Journal of Bacteriology and Parasitology

Freidlin, J Bacteriol Parasitol 2013, 4:3
DOI: 10.4172/2155-9597.1000169

Research Article Open Access

A New Paradigm for Genus *Mycobacterium* Population Structure

Paul Jeffrey Freidlin*

National Mycobacterium Reference Laboratory, National Public Health Laboratory, Ministry of Health, Tel-Aviv, Israel

Abstract

The genus *Mycobacterium* contains the pathogenic species *tuberculosis* and *leprae,* and additionally at least 148 sometimes opportunistic pathogenic species that occupy environmental niches. The genus *Mycobacterium* is in the order *Actinomycetes* along with other acid-fast genera with mycolic acids in their cell walls, such as *Corynebacterium*, *Nocardia*, and *Rhodococcus*. The DNA-dependent RNA polymerase subunit b gene *rpoB*, is relatively conserved among these bacteria. Sequence polymorphism in a 342bp fragment of *rpoB* is sufficient to differentiate among most species of *Mycobacterium*. However, construction of a robust and testable phylogeny or population structure based on this polymorphism, or polymorphism in other genomic regions, has remained elusive. This study presents a new paradigm for the resolution of genus *Mycobacterium* population structure. A minimum spanning tree (MST) was constructed on restriction enzyme site polymorphisms detected *in silico* for 5 enzymes, on the 342bp *rpoB* fragment obtained from GenBank data for 47 *Mycobacterium* species, and one species each from *Corynebacterium, Nocardia and Rhodococcus*. The MST was empirically divided into 3 regions. All the rapidly growing mycobacteria (RGM), with a halo of slowly growing mycobacteria (SGM), were found in MST-region 1. The correctness of the MST-regions model of genus *Mycobacterium* population structure was validated by statistically confirmed linkage disequilibrium of certain restriction site alleles. For certain alleles, SGM in MST-region 1 resembled the RGM of MST-region 1 more than the SGM of MST-region 3. The model provided a framework consistent with published genotypic and phenotypic observations, including expectations from comparative genomics and ribosomal RNA (rRNA) gene properties' distribution among species, and reported cladistic groupings of species. *Corynebacterium diphtheriae*, *Nocardia nova* and *Rhodococcus equi* belonged to MST-regions 2, 3, and 1 respectively. In conclusion, the MST-regions model of genus *Mycobacterium* population structure was robust, unambiguous, transparent for alleles, consistent with genotypes and phenotypes, and statistically testable.

Keywords: *Mycobacterium*; NTM; Tuberculosis; Population structure; *Actinomycetes*; *Corynebacterium diphtheriae*; *Nocardia nova*; *Rhodococcusequi*; *rpoB*; Restriction enzyme site alleles; Single nucleotide polymorphism; Minimum spanning tree; Minimal spanning tree; Rapidly growing mycobacteria; Slowly growing mycobacteria; Quantitative trait; Linkage disequilibrium; Molecular biology; Molecular genetics; Molecular microbiology; Species; Genus; Taxonomy; Genus *mycobacterium* population structure; Perl; JavaScript; Regular expressions; Genomics; Molecular epidemiology; Molecular diagnostics; Molecular taxonomy

Introduction

The genus *Mycobacterium* is comprised of over 150 parasitic and free living acid-fast species of bacteria [1]. It has been estimated that about a third of the world's population is latently infected with *Mycobacterium tuberculosis* [2], with active tuberculosis in 2011 causing an estimated 8.7 million new cases of TB and 1.4 million deaths [3]. Diagnoses of non-tuberculous *Mycobacterium* (NTM) associated diseases have been increasing [4-6]. Therapeutic modalities for the treatment of *Mycobacterium*-associated diseases are constantly being challenged by the development of multiple drug resistance [7-10] and in particular for *Mycobacterium tuberculosis*, the ability to evade (avoid or subvert) various components of the immune system [11-18]. The discovery of new therapeutics and vaccines is of paramount importance, but is dependent on deeper understanding of the structural and functional nature of the molecular biology of *Mycobacterium* species [19].

