such as MIRU24 typing and spoligotyping provide low-resolution power (4, 5). Although IS6110 restriction fragment length polymorphism (RFLP) can relatively improve the discrimination power, it has also proven to be of low informative value, as the majority of isolates share the same pattern (4). In a recent study, Frink and colleagues developed a deletion-based subtyping assay to further parse out members of the Manila family. Deletion-based subtyping proved to decrease the clustering rate compared to that of MIRU24 typing and spoligotyping alone, and when used in combination with these two methods, the level of discrimination was even higher (5). Although relatively simple, this method requires the combination of several genotyping techniques in order to obtain meaningful results. Given the inability of currently used genotyping techniques to discriminate between unrelated isolates of the Manila family, it is difficult to draw meaningful conclusions during contact investigations for TB cases due to these isolates.

Several reports have shown that whole-genome sequencing (WGS) permits a much finer resolution than current genotyping techniques (6–9). WGS may also be used to identify potential genomic markers such as single-nucleotide polymorphisms (SNPs) that can be incorporated into a subtyping schema (10, 11).

The main goal of this study was to evaluate the utility of WGS and phylogenetic analysis to improve the genotyping of the Manila family of *M. tuberculosis*. Thirty-six clinical isolates with MIRU24 typing and spoligotyping patterns consistent with those of the Manila family were selected for WGS. Most isolates were part of two major Manila family subgroups found in our patient population and designated ON-1 ($n = 25$) or ON-2 ($n = 21$). The remaining isolates ($n = 9$) were also part of the Manila family but did not match the ON-1 or the ON-2 genotype. DNA extraction and WGS were performed as previously described (12).

Polymorphism discovery was performed with the VAAL algorithm, as previously described, using default parameters (13). In order to detect polymorphisms in regions not present in the H37Rv reference strain, we constructed a custom-made reference sequence. Briefly, the reference strain for polymorphism calling was custom made by joining the full sequence of *M. tuberculosis* reference strain H37Rv (GenBank accession number NC_018143.1), and contigs were *de novo* assembled using the A5 pipeline (14) from all short reads of a randomly selected Manila isolate that did not align to the H37Rv reference strain, as determined by the Mosaik assembler (<https://code.google.com/p/mosaik-aligner/>). These contigs were concatenated after the H37Rv sequence using the string NNNNNCATTCCATTCATTAATTAATTAATGAATGAATGNNNNN as a separator.

Using this approach, 3,786 polymorphic positions were identified in at least one of the 36 isolates compared to the reference. Of these, 263 (7%) were indels (insertions-deletions), and 3,523 (83%) were single nucleotide polymorphisms (SNPs).

Of the 3,786 polymorphic positions, 45.6% (1,728) were pres-

Received 18 June 2014 Returned for modification 27 June 2014

Accepted 23 July 2014

Published ahead of print 30 July 2014

Editor: G. A. Land

Address correspondence to C. Mehaffy, carolina.mehaffy@colostate.edu.

* Present address: C. Mehaffy, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado, USA.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.01726-14

FIG 1 Phylogenetic tree of 36 *M. tuberculosis* isolates from the Manila family, built in SplitsTree4 version 4.13.1 using the BioNJ algorithm. (A) Generated using all polymorphic sites. (B) Reconstructed using only parsimony-informative sites. Groups 1a to 1f constitute the major ON-1 group. Groups 2a to 2d constitute the major ON-2 group. Numbers shown with a gray background correspond to strains that did not belong to ON-1 or ON-2 by 24-locus MIRU-VNTR typing and spoligotyping.

ent in all 36 isolates, and 42% (1,590) were singletons, that is, variations present only in single isolates. The remaining 468 substitutions were either subcluster associated (218, 5.8%) or represented homoplastic variations (250, 6.6%), probably due to convergent evolution.

The 3,523 SNPs were concatenated in order of occurrence relative to the reference genome, used to reconstruct the phylogeny of Manila isolates using SplitsTree4 version 4.13.1 software (15) and the BioNJ algorithm (16), and reconstructed using the equal-angle algorithm (15) with equal-daylight and box-opening optimization (17). The phylogenetic tree was reconstructed with only parsimony-informative sites.

Our analysis confirmed the two major genotypes, ON-1 and ON-2. In addition, SNP-based phylogenetic analysis demonstrated the presence of at least 10 subclusters (1a to 1f and 2a to 2d), 6 within the ON-1 group and 4 within the ON-2 group (Fig. 1).

