Distinct genotypic profiles of the two major clades of \textit{Mycobacterium africanum}

Sidra E Gonçalves Vasconcellos\(^1\), Richard C Huard\(^2\), Stefan Niemann\(^3\), Kristin Kremer\(^4\), Adalberto R Santos\(^1\), Philip N Suffys\(^1\), John L Ho\(^5\)

\textbf{Abstract}

\textbf{Background:} \textit{Mycobacterium tuberculosis} is the principal etiologic agent of human tuberculosis (TB) and a member of the \textit{M. tuberculosis} complex (MTC). Additional MTC species that cause TB in humans and other mammals include \textit{Mycobacterium africanum} and \textit{Mycobacterium bovis}. One result of studies interrogating recently identified MTC phylogenetic markers has been the recognition of at least two distinct lineages of \textit{M. africanum}, known as West African-1 and West African-2.

\textbf{Methods:} We screened a blinded non-random set of MTC strains isolated from TB patients in Ghana (\(n = 47\)) for known chromosomal region-of-difference (RD) loci and single nucleotide polymorphisms (SNPs). A MTC PCR-typing panel, single-target standard PCR, multi-primer PCR, PCR-restriction fragment analysis, and sequence analysis of amplified products were among the methods utilized for the comparative evaluation of targets and identification systems. The MTC distributions of novel SNPs were characterized in both the Ghana collection and two other diverse collections of MTC strains (\(n = 175\) in total).

\textbf{Results:} The utility of various polymorphisms as species-, lineage-, and sublineage-defining phylogenetic markers for \textit{M. africanum} was determined. Novel SNPs were also identified and found to be specific to either \textit{M. africanum} West African-1 (\(Rv1332^{523}\); \(n = 32\)) or \textit{M. africanum} West African-2 (\(nat^{751}\); \(n = 27\)). In the final analysis, a strain identification approach that combined multi-primer PCR targeting of the RD loci RD9, RD10, and RD702 was the most simple, straightforward, and definitive means of distinguishing the two clades of \textit{M. africanum} from one another and from other MTC species.

\textbf{Conclusion:} With this study, we have organized a series of consistent phylogenetically-relevant markers for each of the distinct MTC lineages that share the \textit{M. africanum} designation. A differential distribution of each \textit{M. africanum} clade in Western Africa is described.

\textbf{Background}

Mycobacteria that cause human and/or animal tuberculosis (TB) are grouped together within the \textit{Mycobacterium tuberculosis} complex (MTC). The MTC is comprised of the classical species \textit{M. tuberculosis}, \textit{Mycobacterium africanum}, \textit{Mycobacterium microti}, and \textit{Mycobacterium bovis} (along with the widely used vaccine strain \textit{M. bovis} bacillus Calmette-Guérin [BCG]) [1-3], as well as newly recognized additions \textit{Mycobacterium caprae} and \textit{Mycobacterium pinnipedii} [4,5]. Although they are not presently officially described microorganisms, “\textit{Mycobacterium canettii}” (proposed name), the oryx bacillus, and the dassie bacillus are additional widely-accepted members of the MTC [6-8]. \textit{M. tuberculosis} is the predominant cause of human TB worldwide but \textit{M. africanum} and \textit{M. bovis} remain important agents of human disease in certain geographical regions. Of note, \textit{M. bovis} is naturally resistant to pyrazinamide, a first-line anti-TB drug [9], and so treatment of human TB caused by \textit{M. bovis} should not include pyrazinamide. Therefore, the correct identification of MTC isolates to the species level is important to ensure appropriate patient treatment, as well as for the
collection of epidemiological information and for implementing necessary public health interventions.

Mycobacteriological laboratory methods have traditionally utilized a series of tests based upon growth, microscopic, phenotypic, and biochemical properties in order to segregate the classical members of the MTC [10]. However, these tests can be slow-to-results, cumbersome, imprecise, non-reproducible, time-consuming, may not give an unambiguous result in every case, and may not be performed by every clinical microbiology laboratory. The relatively recent identification of DNA sequence level differences amongst the species of the MTC has greatly improved our capacity for performing molecular epidemiology, phylogenetic structuring of the MTC evolutionary tree, and MTC species determination. Molecular techniques, such as PCR, either alone or followed by sequence analysis or restriction fragment analysis (RFA), have proven particularly useful for the characterization of single nucleotide polymorphisms (SNPs) and/or chromosomal region-of-difference (RD) loci (such as insertions, deletions, and rearrangements) that are either lineage-, species-, or strain-specific [7]. Several groups have reported on the development of molecular protocols for the definitive identification of unknown MTC isolates to the species level by RD and/or SNP analysis [2,7,11-13] and clinical laboratories are now beginning to integrate such home-brew protocols into their routine identification protocols for acid-fast bacilli. The only currently available commercial protocol for MTC species identification is the GenoType MTBC® assay (Hain Lifescience, Nehren, Germany) that can differentiate M. tuberculosis from M. africanum, M. microti, M. caprae, M. bovis, and M. bovis BCG [14-16]. However, this test is limited in that it cannot differentiate all species of the MTC and it is not commercially available for diagnostic purposes in the USA.

In the past, M. africanum strains were generally identified by default, having first ruled-out both M. tuberculosis and M. bovis by the traditional battery of tests. Two biovars of M. africanum were commonly described that lay along the phenotypic continuum between M. tuberculosis and M. bovis [17]. We now understand that most strains formerly designated as M. africanum subtype II strains were actually M. tuberculosis [1,2,7,18-23], while strains formerly characterized as M. africanum subtype I can be segregated into two distinct genealogical clades on the basis of multiple genome sequence-level differences [1,2,7,23]. Several names have been given to each of the subtype I lineages in order to distinguish them. In this report we refer to the subtype I groupings as M. africanum West African-1 and M. africanum West African-2 [24,25]. For reference, as first described by Mostowy et al. [23], strains of M. afric-

uniquely possess the long sequence polymorphism (LSP) RD713, while M. africanum West African-2 (also known as clade 2 [26]) carries the defining LSPs RD701 and RD702. Huard et al. [7], recently confirmed the clade specificity of these RDs, identified and validated the first SNPs restricted to either M. africanum West African-1 or M. africanum West African-2, and placed several additional previously known and novel polymorphisms into a unified phylogenetic context vis à vis M. afric-

num West African-1 and M. africanum West African-2.

In the present study, we characterized the content of known phylogenetically relevant RDs and SNPs in a blinded, and M. africanaum-enriched, set of MTC strains isolated from TB patients in Ghana. The results of this evaluation established the utility of several consistent RD and SNP markers for M. africanaum identification and clade differentiation and allowed us to settle upon a focused approach for future evaluations. In addition, novel SNPs were identified and validated against a large and diverse collection of MTC species and found to be specific to either M. africanaum West African-1 (Rv1332^225) or M. africanaum West African-2 (nat^251), thereby further expanding the limited number of genetic markers that can be used to unambiguously differentiate the two M. africanaum lineages.

(This study contributed to the fulfillment of the Master’s degree requirements by S.E.G.V.)

Methods

MTC strains analyzed
A total of 175 unique isolates that represent all of the presently described members of the MTC were included in the analysis and were derived from three strain collections, maintained at different institutions. One set of strains (n = 47) came from the National Reference Center for Mycobacteria in Forschungszentrum, Borstel, Germany and was collected in 2001-2003 from patients with pulmonary TB in Ghana. This set of Ghana strains was provided in a non-random blinded fashion but was known to contain both M. africanum and M. tuberculosis (as controls). All strains were previously characterized using the GenoType MTBC® assay, as per the manufacturer’s instructions, and these results were provided subsequent to the derivation of species identity using RD markers. A complete listing of the Ghana collection isolates by strain number accompanies a recent article by Wirth et al. [24] (excepting all M. bovis from Ghana and the non-M. bovis strains 10514/01, 1473/02, and 5357/02) and was recently made available as part of the MIRU-VNTRplus database http://www.miru-vntrplus.org/MIRU/index.faces[27]. Another 124 isolates were of a well-described strain collection from the Weill Medical College of Cornell University, New York. The extensive molecular characterization of the Cornell
collection, and a complete listing by MTC species, unique identifier, and origin, was previously reported [7]. Only one isolate from that collection (M. tuberculosis strain W) was not included in the current evaluation. This sampling was composed of “M. canetti” (n = 5), M. tuberculosis (n = 44), M. africanum West African-1 (n = 12) (note: given previously as M. africanum subtype Ib), M. africanum West African-2 (n = 18) (note: given previously as M. africanum subtype Ia), the dassie bacillus (n = 4), the oryx bacillus (n = 2), M. microti (n = 10), M. pinnipedii (n = 7), M. caprae (n = 1), M. bovis (n = 14), and M. bovis BCG (n = 8). Lastly, 15 DNA samples were provided from the collection of the National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands [28]. These included strains of M. tuberculosis (strains 13 and 22), “M. canetti” (strains 116 and 119), M. africanum West African-1 (strain 92), M. africanum West African-2 (strains 6 and 85), M. microti (strains 25 and 62), M. pinnipedii (stains 76 and 81), M. bovis (117 and 128), and M. bovis BCG (2 and 71) (note: some strain identities are corrected as per [2,22]). The 4 strains underlined in the above were unique and the remaining 11 were also included in the Cornell collection [7]. All strains from the Ghana collection were screened for every marker of interest while strains of the Cornell and RIVM collections were screened selectively, as described in each respective section of the Results.