Despite spectacular advances in the ability to differentiate among *Mycobacterium* species and sub-species on the basis of molecular genetic characteristics [1,6,20-22], previous attempts to model *Mycobacterium* population structure have been only partially successful [23,24]. An often used test for the statistical correctness of phylogenetic tree branches is to check whether the branch appears in 95% or more bootstraps [25]. If even one branch appears in less than 95% of the bootstraps, the model is not "complete" because the topology of relationships among the branches will be statistically uncertain. However, even when using concatenated sequences [26], incorrect phylogenies can be obtained that are supported by significantly high bootstrap values [26]. Therefore, acceptable models of population structure must meet two sine-qua-non criteria: they must be biologically correct – consistent with "gold-standard" determinations of the taxonomic rank being modeled, and they must be transparent with respect to traceability of alleles in order to allow rigorous statistical testing of hypotheses [27]. Recent studies focus on the use of single nucleotide polymorphisms (SNPs) obtained from whole genome or multiple conserved loci (such as core or housekeeping genes), to differentiate and resolve phylogeny among *Mycobacterium tuberculosis* sub-species [28,29]. While use of SNPs is likely to lead to better inference of population structure than previous methods, the choice of informative SNPs is still important [30]. Restriction enzyme site polymorphism on the 342bp *rpoB* fragment, has been largely ascribed to generation of a synonymous SNP at any given restriction site [20,21]. Therefore, this type of *rpoB* fragment polymorphism can be considered a subset of SNP-generated polymorphism, and in this article it is employed to investigate genus *Mycobacterium* population structure via minimum spanning tree analysis.

It was realized early on that even one conserved gene, gene

***Corresponding author:** Paul Jeffrey Freidlin, National Mycobacterium Reference Laboratory, National Public Health Laboratory, Ministry of Health, Tel-Aviv, Israel, Tel: +972 3 5158687; Mobile: +972 50 6246668; Fax: +972 3 5185537; E-mail: paul.freidlin@phlta.health.gov.il

Received May 29, 2013; **Accepted** June 27, 2013; **Published** July 05, 2013

Citation: Freidlin PJ (2013) A New Paradigm for Genus *Mycobacterium* Population Structure. J Bacteriol Parasitol 4: 169. doi:10.4172/2155-9597.1000169

fragment, or corresponding polypeptide sequence, could be used as a "marker" for the quantitative trait associated with population structure and phylogeny at a given taxonomic level for example, cytochrome c [31]. The "gold standard" today for resolution of many bacterial species is sequence analysis of 16S ribosomal RNA (rRNA) or the DNA fragment coding for part of the 16S rRNA [32,33], although the 16S rRNA gene fragment is too conserved to resolve all the *Mycobacterium* species. The implicit assumption for the acceptance of DNA-marker-associated resolution of phylogeny, or population-structure (note that the criteria for phylogeny and population-structure are not the same – population-structure algorithms do not assume ancestor/descendent relationships), is that an allele of the marker is in linkage disequilibrium with the consensus-designated quantitative trait "species (or whatever taxonomic rank is being examined)" [27].

This study presents the first minimum spanning tree (MST) model of genus *Mycobacterium* population structure, a new paradigm. The model was constructed from restriction enzyme site alleles generated *in silico* by 4 restriction enzymes [21] on a 342bp *Mycobacterium rpoB* gene fragment [20,21], plus the unique invariant restriction site for the enzyme BclI. It was shown that the genus *Mycobacterium* has a highly polarized population structure, which can be empirically divided into three regions. The first region (MST-region 1) contained the compact group of rapidly growing mycobacteria (RGM) along with a halo of related slowly growing mycobacteria (SGM). The intermediate second region (MST-region 2) contained SGM with intermediate characteristics. The third region (MST-region 3) contained SGM, including *Mycobacterium tuberculosis* and *Mycobacterium leprae*, separated at the polar extreme from the RGM-containing MST-region 1. The model was testable, showing statistically significant linkage disequilibrium of certain restriction site alleles. The model was validated by its consistency with, and providing a framework for, observations such as the previously observed separation of RGM and most SGM, the appearance of previously observed clades each consisting of two or three species, and the distribution of certain SGM that have additional (more than one) ribosomal RNA operons into the RGM-containing MST-region 1. It is anticipated that this tested validated MST-regions model for genus *Mycobacterium* population structure will facilitate the development of testable hypotheses, possibly in conjunction with the new high throughput whole genome and epigenetic methylome profiling [34], which will advance our understanding of the genus *Mycobacterium*.