Eight of the nine isolates not categorized as ON-1 or ON-2 (by combined MIRU24 typing and spoligotyping) were grouped within the two large clusters; one was placed within ON-2 and the remaining seven within ON-1, with four making up the distinct subcluster 1e.

The two main groups, ON-1 and ON-2, were separated by 17 SNPs. ON-1 was further subdivided into 6 subclusters, 1a to 1f. Twenty-seven unique variants defined group 1a, subcluster 1b was defined by 5 SNPs, subcluster 1c was defined by 4 SNPs, subcluster 1d was defined by 31 SNPs, subcluster 1e was defined by 6 SNPs, and subcluster 1f was defined by 12 SNPs (Table 1).

ON-2 was further subdivided into 4 subclusters (2a to 2d).

Twenty-five unique SNPs defined group 2a, subcluster 2b was defined by 41 unique SNPs, and subcluster 2c was defined by 21 SNPs. Finally, subcluster 2d was defined by 12 unique SNPs (Table 1).

Our results demonstrate that WGS is a more powerful discriminatory technique than 24-locus MIRU-VNTR typing and spoligotyping, dramatically reducing the clustering rate and, at the same time, highlighting isolate relatedness otherwise missed by MIRU-VNTR typing and spoligotyping.

However, WGS is still a costly technology that is not available in many public health settings, and alternative and less expensive methods may be easier to implement. Since *M. tuberculosis* dem-

TABLE 1 Distribution of unique SNPs in each subcluster

Cluster	No. of unique SNPs		
	Synonymous	Nonsynonymous	Noncoding
ON-1/ON-2	5	7	5
1a	8	16	3
1b	1	4	0
1c	0	2	2
1d	12	18	1
1e	4	1	1
1f	6	6	0
2a	10	10	5
2b	11	24	6
2c	7	12	2
2d	7	4	1

TABLE 2 List of SNP targets and primers for SNP-based genotyping

Cluster	Position	GeneID	Position in gene	Ref_base	Cluster base	Primers (5' to 3')	
						Reverse	Forward
ON-1	4027458	Rv3585	836	G	T	CCACGGTGGACAGGTAGATG	GTCGACGTTGTGCTGCATTT
	2035794	Rv1797	305	G	A	AATGTGGAACCTGGCGCAAAC	GTGACGTTACTGTGCGGTGGT
	3004512	Rv2687c	245	C	T	AACGGCAACGAGGAACGTAA	TCTCCGGACTGATTTGGCTG
1a	4015527	Rv3573c	863	C	T	CACCGTTCTGGAGGTGTCG	GAGGAACCCACCGATTCCG
	110332	Rv0101	335	C	T	ATTCTTCACCTTGCCCGTC	TCGGCAAGAGCTATCGGTTC
1b	4231188	Rv3784	755	C	T	AGAACGCTGGAACCGCTAAA	TCTCAAGGAAGGCGAACGAC
	3776290	Rv3365c	1381	G	A	AGCTCGCAAGCAGATGAAT	CGACTGCTGGTCAACGAGAT
1c	2422395	Rv2161c	749	C	A	GGTGACGGTGGACAAGCTC	CGCTAGGCATGGCTTTTGAC
	2380124	Rv2121c	543	T	C	AAGGGAATCGAAGCAACGGT	CCACAGGATCAGAATCGGCA
1d	4120034	Rv3680	62	G	T	GTCTAAGAAGTCCAGCGCGT	CCTGCTTACCGAGACCATCC
	3578790	Rv3202c	1406	C	T	CTGATCGATGGGGTGCCTTG	CAGACTTCCTGAGCGGTGG
1e	962447	Rv0862c	167	G	A	ACCTTCAGCAATGTCAGCCA	GCCAGGGCACGTTGTTTAAAG
	83610	Rv0074	866	G	A	GGGTCATGGTCCACACCTG	CTGCACGTTCTTGAGCGAAG
1f	3113987	Rv2807	320	C	T	GCAGTCCGAATCGAAAACCG	GATCTTGACAAGCCGTTTCGC
	824225	Rv0731c	416	C	T	CGTTGAGCCGATTGATGTCG	AACAGCTGGTATGGACTGGC
2a	1118560	Rv1002c	1382	C	T	GATGGGCATTTTCGCTGGTTC	ATCGACAAGCTGCGGGTAAT
	2700340	Rv2402	1802	T	C	CTCGGCATCGGAAGTCCTAC	ACGTTACCATCTGCTCGTT
2b	1448012	Rv1292	1628	C	T	TCCTCGTCGATGACGAACAG	TTAACCACGACAAGGAGGGC
	877663	Rv0783c	779	C	T	GATGGGCATTTTCGCTGGTTC	GCGATCGTGTCCCAAGAGA
2c	972357	Rv0873	1850	C	T	CGCCCTAAAGTTTCGCTTCG	GGTGAGCAGGCATACGAACT
	3782513	Rv3370c	2406	G	A	GGTGATCGACCGGATCTACG	ACGGAAGCTGCACCCG
2d	2017033	Rv1781c	449	C	T	GATCTTGACGACGCTGGAT	TCCACCCGAAGGTAGAGAGG
	521289	Rv0433	963	C	A	GCATCGGCTAAATCACCAGC	AGAAGACCGGCATCATCGAC