**PGG Analysis**

Frequently observed SNPs in the genes katG463 and gyrA35 are routinely assessed in order to broadly categorize isolates into defined MTC phylogenies, known as principal genetic groups (PGG) [29]. The distribution of SNPs in katG463 and gyrA35 suggests that PGG1 M. tuberculosis strains more closely resemble the most recent common ancestor of all M. tuberculosis strains than PGG2 strains, and PPG2 strains more so than PGG3 strains. MTC species along the M. africanum→M. bovis evolutionary track are also PGG1 [1,2]. SNP analysis of katG263 was used to further segregate PGG1a isolates from PGG1b strains [7,30]. Representatives of each PGG were included in the Cornell collection of MTC strains.

**MTC PCR-typing Panel**

In previous reports we described [2], and then expanded upon [7], a PCR-based protocol for the differentiation of the various MTC species on the basis of genomic deletions. This MTC PCR-typing panel targets eight independent loci for amplification (16S rRNA, cfp32 [Rv0577], Mid3 [IS1561], RD4 [Rv1510], RD7 [Rv1970], RD1 [Rv3877-Rv3878], RD9 [Rv2073c], and RD12 [Rv3120]), each of which either results in an amplicon of an expected size or fails, depending upon the genomic content of the MTC strain being evaluated. The resulting band-pattern that is observed following agarose gel electrophoresis is indicative of MTC species identity. Of note, the RD12 target region in M. bovis and M. caprae overlaps a specific LSP in “M. canetti” (RD12an), while the RD1 target region in M. bovis BCG overlaps a specific LSP in the dassie bacillus (RD1da). With this protocol, the pattern of bands for M. microti and M. pinnipedii are identical, while the pattern of bands for the orxy bacillus is the same as that of M. africanum West African-2. The MTC PCR-typing panel has been successfully applied to collections of MTC strains from Rio de Janeiro, Brazil, and Kampala, Uganda, in order to characterize the diversity of MTC species within these locales [21,31].

**PCR amplification primers and conditions**

Purified DNA was prepared for PCR as previously described [2]. For some strains, culture thermolysates (80°C for 30 min) were used as the source of DNA in PCR amplifications. The primers used for the MTC PCR-typing panel, the RDrrio flank multiplex, RD174, RD701, RD702, RD711, RD713, in addition to targets containing the the pks15/1 micro-deletions and SNPs at aroA285, cfp32311, gyrA95, gyrB1450, hsp65540, katG203, katG463, PPE552148, PPE552154, narGHJI251, RD13174, rpoB1049, rpoB1163, Rv15101129, and ThdB1197, were the same as described earlier [7,32,33]. For analysis of the loci RD8, RD9, RD10, RD701, and ThdB1 additional new site-specific 3-primer combinations were designed for each, similar to as previously detailed [32], and each included two deletion flanking primers and one primer internal to the deletion. The 3-primer PCR’s were each designed to amplify a product of one size when the target locus is intact or to produce a different band size when a known LSP is present. New primers were also designed to amplify a 1069-bp nat gene fragment and the SNP-containing targets in nat751 and Rv1332223. New primers, along with expected band sizes and the PCR program used to amplify, are listed in Table 1. The general PCR protocol was identical to that used previously [2,7]. PCR amplification from purified DNA was performed using the following cycling conditions: Program 1a (with an initial denaturation step of 5 min at 94°C, followed by 45 cycles of 1 min at 94°C, 1 min at 60°C, and 1 min at 72°C, and ending with a final elongation step for 10 min at 72°C) or program 2a (similar to program 1a but with an annealing temperature of 65°C). PCR testing of DNA thermolysates was performed in a similar manner using the following cycling conditions: Program 1b (with an initial denaturation step of 5 min at 94°C, followed by 45 cycles of 1 min at 94°C, 1 min at 60°C, and 4 min at 72°C, and ending with a final elongation step for 10 min at 72°C) or program 2b
(similar to program 1b but with an annealing temperature of 65°C). Programs 1b and 2b were also used to amplify from purified DNA when potential target PCR fragments were greater than 1,250 bp. PCR products were visualized as previously described by agarose gel electrophoresis [2]. Negative or unexpected positive PCR results were repeated at least once for confirmation. Importantly, all PCR tests included parallel samples containing DNA of *M. tuberculosis* strain H37Rv (ATCC 27294T) and either *M. africanum* West African-1 strain Percy16, *M. africanum* West African-2 strain ATCC 25420T, or *M. bovis* strain ATCC 19210T, where appropriate, as controls. All controls consistently provided the expected results for each particular marker screened. Negative control PCRs, lacking input DNA, were also included to control for DNA contamination.

PCR of the *nat* gene (1069 bp) was performed by a slightly different protocol. PCR Program 2a and a PCR reaction mix in 50 μl, with 40 pmol of each primer, 5 mM MgCl₂, 0.2 mM dNTPs, 1U Taq polymerase (Invitrogen, Brazil), PCR-buffer (10 mM Tris-HCl, 1.5 mM MgCl₂, 50 mM KCl, pH 8.3) (Invitrogen, Brazil), 10% glycerol, and 10 ng of target DNA were used in this case.

It should be noted that the *M. africanum* West African-1 and *M. africanum* West African-2-restricted LSPs were amplified by RD flanking primers [23] and analyzed as previously described [7] with the results based upon a size estimation of the PCR products on agarose gel. PCR amplification of RD713 in *M. africanum* clade 1 strains typically yields a 2,798 bp amplicon, while amplification of this locus in other MTC strains either results in a 4,248 bp product (PGG2 and PGG3 *M. tuberculosis*) or no PCR product (PGG1a MTC species with the partially overlapping RD7 deletion and PGG1b *M. tuberculosis* which possess additional genomic content at this locus [7]). In PCR amplification of RD711, most, but not all, *M. africanum* clade 1 strains are expected to yield a 944 bp amplicon while the remaining *M. africanum* West African-1 strains and