Materials and Methods

Sequence acquisition

55 reference sequences, each corresponding to a 306 bp (342bp with flanking primers) fragment of the *rpoB* gene from 52 different *Mycobacterium* species or 1 species from each of 3 related control genera [originally proposed, sequenced, characterized and submitted to GenBank by Kim et al. [20,21], were obtained from GenBank [35]. The GenBank ID of each sequence is shown in Table SI. It should be noted that although the separate sequence for GI 5902505 is still listed in GenBank, it is designated as being from *Mycobacterium celatum*, and not from *Mycobacterium celatum type II*. The extent to which these 55 reference sequences represented the world-wide spectrum of clinical and environmental species and subspecies of *Mycobacterium* and the 3 control genera, is considered in the Discussion section of this article.

Sequence Processing

The full-length forward 5` -> 3` 342 bases *rpoB* segment was constructed *in silico* [Application S2] by adding to the 306bases forward fragment, 1) the leading 5' 18 bases MF primer [20] and 2) the ending 3' 18 bases primer MR-complement [20]. ***Please note that Application S2 is copyright 2013 by Paul Jeffrey Freidlin,*** and is distributed as supplementary material accompanying this article under the terms of the Creative Commons Attribution License. Application S2 is offered as supplementary material to help the reader confirm the Application S2-generated results presented in this paper, and as a teaching device to introduce the reader to this type of computer program. There are, as of the submission date for this paper, no other validated uses for Application S2. The author (Paul Jeffrey Freidlin) recommends use of commercially or publicly distributed software, if available, for *in silico* restriction analysis of DNA sequences, a major advantage being the acquisition of already validated software. Use of the 342 bases long *rpoB* fragment allowed comparison of 5' -> 3' ordered restriction fragment lengths to previously published sequences and fragment lengths of the enzyme-restricted 342bp (306bp) *rpoB* fragment [20,21]. 5' -> 3' ordered restriction fragment lengths for *in silico* (PCR) restriction fragment length polymorphism analysis (PRA, or perhaps better IRA for "*in silico* restriction analysis") of the 342bp *rpoB* gene fragment [21], were generated *in silico* by Application S2, an in-house client-side Perl via JavaScript program. Perl regular expressions embedded in JavaScript objects [36] were used to identify and cut at restriction sites. The enzyme restriction sites were: BclI T^GATCA (6 cutter, type II restriction enzyme), AccII CG^CG (4 cutter, type II), HaeIII GG^CC (4 cutter, type II), HindIIGTY^RAC (6 cutter, type II), and MvaICC^WGG (5 cutter, type II). Information on restriction enzymes and specificities can be found in reference [37]. Nucleotide symbolism is according to the rules of the Nomenclature Committee of the International Union of Biochemistry [38]. Excel software (Microsoft Corporation) was used to generate the sequential 5' -> 3' ordered cumulative lengths of Application S2-generated 5' -> 3' ordered *rpoB* restriction fragments [Table S3], which corresponded to the restriction enzyme sites [Table S3]. The program in the form provided as supplementary material [Application S2] failed to cut properly at tandem HaeIII restriction sites (such as found in *celatum* type 2 and *kansasii* type II). Therefore, the program informed the user that a sequence contains tandem HaeIII restriction enzyme sites, and consequently the user must manually inspect and process the sequence for those sites. Note that in the current GenBank database, the strain with the ID shown in Table S1 is no longer known as *celatum* type 2, but its sequence still retains the tandem HaeIII sites.