onstrates low levels of homoplasy (18), SNPs are considered strong phylogenetic markers that can be used for strain classification (19). The usefulness of an SNP-based typing method to discriminate *M. tuberculosis* Manila isolates can be evaluated using synonymous SNPs which are considered neutral and are under less stringent selective pressure (20). For these reasons, synonymous SNPs are preferred over nonsynonymous SNPs for the purposes of SNP typing (21–25). We selected cluster-specific synonymous SNPs obtained from the WGS analysis and tested them using PCR and Sanger sequencing in 19 additional isolates classified as belonging to the Manila family by MIRU24 typing and spoligotyping (10 to ON-1 and 9 to ON-2). The primers were designed using Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm>) (Table 2).

Of the 23 selected SNP regions, 21 did not amplify after several attempts and therefore were removed from the SNP-based typing method. Of the 19 isolates, 12 (63%) were placed in one of the 10 subclusters (1a to 1f and 2a to 2d), while the remaining 7 (37%) were identified as only ON-1 or ON-2 but did not match any of our previously identified SNP-specific subclusters.

Although the data generated from WGS can be used to develop SNP-based typing assays to increase the level of discrimination obtained by classical genotyping techniques, even after including more than 20 targets, PCR-SNP-based typing was shown to be

inconclusive, possibly due to a higher genetic variability than expected. WGS is, therefore, a much more rapid and sensitive genotyping technique that yields accurate and highly discriminatory results.

In summary, we validated WGS as a better genotyping tool for identifying the relatedness of closely related Manila isolates. If implemented, information gathered from this technology can support public health teams during contact investigations of TB cases associated with genetically homogeneous *M. tuberculosis* isolates, such as those from the Manila family.

Nucleotide sequence accession number. The raw sequence data generated in this study were submitted to the Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/Traces/sra/>) under accession number SRP044223.

ACKNOWLEDGMENT

Two of the 36 isolates sequenced in this study were kindly provided by David Alexander, Saskatchewan Disease Control Laboratory.

REFERENCES

1. Mazars E, Lesjean S, Banuls AL, Gilbert M, Vincent V, Gicquel B, Tibayrenc M, Locht C, Supply P. 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc. Natl. Acad. Sci. U. S. A.* 98: 1901–1906. <http://dx.doi.org/10.1073/pnas.98.4.1901>.