### Table 1 New primers used in this study.

| Target Locus | Primer Name | Nucleotide Sequence | PCR Program | Size (bp) |
|--------------|-------------|---------------------|-------------|-----------|
| New 3-primer combinations | RD8 | 5’ CAT GCT AAG CAG ATC GTC AGT TTT GA 3’ | 1a, 1b | 289/485 |
| | RD8FlnkF | 5’ GCC GCA TTG TCG GGG TGC CAT CAC ACC 3’ | | |
| | RD8FlnkR | 5’ CCG TTC CGG CGG GGT GAG TGC ACT 3’ | | |
| | RD9 | 5’ ACT CCC AGC GCT CGG TCG TGA CAG TAT CTT 3’ | 1a, 1b | 293/499 |
| | RD9FlnkF | 5’ ATT CGG TGG GCG CTG CGG CCA ATG TTT GGT 3’ | | |
| | RD9FlnkR | 5’ GTG TCT CGG CAC GAC CAA CTC GCT TGA CAG 3’ | | |
| | RD10 | 5’ CCG CCA CCT CGG GGT GAT TCC AGG AAC CAC 3’ | 1a, 1b | 291/478 |
| | RD10FlnkF | 5’ TCC GGC CCT GCC GCT GCA TAA TAG CGA 3’ | | |
| | RD10FlnkR | 5’ CTC GCC GGC AAG TCG GGC GGC ATC ATT CTC 3’ | | |
| | RD701 | 5’ ACT CGG CGG CTG AGG TGG TGG TT TGC 3’ | 1a, 1b | 350/487 |
| | RD701FlnkF | 5’ TCC AAA TTG TCG CCC TTC AGT GCG GTA CCC TCC 3’ | | |
| | RD701FlnkR | 5’ GAG GGG CAG CGG GGG GAA GTC G 3’ | | |
| | Tbd1 | 5’ AAG GAA CTG CGA GAT AGG ATC GCC AAT TTC 3’ | 1a, 1b | 298/475 |
| | Tbd1FlnkF | 5’ CAT AGA TCC CGG ACA TGG TG 3’ | | |
| | Tbd1FlnkR | 5’ CTA CCT CAT CTT CGG GTC CA 3’ | | |
| New primers for PCR-RFA | Rv1332 | 5’ GCC CTG CGC AGC CTG CAC GAA CCT GAG ATT 3’ | 1a, 1b | 344 |
| | Rv1332F | 5’ GGA TGG CCC CGA CTC GCG TGA AGT ATG 3’ | | |
| | Rv1332R | 5’ ACC CGG CAT CGA AGT TCG TCA CGG GAC TGA 3’ | 2a, 2b | 766 |
| | nat751 | 5’ TGG TGT ACC AAG GGG CAC CGC AAA CCA G 3’ | | |
| | nat751F | 5’ ATC GGT GCG ACA TAG TTG G 3’ | 2a, 2b | 1069 |
| | nat751R | 5’ GCC TTC TGC TCA AAG TGG CT 3’ | | |
| New amplification primers | nat | 5’ GAC CGA CCT CGG TTC 3’ | | |
| | natF | 5’ CAC CGA CCT CGG TTC 3’ | | |
| | natR | 5’ GTC CCT GAG CGG ATA AGG TT 3’ | | |
| Additional sequencing primers | nat | 5’ GAC CGA CCT CGG TTC 3’ | | |
| | natF | 5’ CAC CGA CCT CGG TTC 3’ | | |
| | natR | 5’ GTC CCT GAG CGG ATA AGG TT 3’ | | |
| Corrected primer from ref. 7 | katG203 | 5’ CAA GAA GTC CCT ATG GCC GGA CCA TGG TG 3’ | | |

1 for RD loci, expected band sizes given as intact/deletion present.

Vasconcellos et al. *BMC Infectious Diseases* 2010, 10:80
http://www.biomedcentral.com/1471-2334/10/80
Page 4 of 16
MTC species amplify a 2,885 bp product. With respect to RD701, all *M. africanum* West African-2 strains are expected to generate a 340 bp amplicon while strains from the other MTC species amplify a 2,081 bp PCR fragment. Likewise, for RD702, all *M. africanum* West African-2 strains are expected to amplify a 732 bp product while strains from the other MTC species produce a 2,101 bp PCR fragment. For this study, RD9 (as part of the MTC PCR typing panel), Tbd1, and RD701 were evaluated by both 2-primer and 3-primer PCR tests.

**SNP analysis by PCR-RFA and sequencing**

Characterization of the SNPs at *gyrA*95, *gyrB*1450, *hsp65*240, *katG*73, *katG*463, *narGHJI*251, *rpoB*049, *rpoB*1163, and *Rv1510*129 was performed by PCR-RFA [7,33]. For the analysis of the SNPs in *aroA*177, *3’cfp32*111, *mmpL*555, *nat*751, and Tbd1197 novel PCR-RFA procedures were developed, similar to those previously detailed [2,7]. The restriction enzymes and expected digest band sizes for each PCR-RFA are listed in Table 2. Amplified products from *M. tuberculosis* H37Rv and a second appropriate MTC species (see above) were included in all digest reactions as controls. All unexpected digestion results were repeated least once for confirmation. For each PCR-RFA evaluation, the PCR fragments from at least one strain of each digest pattern were sequenced in order to confirm the presence or absence of the target SNP.

Because it was not possible to develop a PCR-RFA based approach for characterization of the SNPs at *PPE55*275, *PPE55*225, *RD13*74, and *Rv1332c*223, SNP analysis for these markers was performed by direct sequencing of the PCR products. The same procedure was used for verification of micro-deletions in the *pks15* locus [34]. In most cases, the primers for PCR amplification primers were also used for sequencing, as previously described [2,7], with the exception of the 1069 bp *nat* fragment which was also sequenced using internal primers (Table 1). Sequencing was performed using the BigDye Terminator kit (PE Applied Biosystems) on an ABI 3730 DNA Analyzer, either at the Cornell University BioResource Center (Ithaca, NY) http://www.brc.cornell.edu or at the Oswaldo Cruz Foundation (PDTIS DNA Sequencing Platform/FIOCRUZ, Rio de janeiro, RJ); http://www.dbbm.fiocruz.br/PDTIS_Genomica/) and the results were analysed as previously described [2,7].

**Nucleotide sequence accession numbers**

Gene fragment sequences containing novel SNPs were submitted to GenBank for *M. africanum* West African-1 (Rv1332c*523; accession number FJ617580) and *M. africanum* West African-2 (*nat*51*; accession number FJ617579). Previously identified polymorphic gene fragment sequences are now available for *M. africanum* West African-1 (aroA*285 [FJ617581] and Tbd1197 [FJ617582]) and *M. africanum* West African-2 (*hsp65*240 [FJ617583]; *Rv1510*129 [GU270931]; and *rpoB* variants [FJ617584, FJ617585, FJ617586]).

**Results**

**Genetic characterization of MTC isolates by PCR deletion analysis**

For this study we applied the MTC PCR-typing panel to a blinded, *M. africanum*-enriched, challenge collection of MTC strains isolated from patients with TB in Ghana (n = 47). As a result, 18 *M. tuberculosis* isolates, 20 strains of *M. africanum* West African-1, and 9 *M. africanum* West African-2 strains were putatively differentiated [7]. Strains were identified as *M. tuberculosis* by the successful amplification of targets internal to the RD9 and RD12/RD12cano loci. Strains were identified as *M. africanum* West African-1 on the basis of failure of amplification of the RD9 locus but the successful amplification of the RD7 target region, while *M. africanum* West African-2 strains were putatively identified on the basis of failure of amplification of the RD13 locus.

**An exception to the common** *M. tuberculosis* MTC PCR-typing panel profile occurred with 9 *M. tuberculosis* strains from Ghana, which failed to amplify the IS1661′ target (see Fig. 1B). Previously, strains with this particular band pattern were found to share a clonal deletion called RD190 that defines a major, newly
**Table 2 Summary of PCR-RFA protocols used in this study**

| Locus   | Restriction enzyme | MTC species                                      | Predicted digest pattern (bp) |
|---------|--------------------|--------------------------------------------------|------------------------------|
| katG    | BstNI              | PGG1, MTC                                        | 12, 59, 106, 174             |
|         |                    | PGG2, PGG3, M. tuberculosis                      | 12, 59, 280                  |
| gyrA    | Ael                | PGG1, PGG2 MTC                                   | 161, 193                     |
|         |                    | PGG3, M. tuberculosis                            | 354                          |
| narGHJI | Sau3AI             | "modern" M. tuberculosis                         | 155                          |
|         |                    | Remaining MTC                                    | 69, 86                       |
| gyrB    | Taq1α              | "M. canetti", M. tuberculosis                    | 6, 21, 74, 96, 129, 270, 444 |
|         |                    | Remaining MTC                                    | 6, 21, 74, 96, 129, 163, 107, 444 |
| katG    | BstNI              | "M. canetti", M. tuberculosis, M. africanum WA-1 | 140, 230                     |
|         |                    | Remaining MTC (PGG1a)                            | 370                          |
| 3’tcp32 | BstNI              | M. tuberculosis, M. africanum WA-1              | 28, 34, 75, 235²             |
|         |                    | Remaining MTC (PGG1a)                            | 34, 75, 263                  |
| aroA    | BssHII             | M. africanum WA-1                               | 177, 254                     |
|         |                    | Remaining MTC                                    | 104, 150, 177                |
| TbD1    | TseII              | M. africanum WA-1                               | 46, 454                      |
|         |                    | Remaining MTC                                    | 46, 193, 261³                |
| Rv1510  | NruI               | M. africanum WA-2, dassie bacillus              | 192, 841                     |
|         |                    | Remaining MTC                                    | 1033⁴                        |
| hsp65   | AflIII             | M. africanum WA-2                               | 49, 392                      |
|         |                    | Remaining MTC                                    | 441                          |
| narJ    | BclI               | M. africanum WA-2                               | 111, 177, 208, 270           |
|         |                    | Remaining MTC                                    | 111, 208, 447                |
| rpoB    | Sau3AI             | M. africanum WA-2                               | 11, 12, 18, 69, 87, 163      |
|         |                    | Remaining MTC, M. africanum WA-2                 | 11, 12, 18, 69, 250          |
| rpoB    | BstUI              | M. africanum clade 2                             | 28, 332                      |
|         |                    | Remaining MTC, M. africanum WA-2                 | 28, 79, 253                  |
| mmpL6   | Hpy166III          | "M. canetti", M. tuberculosis, M. africanum WA-1, M. africanum WA-2, dassie bacillus | 155, 298³                     |
|         |                    | Remaining MTC                                    | 453                          |