Phylogenetic analysis

Restriction sites data for 5 enzymes for each of 47 reference sequences of *Mycobacterium* species and 3 reference sequences [20,21] of non-*Mycobacterium* species from related genera [Table S4] were analyzed using relevant programs of the PHYLIP package [39] (results not shown). Threeof the *Mycobacterium tuberculosis* complex species, *tuberculosis, bovis,* and *africanum* (the other species, *M. microtii, M. canetti*, and a new sixth species, were not examined) had identical in silico 342bp *rpoB* PRA profiles, and so were analyzed as one species. Similarly, *M. goodie* and *M. wolinskyi* [obtained from GenBank, but not studied by Kim et al. [20,21] had identical in silico 342bp *rpoB* PRA profiles, and so were analyzed as one species.

Population structure analysis

BioNumerics version 3.5 (Applied Maths, Belgium) was used under default settings [15] to construct a minimum spanning tree (MST, Figure 1) from the restriction site data (Table S3) on *Mycobacterium*

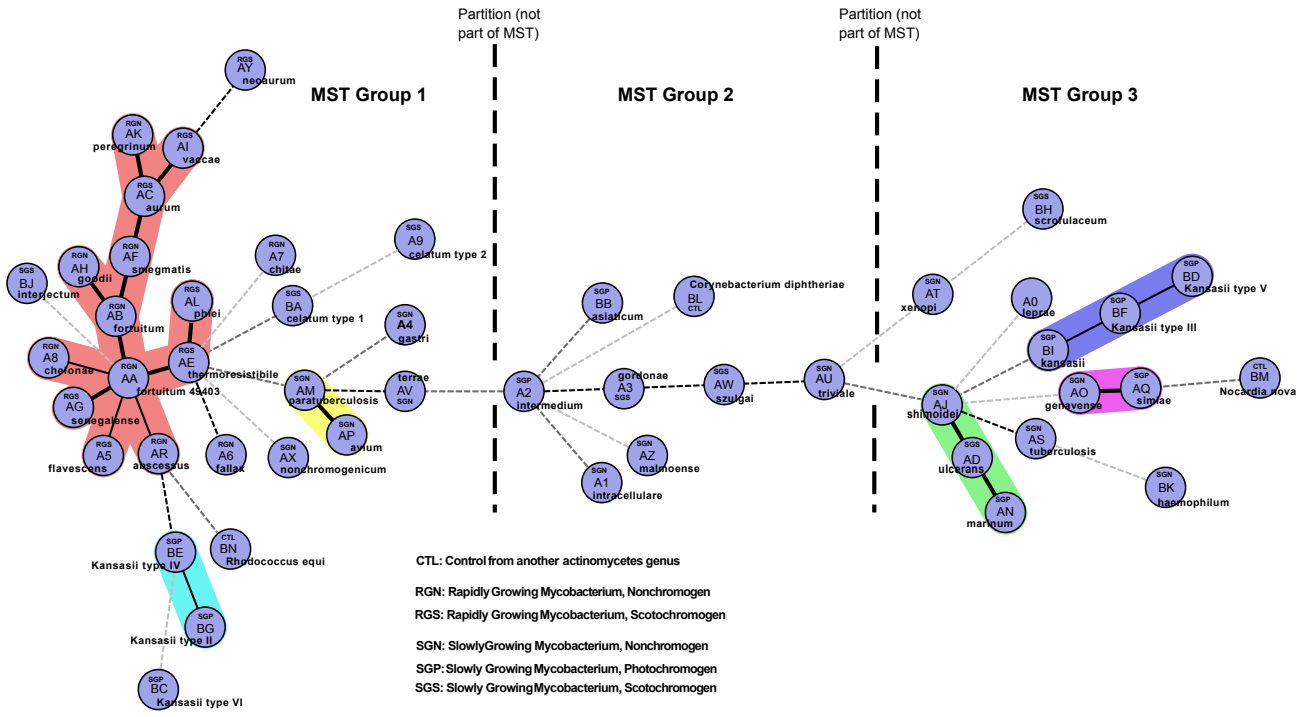

Figure 1: The minimum spanning tree (MST) model of genus *Mycobacterium* population structure.