2. Kanduma E, McHugh TD, Gillespie SH. 2003. Molecular methods for *Mycobacterium tuberculosis* strain typing: a users guide. *J. Appl. Microbiol.* 94:781–791. <http://dx.doi.org/10.1046/j.1365-2672.2003.01918.x>.
3. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J. Clin. Microbiol.* 39:3563–3571. <http://dx.doi.org/10.1128/JCM.39.10.3563-3571.2001>.
4. Douglas JT, Qian L, Montoya JC, Musser JM, Van Embden JDA, Van Soolingen D, Kremer K. 2003. Characterization of the Manila family of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 41:2723–2726. <http://dx.doi.org/10.1128/JCM.41.6.2723-2726.2003>.
5. Frink S, Qian L, Yu S, Cruz L, Desmond E, Douglas JT. 2011. Rapid deletion-based subtyping system for the Manila family of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 49:1951–1955. <http://dx.doi.org/10.1128/JCM.01338-10>.
6. Gardy JL, Johnston JC, Sui SJH, Cook VJ, Shah L, Brodtkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJM, Brinkman FSL, Brunham RC, Tang P. 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* 364:730–739. <http://dx.doi.org/10.1056/NEJMoa1003176>.
7. Kato-Maeda M, Ho C, Passarelli B, Banaei N, Grinsdale J, Flores L, Anderson J, Murray M, Rose G, Kawamura LM, Pourmand N, Tariq MA, Gagneux S, Hopewell PC. 2013. Use of whole genome sequencing to determine the microevolution of *Mycobacterium tuberculosis* during an outbreak. *PLoS One* 8:e58235. <http://dx.doi.org/10.1371/journal.pone.0058235>.
8. Török ME, Reuter S, Bryant J, Koser CU, Stinchcombe SV, Nazareth B, Ellington MJ, Bentley SD, Smith GP, Parkhill J, Peacock SJ. 2013. Rapid whole-genome sequencing for investigation of a suspected tuberculosis outbreak. *J. Clin. Microbiol.* 51:611–614. <http://dx.doi.org/10.1128/JCM.02279-12>.
9. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE. 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.* 13:137–146. [http://dx.doi.org/10.1016/S1473-3099\(12\)70277-3](http://dx.doi.org/10.1016/S1473-3099(12)70277-3).
10. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nübel U, Niemann S. 2012. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PLoS One* 7:e39855. <http://dx.doi.org/10.1371/journal.pone.0039855>.
11. Stucki D, Malla B, Hostettler S, Huna T, Feldmann J, Yeboah-Manu D, Borrell S, Fenner L, Comas I, Coscollà M, Gagneux S. 2012. Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. *PLoS One* 7:e41253. <http://dx.doi.org/10.1371/journal.pone.0041253>.
12. Jamieson FB, Guthrie JL, Neemuchwala A, Lastovetska O, Melano RG, Mehaffy C. 2014. Profiling of *rpoB* mutations and MICs to rifampicin and rifabutin in *Mycobacterium tuberculosis*. *J. Clin. Microbiol.*, in press.
13. Nusbaum C, Ohsumi TK, Gomez J, Aquadro J, Victor TC, Warren RM, Hung DT, Birren BW, Lander ES, Jaffe DB. 2009. Sensitive, specific polymorphism discovery in bacteria using massively parallel sequencing. *Nat. Methods* 6:67–69. <http://dx.doi.org/10.1038/nmeth.1286>.
14. Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012. An integrated pipeline for *de novo* assembly of microbial genomes. *PLoS One* 7:e42304. <http://dx.doi.org/10.1371/journal.pone.0042304>.
15. Dress AWM, Huson DH. 2004. Constructing splits graphs. *IEEE/ACM Trans Comput. Biol. Bioinform.* 1:109–115. <http://dx.doi.org/10.1109/TCBB.2004.27>.
16. Gascuel O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14:685–695. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a025808>.
17. Gambette P, Huson DH. 2008. Improved layout of phylogenetic networks. *IEEE/ACM Trans Comput. Biol. Bioinform.* 5:472–479. <http://dx.doi.org/10.1109/tcbb.2007.1046>.
18. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S. 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* 6:e311. <http://dx.doi.org/10.1371/journal.pbio.0060311>.
19. Gagneux S, Small PM. 2007. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect. Dis.* 7:328–337. [http://dx.doi.org/10.1016/S1473-3099\(07\)70108-1](http://dx.doi.org/10.1016/S1473-3099(07)70108-1).
20. Kimura M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, United Kingdom.
21. Baker L, Brown T, Maiden MC, Drobniowski F. 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg. Infect. Dis.* 10:1568–1577. <http://dx.doi.org/10.3201/eid1009.040046>.
22. Gutacker MM, Smoot JC, Migliaccio CAL, Ricklefs SM, Hua S, Cousins DV, Graviss EA, Shashkina E, Kreiswirth BN, Musser JM. 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* 162:1533–1543.
23. Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, Musser JM. 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J. Infect. Dis.* 193:121–128. <http://dx.doi.org/10.1086/498574>.
24. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbon MH, Bobadilla del Valle M, Fyfe J, Garcia-Garcia L, Rastogi N, Sola C, Zozio T, Guerrero MI, Leon CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendon A, Sifuentes-Osorio J, Ponce de Leon A, Cave MD, Fleischmann R, Whittam TS, Alland D. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* 188:759–772. <http://dx.doi.org/10.1128/JB.188.2.759-772.2006>.
25. Schürch AC, Kremer K, Hendriks ACA, Freyee B, McEvoy CRE, van Crevel R, Boeree MJ, van Helden P, Warren RM, Siezen RJ, van Soolingen D. 2011. SNP/RD typing of *Mycobacterium tuberculosis* Beijing strains reveals local and worldwide disseminated clonal complexes. *PLoS One* 6:e28365. <http://dx.doi.org/10.1371/journal.pone.0028365>.