**WA, West African**

¹ This table contains the correction of errors from ref. [7]
² "M. canetti" fails to PCR amplify
³ "Modern" M. tuberculosis fails to PCR amplify
⁴ M. bovis fails to PCR amplify

[7] Vasconcellos et al. BMC Infectious Diseases 2010, 10:80
http://www.biomedcentral.com/1471-2334/10/80
Figure 1 The composite MTC PCR typing panel. Illustrated is the MTC PCR typing panel output pattern for A) a typical *M. tuberculosis* strain, B) a secondary pattern seen in some Cameroon genotype *M. tuberculosis* strains from Ghana, C) a typical *M. africanum* West African-1 strain, D) a typical *M. africanum* West African-2 strain, and E) a typical *M. bovis* strain [7]. PCR products and the 100-bp ladder (unlabelled lanes) were visualized by agarose gel electrophoresis and ethidium bromide staining. Lanes: 1, 16S rRNA; 2, cfp32 (Rv0577); 3, MiD3 (IS1561') ; 4, RD4 (Rv1510); 5, RD7 (Rv1970); 6, RD1 (Rv3877-Rv3878); 7, RD9 (Rv2073c); 8, RD12 (Rv3120). WA - West African.
recognized, lineage of *M. tuberculosis* that is the predominant cause of TB in Rio de Janeiro, Brazil, and that has disseminated to many countries around the world [7,31,32]. However, multiplex PCRs for both the RD<sup>Ri</sup> LSP and the coincident RD174 deletion [32] showed that these Ghanaian strains were not RD<sup>Ri</sup> genotype *M. tuberculosis*. Rather, data from the MIRU-VNTRplus website identified these strains as being of the RD726-harboring Cameroon genotype (ST61 and variants) and lists the strains as lacking IS<sup>1561</sup>' [27]. The Cameroon genotype therefore appears to possess an undefined LSP of IS<sup>1561</sup>' that overlaps RD<sup>Ri</sup> (Fig. 2; see node 4) and the MiD3 locus in *M. microti* and *M. pinnipedii* (Fig. 2; see node 16) [7,31].

### Table 3 Summary of results targeting MTC polymorphic loci

| Target Locus | Node #(s) in Fig. 2 | *M. tuberculosis* (n) | *M. africanum* WA-1 (n) | *M. africanum* WA-2 (n) | Other MTC (n) |
|--------------|---------------------|-----------------------|-------------------------|------------------------|-------------|
| RD loci      |                     |                       |                         |                        |             |
| RD12<sup>−</sup>/RD12 1 | 1, 17               | intact (18)           | intact (20)             | intact (9)            |             |
| 3cfp32       | 1                   | intact (18)           | intact (20)             | intact (9)            |             |
| TbD1         | 2                   | deleted (18)          | intact (20)             | intact (9)            |             |
| pks15/1      | 3, 9                | 7-bp deletion (18)    | intact (20)             | 6-bp deletion (9)     |             |
| PPESS 2      | 4, 16               | intact (9)            | intact (20)             | intact (9)            |             |
| MiD3/RD<sup>No</sup> 2 (IS<sup>1561</sup> + PPESS) | 4, 16               | intact (9)            | intact (20)             | intact (9)            |             |
| RD9          | 6                   | intact (18)           | deleted (20)            | deleted (9)           |             |
| RD713 3      | 7                   | fail (18)             | deleted (20)            | fail (9)              |             |
| RD711        | 8                   | intact (18)           | deleted (20)            | intact (9)            |             |
| RD7         | 3                   | intact (18)           | intact (20)             | deleted (9)           |             |
| RD8          | 9                   | intact (18)           | intact (20)             | deleted (9)           |             |
| RD10         | 9                   | intact (18)           | intact (20)             | deleted (9)           |             |
| RD701        | 11                  | intact (18)           | intact (20)             | deleted (9)           |             |
| RD702        | 11                  | intact (18)           | intact (20)             | deleted (9)           |             |
| RD<sup>148</sup>/RD<sup>159</sup> 4 | 14, 19              | intact (18)           | intact (20)             | intact (9)           |             |
| RD13         | 17                  | intact (18)           | intact (20)             | intact (9)            |             |
| RD4          | 18                  | intact (18)           | intact (20)             | intact (9)            |             |
| SNP loci     |                     |                       |                         |                        |             |
| narGHJI<sup>−215</sup> | 2                   | C→T (18)              | no Δ (20)               | no Δ (9)              |             |
| katG<sup>−463</sup> | 3                   | CGG (18)              | no Δ (20)               | no Δ (9)              |             |
| gyn<sup>−455</sup> | 5                   | no Δ (18)             | no Δ (20)               | no Δ (9)              |             |
| gyn<sup>−465</sup> | 6                   | G→T (18)              | G→T (20)               | G→T (9)              |             |
| TbD1<sup>−97</sup> | 7                   | fail (18)             | C→T (20)               | no Δ (9)              |             |
| arao<sup>−265</sup> | 7                   | no Δ (18)             | G→A (20)               | no Δ (9)              |             |
| Rv1332<sup>−223</sup> | 7                   | no Δ (27)             | G→T (32)               | no Δ (11)             | no Δ (15) |
| katG<sup>−203</sup> | 9                   | no Δ (18)             | no Δ (20)               | ACC→ACT (9)           |             |
| 3cfp32<sup>−311</sup> | 9                   | no Δ (18)             | no Δ (20)               | G→A (9)              |             |
| RD13<sup>−174</sup> | 9                   | no Δ (18)             | no Δ (20)               | G→A (9)              |             |
| PPESS<sup>−148</sup> | 9                   | no Δ (9)              | no Δ (20)               | A→G (9)              |             |
| PPESS<sup>−154</sup> | 9                   | no Δ (9)              | no Δ (20)               | A→G (9)              |             |
| Rv1510<sup>−120</sup> | 10                  | no Δ (18)             | no Δ (20)               | G→A (9)              |             |
| hsp65<sup>−340</sup> | 11                  | no Δ (18)             | no Δ (20)               | C→G (9)              |             |
| nar<sup>−251</sup> | 11                  | no Δ (63)             | no Δ (32)               | G→A (27)             | no Δ (53) |
| tpo<sup>−4163</sup> | 12                  | no Δ (18)             | no Δ (20)               | C→T (9)              |             |
| tpo<sup>−1049</sup> | 13                  | no Δ (18)             | no Δ (20)               | no Δ (9)              |             |
| mmap<sup>−551</sup> | 15                  | fail (18)             | no Δ (20)               | no Δ (9)              |             |
| Other targets |                     |                       |                         |                        |             |
| cfp32        |                     | intact (18)           | intact (20)             | intact (9)            |             |

<sup>1,2,3,4</sup> pairs of partially overlapping LSPs
WA, West African; Fail, no PCR amplification; bp, base pair; Δ, change
In addition to the MTC PCR-typing panel, some PCR targets used in SNP analysis, as will be described below, amplify from genomic regions that are deleted in some MTC species or lineages [7]. The successful amplification of the 3' cfp32 and RD13 loci in all the strains of the Ghana collection confirmed the species distribution obtained using the MTC PCR-typing panel, as these targets are deleted in either "M. canettii" (Fig. 2; see node 1) or both M. caprae and M. bovis (Fig. 2; see node 17), respectively [7]. Furthermore, PPE55 is located proximal to IS1561' and so the failure to amplify PPE55 from the 9 Cameroon genotype M. tuberculosis isolates is consistent with a single genomic deletion in the region of IS1561' (Fig. 2; see node 4). Lastly, TbD1 is an important phylogenetic marker that categorically divides M. tuberculosis into two major lineages [1]. All M. tuberculosis isolates in the Ghana collection failed to amplify from targets internal to TbD1 (Fig. 2; see node 2), while all M. africanum clades 1 and 2 strains yielded an amplicon of the correct size, consistent with the previous finding that isolates from the M. africanum→M. bovis evolutionary tract are all TbD1-positive and likely