species and control species from 3 related mycolic acid-containing genera. *M. tuberculosis* complex species were given one entry due to identical restriction sites on the given species, and likewise for *M. goodie* and *M. wolinskyi* which had identical restriction sites. An Excel database was constructed from MST data based on an empirical division of the minimum spanning tree into 3 regions (Figure 1, Table S5). Then the Excel database was summarized with pivot tables for an overall summary of the data (Table S6). Linkage disequilibrium of restriction site alleles with groups of *Mycobacterium* species appearing in different MST-regions (Table S7) was confirmed by 1 parametric and 2 non-parametric statistical method (Table S8). The parametric method was student's t for unpaired unequal samples (Excel statistical function, Microsoft). The non-parametric methods were: 1. The Fisher's Exact Mean [40] and 2. The Wilcoxon Mann-Whitney Test (U) [41,42]. Data taken from Table S6 also were used to graphically visualize with an Excel chart, the distribution of total restriction sites among the *Mycobacterium* species and controls (Table S9).

Results

Generation and labeling of rpoB fragment restriction sites from GenBank sequences

The 342bp *rpoB* fragment (306 bp plus 2 flanking PCR primers MF and MR complement of 18 bp each) is polymorphic resulting in unique restriction allele profiles for most *Mycobacterium* species(306 bp plus 2 PCR primers MF and MR of 18 bp each) [20,21]. The polymorphic sequences were accessed from GenBank for 47 unique reference sequencesfrom over 50 *Mycobacterium* species and an additional 3 sequences from 3 related species [20,21] (gene IDs are shown in Table S1). The *rpoB* restriction enzyme sites for enzymes *AccII, BclI, HaeIII, HindII, andMvaI* were determined with an in-house (written by the author) JavaScript program that utilized Perl regular expressions to detect the sites (Application S2). Table S1 presented the 5' ->3' ordered restriction fragment lengths obtained with Application S2 for each species for each enzyme. Table S3 concatenated the sites found on the 47 unique *Mycobacterium* reference sequences and reference sequences from 3 related species, *Corynebacterium diphtheriae*, *Nocardia nova* and *Rhodococcusequi*. In total for all species, 30 different restriction site alleles for 5 different restriction enzymes were found. A binary profile (0: site not present, 1: site present) of 30 digits was constructed for each species (Table S3) and presented in a form (Table S4) that could be analyzed with relevant programs in the PHYLIP suite of programs [39]. The phylogenetic trees are not shown, because they resembled results previously published [20,24,43] and added no new information. The species were distinctly separated one from the other, rapidly growing mycobacteria (RGM) were separated as a group from most slowly growing mycobacteria (SGM), certain species tended to cluster together in clades, and the technique could not differentiate among the species of the *Mycobacterium tuberculosis* complex because they had identical 342bp *rpoB* fragment sequences and consequently identical restriction sites. Similarly, *Mycobacterium goodie* could not be differentiated from *Mycobacterium wolinskyi* due to identity between their 342bp *rpoB* fragment sequences. However, a different region in the complete *rpoB* gene sequence, or use of the *hsp65* fragment, does allow the differentiation of *M. goodie* from *M. wolinskyi* [1,44-46].

Construction and partitioning of the minimum spanning tree

A minimum spanning tree (MST) was constructed from the restriction sites using Bionumerics version 3.5 (Figure 1). The 30 digit binary profiles (Table S3) were entered as character data into Bionumerics, and analyzed using the default settings [15] for generation of a minimum spanning tree. The MST was manually (empirically) partitioned into three regions, MST-region 1, MST- region 2, and MST-region 3 (Figure 1).

Characterization of the minimum spanning tree

The genus *Mycobacterium* was for the first time revealed to have a diffuse and highly polarized population structure that partitioned into 3 MST-regions.