Figure 2 Summary diagram and phylogenetic positions of the genomic markers interrogated against the Ghana MTC strain collection. Shown are the various major divisions of the MTC segregated according to the presence or absence of inter-species-, intra-species-, and sublineage-specific polymorphisms. Circles are placed at points in evolutionary history beyond which each strain that was evaluated possessed a consistent set of polymorphisms. The nodes are numbered in the figure as follows to denote: 1. RD12, 3' cfp32 deletion; 2. TbD1, narGHIJ - 215; 3. pks15/1 (7-bp deletion), katG - 4; undefined deletion at the RD12/MACD locus; 5. gyrA - 6; RD9, gyrB - 10; 7. RD13, TbD1 - 107; aroA - 205; Rv1332c - 205; 8. RD711; 9. RD7, RD8, RD10, pks15/1 (6-bp deletion), katG - 203; 3' cfp32 - 211; RD13 - 174; PPE55 - 146; PPE55 - 154; 10. Rv1510c - 129; 11. RD701, RD702, hsp65 - 240; nat - 51; 12. rpoB - 1163; 13. rpoB - 1049; 14. RD1 - 151; 15. mmpL6 - 155; 16. M93; 17. RD12, RD13; 18. RD4; 19. RD1 BCG. Lineages that include strains from the Ghana collection are terminated with arrowheads. Note that distances are arbitrary and do not reflect the number of phylogenetically relevant polymorphisms present at each juncture. TbD1-positive M. tuberculosis is also known as "ancient" M. tuberculosis and TbD1-negative M. tuberculosis is also known as "modern" M. tuberculosis [1]. WA - West African.
a branch off of a TbD1-positive *M. tuberculosis* lineage [1,24].

We next evaluated the Ghana strain collection by PCR (using LSP flanking primers) for RDs that have been described previously as being either specific to *M. africanaum* West African-1 (RD713), restricted to a subgroup of *M. africanaum* West African-1 (RD711), or specific to *M. africanaum* West African-2 (RD701 and RD702) [7,23]. All *M. africanaum* West African-1 strains (\(n = 20\)) yielded amplification products for RD711 and RD713 of shorter band sizes that were consistent with amplicons that bridge a deletion (Fig. 2; see nodes 7 and 8). All *M. tuberculosis* strains (\(n = 18\)) contained the RD711 and RD713 regions, while each *M. africanaum* West African-2 strain (\(n = 9\)) yielded PCR fragments suggestive of intact RD711. Each *M. africanaum* West African-2 strain (\(n = 9\)) also failed to produce any amplification products from the RD713 locus region, as expected, owing to the overlapping RD7 [7].

Likewise, all *M. africanaum* West African-2 strains produced shortened RD701 and RD702 amplicons (Fig. 2; see node 11), while each *M. tuberculosis* and *M. africanaum* West African-1 strain exhibited PCR fragments representative of intact sequences within these loci. The *M. africanaum* clade-specific bridge-deletion PCR results were therefore congruent with the MTC PCR-typing panel data.

A drawback, however, of the MTC PCR-typing assay as it was designed is that overlapping polymorphisms may occur in the target regions of the panel. Such hypothetical LSPs would therefore have the potential to cause a failure in amplification and to confuse the interpretation of banding patterns which may, in turn, lead to erroneous species determinations. To begin to address this issue, with respect to loci relevant to the species within the current Ghana collection, we developed new 3-primer combination sets for RD8, RD9, RD10, RD701, and TbD1 (Table 1). As was expected from previous phylogenetic evaluations [1,3,7], each of the test loci were found to be intact in the Ghana collection PGG2 *M. tuberculosis* strains, excepting TbD1. Moreover, excepting RD9, each of the studied RDs were intact in the *M. africanaum* West African-1 strains, while in the *M. africanaum* West African-2 strains only TbD1 remained intact, i.e. the RDs 8-10 and RD701 were deleted. Overall, no inconsistencies were observed with respect to species identification within the Ghana MTC strain collection across the different strategies for PCR deletion analysis that were employed.

**Genetic characterization of MTC isolates by SNP analysis**

For the second stage of this study we screened the Ghana MTC collection for known phylogenetically relevant SNPs. With respect to the *M. tuberculosis* strains, we determined that all were PGG2 (\(n = 19\)) (Fig. 2; see nodes 3 and 5). Consistent with this determination, the 7-bp *pks15/1* micro-deletion was observed in all the *M. tuberculosis* strains; this polymorphism is positioned at the same point along the MTC evolutionary tree as the *katG* CTG→CGG SNP that marks PGG2 *M. tuberculosis* strains (Fig. 2; see node 3). Likewise, an SNP in the *narGHJI* operon promoter (−215 C→T), that is phylogenetically coincident with TbD1 [33] was also present in all of the Ghanaian *M. tuberculosis* isolates evaluated (Fig. 2; see node 2). Lastly, the *gyrB* G→T polymorphism (also a target of the GenoType MTBC® assay [14-16]) is known to coincide with the RD9 deletion and likewise segregated the *M. tuberculosis* isolates from the strains of the *M. africanaum* strains (Fig. 2; see node 6).

The following considers SNPs that inform the phylogenetic interrelationships among most of the non-*M. tuberculosis* MTC species. First, all the *M. africanaum* strains (\(n = 28\)) were PGG1. Previously, an ACC→ACT SNP at *katG* has been used to segregate PGG1 strains into PGG1a and PGG1b [30]. Huard et al. [7] reported that this SNP is present in *M. africanaum* West African-2 and all downstream species in the MTC evolutionary tree (Fig. 2; see node 9). As expected, the Ghana collection *M. africanaum* West African-1 strains were determined to be PGG1b, while the *M. africanaum* West African-2 strains were PGG1a by *katG* analysis. Additional inter-species-specific SNPs that colocalize with the *katG* SNP and segregate the *M. africanaum* clades (and are also notably coincident with RD7, RD8, and RD10) have also been reported at 3′*cfp32* G→A, *PPE5* G→A, *PPE5* A→G, and RD13 G→T, in addition to a 6-bp *pks15/1* micro-deletion (Fig. 2; see node 9) [7,34]. These loci were interrogated and indeed found to partition the *M. africanaum* West African-2 strains from the *M. africanaum* West African-1 and *M. tuberculosis* strains of the Ghana collection, consistent with previous reports [7,34]. Lastly, we also screened for an inter-species-specific SNP in *mmpL6* (AAC→AAG) [1,7] that is not observed in *M. africanaum* West African-1, *M. africanaum* West African-2, nor the dassie bacillus, but is present in all of the remaining distal species along the oxyr bacillus→*M. bovis* evolutionary track of the MTC phylogenetic tree [1,26]. As was expected, we found *mmpL6* to be unaltered in the *M. africanaum* West African-1 and West African-2 strains of the Ghana MTC collection (Fig. 2; see node 15). The *mmpL6* SNP occurs within a TbD1 locus gene and was thus deleted in the TbD1-negative *M. tuberculosis* strains of the Ghana collection.

We then investigated SNPs that have been previously described to be restricted to either *M. africanaum* West African-1 or *M. africanaum* West African-2 within the MTC [7]. SNPs at *aroA* G→A and TbD1 T→C (Fig. 2; see node 1)
were found to be limited to the *M. africanum* West African-1 strains of the Ghana MTC collection, thereby coinciding with the *M. africanum* West African-1-specific LSP RD713 (Fig. 2; see node 7). Point mutations at *Rv1510* (G→A), *hsp65* (C→G), and *rpoB* (C→T) were also screened and found to be restricted to the *M. africanum* West African-2 strains (Fig. 2; see nodes 10; 12); a previously noted sublineage-specific SNP at *rpoB* (C→T) was not observed (Fig. 2; see node 13). However, from previous data [7], only *hsp65* has been shown to be truly *M. africanum* West African-2-specific and to associate pylogenetically with RD701 and RD702. In fact, *Rv1510* was previously found to be an inter-species-specific SNP that *M. africanaum* West African-2 shares with the dassie bacillus, and is indicative of a common ancestor between these species, while not all *M. africanum* West African-2 strains possess the *rpoB* and *rpoB* (C→T) point mutations appear to have been acquired in a stepwise sequential order and to define the branch points of sublineages within the *M. africanum* West African-2 species. All Ghana *M. africanum* West African-2 strains evaluated in this study therefore fell into the second of three potential *rpoB* sequence-based sublineage branches. Overall, each of the known MTC inter-species-specific, species-specific, and sublineage-specific SNPs for which the Ghana MTC collection was evaluated were entirely consistent with the current RD analyses and showed a species distribution that paralleled previous descriptions [7].

**Identification of a novel Mycobacterium africanum West African-1-specific ** *Rv1332* SNP

In the process of sequencing the RD711 bridge amplicon to confirm its correct amplification in an *M. africanaum* West African-1 strain, we noted a nonsynonomous G→A SNP in the region 5′ of the RD711 deletion breakpoint and within the *Rv1332* gene, affecting nucleotide 1423. To investigate the distribution of this *Rv1332* SNP amongst the MTC species, we generated a new primer pair to amplify the SNP-containing region upstream of RD711. We then performed PCR and sequence analysis of the amplified products upon samples from select MTC strains of the Cornell collection representing each of the MTC species and major *M. tuberculosis* lineages, i.e., *“M. canetti”* (n = 2), *TbD1*-positive *M. tuberculosis* PGG1 (n = 2), *TbD1*-negative *M. tuberculosis* PGG2 (n = 2), *M. tuberculosis* PGG3 (n = 3), *M. africanaum* West African-1 (n = 12), *M. africanum* West African-2 (n = 2), the dassie bacillus (n = 2), the oryx bacillus (n = 2), *M. microti* (n = 2), *M. pinnipedi* (n = 2), *M. caprae* (n = 1), *M. bovis* (n = 2), and *M. bovis* BCG (n = 2). Only the 12 *M. africanaum* West African-1 strains possessed the *Rv1332* substitution. When the Ghana collection was subsequently evaluated (n = 47), the *Rv1332* SNP was likewise restricted to the 20 *M. africanaum* West African-1 strains. In total, 85 MTC isolates were screened, 32 of which were *M. africanum* West African-1. The data thus supported that the *Rv1332* SNP is a specific marker for *M. africanaum* West African-1 and is only the third such polymorphism reported to date (Fig. 2; see node 7) [7].

**Identification of a novel Mycobacterium africanaum West African-2-specific ** nat*751* SNP

Previously, the nat (*Rv3566c*) gene product aminyl-N-acetyltransferase has been investigated as a potential contributor to reduced isoniazid susceptibility in *M. tuberculosis* [36]. In the course of those investigations, SNPs were identified in the *nat* gene that were restricted to different *M. tuberculosis* lineages. We found a novel nonsynonomous G→A SNP in two *M. africanaum* West African-2 strains at *nat* nucleotide 751 (*nat*540, E251K) upon amplification and sequencing of a 1069-bp fragment using samples from a subset of MTC representative strains (RIVM collection; n = 15). Test sequencing of the 1069-bp *nat* amplicon from 16 MTC strains from the Cornell collection supported the limited distribution of the *nat*540 SNP. We then developed a PCR-RFA protocol for the *nat*540 SNP, amplifying a shorter product using new primers and employing the restriction enzyme *BglI*, and applied the protocol to all strains of both the Cornell (n = 124) and Ghana collections (n = 47). Consistent with the preliminary test results, all MTC isolates amplified *nat* successfully. However, only the 27 *M. africanaum* West African-2 strains possessed the *nat*540 polymorphism, as determined by PCR-RFA. The West African-2 strains showed a 4-band digest pattern on agarose gel electrophoresis as opposed to the remaining MTC strains that showed a 3-band digest pattern (see Table 2). Thus, this SNP appears to be a specific marker for *M. africanaum* West African-2 (n = 175 unique MTC strains evaluated in total) and is only the second SNP reported to be restricted to this clade (Fig. 2; see node 11) [7]. Of note, both the *nat*540 and *hsp65* *M. africanaum* West African-2-specific SNPs are present in the genomic sequencing project of *M. africanaum* strain GM041182 that is currently nearing assembly completion http://www.sanger.ac.uk/sequencing/Mycobacterium/africanaum/.

**Discussion**

*M. africanaum* has been reported to be an important cause of TB in the West African countries of Guinea-Bissau (52%) [37], The Gambia (38%) [38], Sierra Leone (24%) [39], Senegal (20%) [17], Burkina Faso (18.4%) [40], Cameroon (9%) [41], Nigeria (8%) [42], and Côte
D’Ivoire (5% of cases) [22]. *M. africanum* has also been identified in the West African countries of Benin, Mauritania, and Niger [7,43]. Many of the previous *M. africanum* reports appeared, however, before molecular markers distinguished two different clades within this species [1,7,23,25,26]. Therefore, this study is one of the few to use clade-specific molecular markers to investigate the diversity of *M. africanum* strains causing TB within a specific African locale. Previous MTC species surveys that characterized strains using truly informative phylogenetic markers identified *M. africanum* West African-1, but not West African-2, in Cameroon and Nigeria [41,42] or *M. africanum* West African-2, but not West African-1, in The Gambia [38,44] and Guinea-Bissau [23,45]. In contrast, with this study, we highlight the fact that both clades of *M. africanum* are contributing to the TB burden in Ghana [24]. However, because the Ghana MTC collection was not representative, the current study does not allow us to estimate the proportion of TB caused by the various MTC clades in this country. Such a systematic survey of MTC population structure in Ghana is currently in progress.

In actuality, few reports have definitively shown an overlap in the geographic ranges of *M. africanum* West African-1 and *M. africanum* West African-2. Previously, Huard et al. [7] studied isolates derived from patients in Niger that constituted both *M. africanum* clades; both lineages were likewise found to coexist in Sierra Leone [39]. In the absence of a molecular analysis similar to that presented herein, it is not known for certain which *M. africanum* clade predominates in many of the other *M. africanum*-endemic West African countries or if their ranges coincide elsewhere. However, a cross-comparison of molecular epidemiologic evidence presented in some earlier reports [17,46] and more recent data [7,41,43] does suggest that *M. africanum* clades 1 and 2 may both occur in at least Côte D’Ivoire, a country that borders Ghana. The picture that emerges from the combined studies [7,17,22-24,30,37-48] is of a differential geographic distribution of the *M. africanum* lineages, with West African-1 predominating in Eastern-West Africa (Cameroon, Nigeria), West African-2 in Western-West Africa (the Gambia, Guinea-Bissau, Senegal), and the two clades overlapping in Central-West Africa (Côte D’Ivoire, Ghana, Niger, Sierra Leone) (Fig. 3). A conceptually similar gradient of *M. africanum* prevalence across Western Africa was recently hypothesized by de Jong et al., but their analysis did not make a distinction between the two *M. africanum* clades [48]. Lastly, although TB caused by *M. africanum* is concentrated in sub-Saharan West African countries, with immigration and international travel, sporadic cases have also been reported in the USA, the Caribbean, and Europe [28,43,49], including one outbreak of multi-drug resistant *M. africanum* at a Parisian hospital [17,50]. With improved molecular methods of identification, we expect that further cases of infection will be identified outside of the traditional endemic areas of *M. africanum*.