The restriction enzyme data were categorized in an Excel database (Table S5) along with *Mycobacterium* species, MST region, growth rate and pigmentation [20,47-49], and analyzed with appropriate pivot tables (Table S6). *M. szulgai* although designated as a scotochromogen, could also be a photochromogen, depending on conditions of growth [48]. The data were analyzed for linkage disequilibrium of restriction site alleles (Tables S7 and S8). In this manner the concepts "species" and taxonomic ranks in general, were treated as quantitative traits which could be resolved by comparison of whole DNA or protein sequences, sequence fragments, or DNA markers such as restriction site alleles [27]. Table S7 summarizes the linkage disequilibrium ofrestriction site alleles ofslowly (SGM) and rapidly (RGM) growing *Mycobacterium* strains with *rpoB*-based MST-regions 1, 2 and 3. The restriction enzyme site alleles AccII 79, HaeIII 262 and HaeIII 263 of *Mycobacterium* species were in linkage disequilibrium with respect to the SGM distribution among MST regions, such that the SGM in MST region 1 resembled the RGM of MST region 1 significantly more than they resembled the SGM found in MST region 3 (Table S8). The presence of site HaeIII 61 in the 5' -> 3' 342bp *rpoB* fragment was necessary and sufficient to identify a *Mycobacterium* species as belonging to the RGM. That is, it was a unique identifying restriction site allele for the RGM. The presence of site HindII 232 was necessary and sufficient to identify *Mycobacterium tuberculosis* complex members examined in this study (*tuberculosis, bovis, bovis BCG*, and *africanum*). Kim et al. [21] note that PRA of the 342bp *rpoB* fragment generates unique fragments that identify RGMand also *Mycobacterium tuberculosis* complex members.

The suspected linkage disequilibrium of restriction enzyme site alleles was confirmed by 3 different statistical methods (Table S8). That is, the particular polar distribution of certain *rpoB* restriction site alleles supported the validity of the MST partitioning of *Mycobacterium* species into MST-regions 1, 2 and 3.

Species from 3 *Mycobacterium*-related mycolic acid-containing genera (*Rhodococcusequi*, *Corynebacteriumdiphtheriae*, and *Nocardia nova*) were used as controls. The rpob restriction sites associated each control to its own separate *Mycobacterium* MST-region (Figure 1). Intriguingly, *R. equi* which associated with the RGM-containing MST-region 1, also had the RGM-specific (that is, in the genus Mycobacterium, only RGM have this allele) restriction site allele (Table S6). *C. diphtheria* associated with MST region 2, and *N. nova* associated with region 3. Four of the 30 characterized restriction sites were found exclusively in *C. diphtheria*, and one exclusively in *N. nova* and thus have the potential to be used as unique identifiers differentiating these species from *Mycobacterium* species. Finally, restriction enzyme BclI cut at only one site, but this site was invariant, that is, it was found in the same place in all *Mycobacterium* species and in all three control genera; thus it could mark an especially important region of the *rpoB* gene. Surprisingly, 3 *kansasii* species (II, IV, and VI) associated with the RGM in MST-region 1, while the other 3 *kansasii* species (I, III, and V) associated with the SGM in MST-region 3. Each of the 3 pigmentation types, nonchromogen, scotochromogen, and photochromogen were found in each of the MST-regions 1, 2, and 3, although the RGM did not include any photochromogens. All reference species were successfully differentiated, indicating sufficient number and polymorphism of the 342bp *rpoB* fragment restriction sites. In addition, where reference strain sequences were included of known subspecies – for *fortuitum, celatum and kansasii*– the subspecies were successfully differentiated, and subspecies for each respective species were found clustered together. *Kansasii* showed an intriguing distribution into 2 clusters of 3 subspecies each, found in MST-regions 1 and 3. The species most closely related were indicated by thicker, darker, more solid lines and shading joining them, and were consistent with taxonomic clades observed by previous investigators using a variety of molecular methods, as detailed in the Discussion that follows.

Figure S9 graphically presented the number of *rpoB* fragment restriction sites found in each *Mycobacterium* species and in the controls. The number of restriction sites ranged from 3 in *M. kansasii* to 12 in *M. fallax*.