Molecular systems are preferred for the differentiation of *M. africanum* from *M. tuberculosis* and *M. bovis*.
given the heterogeneous phenotypic patterns among *M. africanum* strains, and the prolonged time-to-results and subjectivity inherent to the interpretation of some tests. Importantly, previous data indicate that there are no definitive phenotypic characteristics that can be exploited to differentiate the individual *M. africanum* clades [17,22,45]. In this study, we identified novel *M. africanum* clade-defining SNPs and confirmed the MTC distribution of several other phylogenetically relevant markers among the MTC. Multiple validated intra-species-specific molecular markers are important because they cross-corroborate each other and increase confidence in a given MTC species identification. By the markers described herein, *M. africanum* West African-1 would be defined genotypically as possessing RD713 and SNPs at *aroA* [285], *Rv1335* [223], and *TbD1* [197], while *M. africanum* West African-2 would be defined genotypically by RD701 and RD702, as well as the intra-species-specific SNPs at *hsp65* [540] and *nar* [251]. Other SNPs and RDs that mark particular branches of the MTC phylogenetic tree, such as *gyrB* [1450], *Rv1510* [1229], RD9, and RD10 are also informative of *M. africanum* clade identity and provide further cross-referencing options. However, a streamlined protocol that employs 3-primer PCRs for RD9, RD10, and RD701 was the most rapid, simple, straight-forward and definitive means of differentiating the two clades of *M. africanum* from one another and from other MTC species. This approach limits the number of individual PCR reactions required for identification and eliminates the need for secondary procedures, such as restriction digestion, sequence analysis, or hybridization. Of note, some methods cannot distinguish the two clades of *M. africanum*, such as the GenoType MTBC line-probe assay [14–16]. Because PCR-RFA for SNPs specific to one of the *M. africanum* clades, as described herein, is a relatively simple approach, it may be of benefit for confirmation of species identification in laboratories with limited access to more advanced molecular methods. Other methods for *M. africanum* identification, such as by real-time PCR, microarray analysis, and spoligotyping (a DNA typing method) may also present advantages to laboratories with these capabilities, but these modalities were not evaluated in the current study.

Indeed, all strains of *M. africanum* are also known to lack spacers 9 and 39 in their spoligotype profile, similar to *M. bovis*, but possess one or more spacers that are consistently absent in certain other MTC species [7,25]. Previous data [17,23,37,46] suggest that many, but not all, *M. africanum* West African-1 strains demonstrate an absence of spacer 8 in addition to 9 and 39 (known as spoligotype signature AFRL1.2) [43], while *M. africanum* West African-2 strains may further uniformly lack spacers 7-9 and 39 (known as spoligotype signature AFRL1). As provided on the MIRU-VNTRplus website, all *M. africanum* West African-1 strains from the Ghana collection lacked spacers 8, 9, and 39, while each *M. africanum* West African-2 strain from the Ghana collection lacked spacers 7-9 and 39 [27]. Spoligotyping may therefore provide a preliminary indicator for each *M. africanum* clade [51,52], however, the validity of these associations remains to be conclusively determined using a sample set of isolates with diverse geographical origins.

In addition to identification, MTC species and sublineage specific markers are of importance for genealogical purposes, as they allow the construction of more accurate phylogenetic trees. In recent years, SNP typing has been used to group strains of *M. tuberculosis* [53,54], while LSP analyses and DNA sequencing approaches have been used to establish congruent phylogenies for the *M. tuberculosis* complex [25,51,55]. The species- and sublineage-specific polymorphisms examined in this study for the *M. africanum* clades may therefore be of benefit when characterizing the evolutionary history of MTC strain sets in the future. SNPs in *rpoB*, for instance, demarcate the sequential divergence of sublineages within *M. africanum* West African-2 [7]. Similarly, we previously highlighted that RD711 is deleted in most, but not all of the RD713-harboring *M. africanum* West African-1 strains that were evaluated [7], and so defines a major sublineage within this species. (Studies that would use deletion of RD711 as the single marker to define *M. africanum* West African-1 strains may therefore risk mis-categorizing some isolates.) Nonetheless, all the *M. africanum* West African-1 strains in the Ghana strain collection had RD711 deleted and, as part of another study [24], could be further subdivided phylogenetically based upon differences in mycobacterial tandem repeats numbers. Although not evaluated in this study, Mostowy *et al.* [23] recently reported that RD742 was also variably distributed among *M. africanum* West African-2 strains and a set of phylogenetically informative SNPs for *M. africanum*, different from those screened herein, has been published [51]. Overall, the combined data illustrate the continued evolutionary diversification of the *M. africanum* clades and advance the process of organizing a set of variable markers that may be used to construct meaningful phylogenetic trees for *M. africanum*. To this end, RD715 and RD743 were identified within *M. africanum* West African-1 strains [23] and single nucleotide changes located within the RD1 locus of *M. africanum* West African-2 strains were recently noted in select strains [38], but the utility of these polymorphisms as phylogenetic markers remains to be determined. It should also be mentioned that at least one *M. africanum*-like strain has been described with RD9 deleted, but RD7, RD10,
RD702, RD711, and RD713 intact [56]. Combined, these data indicate that there is greater M. africanum/MTC diversity yet to be characterized.

Our understanding of the nature of M. africanum as a species and its position within the MTC has evolved considerably in recent years. Based upon hard genome level sequence evidence, the name M. africanum subtype II is no longer applied [2,7,20,22,23], while strains denoted as M. africanum subtype I are now, ironically, recognized to constitute two relatively genetically distinct lineages emerging from separate nodes along the MTC evolutionary tree [1,7,25,26]. This opinion is reinforced by the data provided in the current report. Interestingly, the above mentioned unique M. africanum-like strain was isolated from a patient originating from the Democratic Republic of Congo, a central African country [56]. As it has been postulated that the MTC originated near the horn of Africa [57], this strain may therefore be a remnant M. africanum precursor that evolved from M. tuberculosis as humans migrated from Eastern to Western Africa [55]. Indeed, the M. africanum clades possess the phenotypic and genotypic characteristics of sequential intermediary genotypes in the evolution of M. bovis from M. tuberculosis [1,7,24,26]. In so being, there have been suggestions that an M. africanum transmission cycle may exist between humans and an unknown animal reservoir [23]. Reports of M. africanum isolation from a bovine source in Nigeria and from a goat in Guinea Bissau support this hypothesis [37,42]. Therefore, a study of animal MTC isolates employing genetic markers, such as those we have organized herein, should be made a priority effort to rule out M. africanum as an important source of zoonotic and/or anthropozoonotic TB in Western Africa.

Conclusions

With this study, we have organized a series of consistent phylogenetically-relevant markers for each of the distinct MTC lineages that share the M. africanum designation, highlighting those polymorphisms that can be used for specific clade identification. A review of molecular studies of M. africanum reveals a differential distribution of each M. africanum clade in Western Africa. Because M. africanum continues to be an important agent of disease, more M. africanum-focused studies are needed to increase our understanding of MTC pathology, epidemiology, and evolutionary history, all of which could lead to new strategies for TB prevention.

Acknowledgements

The authors thank the PDTIS DNA Sequencing Platform/FIOCRUZ for technical assistance. This work was supported financially by NIH grants R21 A0063147, and R21 A063147 (J.L.H) and Innovative Approaches for TB Control in Brazil (ICOHRTA), U2R TW006885 Fogarty International Center. SEVG was supported by ICOHRTA, as a trainee and CNPq, Brazil and Oswaldo Cruz Foundation (Rio de Janeiro, Brazil. SN was supported by the German Federal Ministry of Education and Research (BMBF) within the Gereman National Genome Research Network (NGFN1, Project D1G50162), and the PathoGenomikPlus Network (Project 0313801J).

Author details

1Laboratory of Molecular Biology Applied to Mycobacteria, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Avenida Brasil 4365, Manguinhos - 21040-900, Rio de Janeiro, Brazil. 2Clinical Microbiology Service and the Department of Pathology, New York-Presbyterian Hospital, Columbia University Medical Center, 622 West 168th Street, New York City, NY, USA. 3National Reference Center for Mycobacteria, Forschungszentrum, Parkallee 18, D-23845, Borstel, Germany. 4Mycobacteria Reference Laboratory, (Cib/LIS), National Institute for Public Health and the Environment, 3720 Bilthoven, the Netherlands. 5Division of International Medicine and Infectious Diseases, Department of Medicine, Joan and Sanford I. Weill Medical College of Cornell University, 1300 York Avenue, New York City, NY, USA.

Authors’ contributions

SEVG and RCH carried out the molecular genetic studies, participated in genotyping studies, analyzed the data and wrote the manuscript. SN: isolation and initial identification of the Ghana collection strains and provided suggestions during manuscript preparation. KP: isolation and identification of control strains and provided critical comments for the manuscript. ARS: provided methodological assistance and critical comments for the manuscript. RCH, PNS and JH: conceived the study and the methodology and supervised the various stages of the research. PNS and JH: coordinated the investigation and provided suggestions during manuscript preparation. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 1 December 2009 Accepted: 29 March 2010
Published: 29 March 2010

References

1. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglemeier K, Garnier T, Gutierrez C, Hewison G, Keter K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci USA 2002, 99(6):3684-3689.
2. Huard RC, Lazzarini LCO, Butler W, van Soolingen D, Ho JL. PCR-based method to differentiate the subspecies of the Mycobacterium tuberculosis complex on the basis of genomic deletions. J Clin Microbiol 2003, 41(4):1637-1650.
3. Mostowy S, Cousins D, Brinkman J, Aranaz A, Behr M. Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. J Infect Dis 2002, 186(1):74-80.
4. Aranaz A, Cousins D, Mateos A, Dominguez L. Elevation of Mycobacterium tuberculosis subsp. caprae Aranaz et al. 1999 to species rank as Mycobacterium caprae comb. nov. sp. nov. Int J Syst Evol Microbiol 2003, 53(8):1785-1789.
5. Cousins D, Bastida R, Catalfi A, Quae V, Redrobe S, Dow S, Duignan P, Murray A, Dupont C, Ahmed N, Collins D, Butler W, Dawson D, Rodriguez D, Loureiro J, Romano M, Alito A, Zamarraga M, Bernardelli A. Tuberculosis in seals caused by a novel member of the Mycobacterium tuberculosis complex: Mycobacterium pinnipedi sp. nov. Int J Syst Evol Microbiol 2003, 53(5):1305-1314.
6. Cousins D, Peet R, Gaynor W, Williams S, Gow B. Tuberculosis in imported hyrax (Procavia capensis) caused by an unusual variant belonging to the Mycobacterium tuberculosis complex. Ver Microbiol 1994, 42(2-3):135-145.
7. Huard RC, Fabre M, de Haas P, Lazzarini LCO, van Soolingen D, Cousins D, Ho JL. Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. J Bacteriol 2006, 188(12):4271-4287.
8. van Soolingen D, Hoogenboezem T, de Haas P, Hermans P, Koedam M, Teppema K, Brennan P, Beira G, Portella F, Top J, Schols L, van Embden J. A novel pathogenic taxon of the Mycobacterium tuberculosis complex,
Differentiation among members of the Mycobacterium tuberculosis complex by molecular and biochemical features: evidence for two pyrazinamide-resistant subtypes of M. bovis. J Clin Microbiol 2000, 38(1):152-157.

Goh K, Legrand E, Sola C, Rastogi N. Rapid differentiation of “Mycobacterium canetti” from other Mycobacterium tuberculosis complex organisms by PCR-restriction analysis of the hsp65 gene. J Clin Microbiol 2001, 39(10):3705-3708.

Parsons L, Brosch R, Cole S, Somoskovi A, Loder A, Bretzel G, Van Vasconcellos J. Molecular evolutionary history of tubercle bacilli assessed by study of the polymorphic nucleotide within the nitrate reductase (narG) operon promoter. J Clin Microbiol 2005, 43(8):4010-4014.

Constant P, Perez E, Malaga W, Laneelle MA, Saurel O, Daffe M, Guilhot C. Role of the pks15/pks9 genes in the biosynthesis of phenolglycolipids in the Mycobacterium tuberculosis complex - Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/pks9 gene. Journal of Biological Chemistry 2002, 277(4):3818-3818.

Haas W, Bretzel G, Arntzh B, Schilke K, Krommes G, Rüsch-Gerdes S. High genetic diversity among Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999, 37(8):2607-2618.

Allin-Béguel C, Harmen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of Mycobacterium tuberculosis complex isolates. J Clin Microbiol 2008, 46(8):2692-2699.

Owusu-Darbo E, Schwander S, Rüsch-Gerdes S. Discovery of a novel Mycobacterium tuberculosis lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. J Clin Microbiol 2007, 45(2):3891-3902.

Conde M, Kritski A, Belisle J, Kreiswirth B, Lapa e Silva J, Ho J. Application of sensitive and specific molecular methods to uncover global dissemination of the major RDRio Sublineage of the Mycobacterium tuberculosis complex. Int J Tuberc Lung Dis 2008, 12(7):798-803.

Watanabe K, Rastogi N, Belchev M, Haas W, Ackman T. Mycobacterial virulence and immune response: the quest for host-specific factors. Clin Microbiol Rev 2008, 21(4):1039-1067.
complex strains from patients with pulmonary tuberculosis in Cameroon. J Clin Microbiol 2003, 41(6):2547-2553.

42. Cadmus S, Palmer S, Okker M, Dale J, Gover K, Smith N, Jahans K, Hewinson R, Gordon S. Molecular analysis of human and bovine tubercle bacilli from a local setting in Nigeria. J Clin Microbiol 2006, 44(1):29-34.

43. Brudey K, Driscoll J, Rigouts L, Prodinger W, Gor J, AH-Hajj S, Allix C, Arintimiuo I, Arau J, Baumanis V, Binder L, Caffene P, Cataldi A, Cheong S, Diet R, Ellemier M, Evans J, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazola L, Gomes H, Gutierrez M, Hawkey P, van Helden P, Kadiwal G, Kreswin B, Krenner K, Kubin M, Kulkami S, Liens B, Lillebaek T, Ho M, Martin C, Mokrousov I, Narsvika Q, Ngowe Y, Naumann L, Niemann S, Parwati I, Rahn Z, Rasolofo-Razanamparany V, Rasololavonta T, Rossetti M, Risch-Gerdes S, Sajuda A, Samper S, Shemyakin I, Singh U, Somosovski A, Sluce R, van Soolingen D, Streicher E, Suffys P, Tortoli E, Traceviska T, Vincent V, Victor T, Warren R, Yap S, Zaman K, Pontaels F, Rastogi N, Sola C. Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol 2006, 6:23.

44. de Jong B, Hill P, Aken A, Jeffries D, Onipede A, Small P, Adenbola R, Corrah T. Clinical presentation and outcome of tuberculosis patients infected by M. africanum versus M. tuberculosis. Int J Tuberc Lung Dis 2007, 11(4):450-456.

45. Koivula T, Eksman T, Löffdahl S, Ghebremichael S, Mostowy S, Behr M, Svenson S, Källenius G. Genetic characterization of the Guinea-Bissau family of Mycobacterium tuberculosis complex strains. Microbes Infect 2004, 6(3):272-278.

46. Brudey K, Gutierrez M, Vincent V, Parsons L, Sallinger M, Rastogi N, Sola C. Mycobacterium africanum genotyping using novel spacer oligonucleotides in the direct repeat locus. J Clin Microbiol 2004, 42(11):5053-5057.

47. Niang M, de la Salmoniere Y, Samb A, Hane A, Cisse M, Gouquel B, Perrat R. Characterization of M. tuberculosis strains from west African patients by spoligotyping. Microbes Infect 1999, 1(14):1189-1192.

48. de Jong B, Antonio M, Awine T, Ogungbemi K, de Jong Y, Gagneux S, De Riemer K, Zoóto T, Rastogi N, Borgdorff M, Hill P, Adenbola R. Use of spoligotyping and large sequence polymorphisms to study the population structure of the Mycobacterium tuberculosis complex in a cohort study of consecutive smear-positive tuberculosis cases in The Gambia. J Clin Microbiol 2009, 47(4):994-1001.

49. Desmond E, Ahmed A, Probert W, Ely J, Jang Y, Sanders C, Lin S, Flood J. Mycobacterium africanum cases, California. Emerg Infect Dis 2004, 10(5):921-923.

50. Gutiérrez M, Galán J, Blázquez J, Bouvet E, Vincent V. Molecular markers demonstrate that the first described multidrug-resistant Mycobacterium bovis outbreak was due to Mycobacterium tuberculosis. J Clin Microbiol 1999, 37(4):971-975.

51. Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis and limitations the highlights of current methodologies. PLoS ONE 2009, 4(11):e7815.

52. Gagneux S, Small P. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis 2007, 7(8):528-37.

53. Filippou J, Motwalia A, Cavatore M, Qi W, Hazbón M, Bobadilla del Valle M, Fyfe J, García-García L, Rastogi N, Sola C, Zoóto T, Guerrero M, León C, Crabtree J, Angiuoli S, Eisenach K, Durma R, Joloba M, Rendón A, Sifuentes-Osorio J, Ponce de León A, Cave M, Fleischmann R, Whittam T, Alland D. Global phylogey 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 2006, 188(2):759-772.

54. Gutacker M, Smoot J, Migliaccio C, Ricklefs S, Hua S, Cousins D, Gravis E, Shashkina E, Kreiswirth B, Musser J. Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics 2002, 162(4):1533-1543.

55. Hershberg R, Lipatov M, Small PM, Shaffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol 2008, 6(12):e311.