Discussion

Corroborating evidence for the validity of the MST-regions model of genus Mycobacterium population structure

Use of the 342bp rpoB fragment: *rpoB* sequence analysis as a novel basis for bacterial identification was published in 1997 [50], and applied to the identification of *Mycobacterium* species in work by Kim et al. [20,21]. The dedicated Perl regular expressions via JavaScript application program used for this study (Application S2), eliminated much of the error-causing tedium in finding *in silico* restriction fragments manually or using multitask programs. Thus an error-prone task became essentially error-free, an important benefit for the clinical laboratory that wishes to introduce routine *in silico* restriction enzyme site analysis of sequenced DNA fragments. Nevertheless, commercial programs adequate to the task have the important quality of pre-validation for the task, and thus commercial programs are to be preferred over in-house software, especially if they have been specifically designed to meet the laboratory's needs. It should be noted that although the BclI restriction site was the same for all species and subspecies examined (Table S6), it was informative for this study in that it provided the cross-species similarity necessary for the construction of phylogenetic trees (not presented), and the minimum spanning tree. Also, sites that are so stringently conserved among species and even across the genus barrier, point to areas of the gene that tolerate little interference, and thus are possible wide-spectrum therapeutic targets. Kim et al. [20,21] reports that restriction enzyme site polymorphisms (for the 4 enzymes they investigated) of the 342bp *rpoB* gene fragment, are due to synonymous single nucleotide changes (synonymous SNPs). The present study is the first published presentation of the corresponding restriction fragments in their 5' to 3' order on the MF primed strand. A 5' -> 3' ordered presentation of fragment lengths was necessary to allow the calculation of restriction site positions (Tables S1 and S3). Furthermore, it facilitated comparisons for identity between restriction fragment length polymorphism profiles from different species, thereby avoiding ambiguity by restriction fragment comparisons by size only, or ambiguity encountered when using percentage-similarity cutoffs in sequence homology comparisons [1]. When the Application S2

restriction fragment lengths were totaled for the products generated by each restriction enzyme, the total length was always 342 bases (except for the previously mentioned 2 species that exhibited tandem HaeIII sites, which had to be evaluated manually) (Table S1). Although other *rpoB* regions show polymorphism [45,51,52], the 342bp region used in this study [20,21] was attractive because it was well characterized for differential polymorphism among more than 44 *Mycobacterium* species for 4 restriction enzymes [21], and for the 3 control species *Rhodococcusequi, Corynebacterium diphtheria,* and *Nocardia nova*, increasing the chance that enough informative restriction sites would be generated for analysis. Possible drawbacks of use of the 342bp *rpoB* fragment included: 1) possible heterogeneity of the sequence for a given species (the existence of subspecies) [53]; however, the fragment appears to be relatively well conserved within species [20,21] and useful for the present study (further comments on this follow in the next section), 2) possible ambiguity for a species due to lateral gene transfer [54]; however, horizontal gene transfer in this fragment appears to be an exceedingly rare phenomenon, and 3) the presence of a *rif* rifampicin resistance hot-spot within the fragment [20]; however, this mainly complicates the identification of rifampicin-resistant tuberculosis complex members [8], and even many of these can be successfully identified [20,21].

Implications and limitations associated with the number and type of sequences which each reference sequence represents, and the number of reference sequences (species) used for the construction of the MST-regions model

Kim et al. [20,21] examined about 300 clinical isolates, of which about half were NTM representing 9 species and sub-species. Concerning their clinical isolates, they had 8 or more isolates for each of the 8 species they identified, and they reported little variation among the sequences representing each of their clinical species. In their articles, Kim et al. [20,21] show that their 306 (342) bp fragment can be used to differentiate the sequences obtained from known reference subspecies, such as those comprising the species *fortuitum, celatum*, and *kansasii.* The MST-regions model presented in this article also succeeded in differentiating these subspecies, while showing the close relationship among subspecies (*kansasii* being a special case to be discussed later). Kim et al. [20,21] did not publish an unambiguous minimum spanning tree model of genus *Mycobacterium* population structure based on restriction site alleles, as presented in this article. They also did not give a 5'->3' ordered sequence of restriction fragments generated by the restriction enzymes they used. The minimum spanning tree (MST) published in this article is the first such published MST tree based on *rpoB* 342bp fragment restriction site alleles, and the first such tree published constructed on unambiguous alleles, traceable and statistically testable for linkage disequilibrium. Kim et al. [20,21] use 4 restriction enzymes on the 342bp *rpoB* fragment, to successfully differentiate among over 44 different species of *Mycobacterium* and 3 control genera. This study employed the same 342bp *rpoB* sequences (with the addition of *goodie* and *wolinskyi*) and the same 3 control genera sequences that Kim et al. [20, 21] submitted to GenBank. Also, this article employed the same 4 restriction enzymes with 1 additional 5th restriction enzyme. Thus the hard work Kim et al. [20,21] invested in showing that their reference sequences were representative of the respective clinically identified 8 species (that is, that at least 8 sequences were examined for each of 8 clinically identified species, except *kansasiisubsp II* which had only 2 isolates), pre-validates the use of these same reference sequences for this article. As stated in this article, the *goodii* and *wolinskyi* species added to the set of species previously analyzed by Kim et al [20,21], could not be differentiated from each other by the *rpoB* 342bp fragment. The addition in this article of 1 invariant restriction enzyme site (generated by a fifth restriction enzyme) was necessitated by the limited amount of loci (30 restriction sites), which would otherwise have led to strains having no loci in common and consequent unsuitability for genetic comparisons (for example, by the minimum spanning tree algorithm).

Itoh et al. [53] used the 342bp *rpo B* sequence (employed in this article) to divide 34 Japanese *M. gordonae* isolates by sequence homology or PRA into 4 clusters representing the 4 known *gordonae* subspecies. It is not known which of the *gordonae* subspecies is represented by the reference sequence employed by Kim et al. [20,21], and this in addition to the fact that only 1 *gordonae* subspecies was analyzed, is a limitation of the MST-regions model presented in this article.

Kim et al. [20,21], whose GenBank sequences are employed in the article, received their *kansasii* subspecies *I-V* reference strains from V. Vincent, who had recently participated in an extensive study of *kansasii* isolates in France [55]. She and coworkers studied 62 *M. kansasii* isolates, including 38 clinical strains and 24 strains from water samples. Both the clinical and environmental isolates yielded 5 subspecies by a variety of molecular methods. Each of the 5 subspecies was comprised of at least 4 isolates, and there was little intra-subspecies variation among the isolates. Thus, it can be assumed that the ability of PCR restriction analysis (PRA) of the 342bp rpoB fragment, to correctly detect each of the 5 subspecies [21] validates the use of this technique, at least for the range of *kansasii* isolates and subspecies seen in France. Another PRA study of 276 *M.kansasii* isolates from various geographic areas within Europe [56], identified the same 5 *kansasii* subspecies as seen by Picardeau et al. [55], thus showing extensive geographic spread of the 5 subspecies. On the other hand, Zhang et al. [56], found by PRA that 78 of 81 clinical *kansasii* isolates in the US belonged to subspecies I.

The current article uses *Mycobacterium* species' reference sequences validated for subspecies appearance in clinical isolates, by the original investigators of 342bp rpoB fragment usage [21]. These clinically isolated species represent the major clinically important *Mycobacterium* species, and thus justify publication of the MST model as being of likely relevance to the relationship among clinically important *Mycobacterium* species. Truly comprehensive coverage of *Mycobacterium* species by the MST-regions model constructed on 342bp *rpoB* fragment restriction sites, would involve a major international effort to collect worldwide clinical isolates, and environmental samples and sequences from representative ecological niches from diverse geographical locations. The current MST model is a step in the right direction, but much work remains to be done.

Concerning the need for more sequences representing the control genera (*Rhodococcus, Corynebacterium,* and *Nocardia*), the original investigators Kim et al. [20,21] used only one reference species' sequence from each control genus. Other investigators have used the same control species, or other species from the same control genera. The MST-regions model employs analysis of the same control reference sequences used by Kim et al. [20,21], but because only one sequence was used per control genus, these results must be considered preliminary, but nevertheless supportive of the results obtained by the original investigators. Again, the MST-regions model provides a testable framework which is a step

in the right direction, but much work outside the scope of the present study remains to be done to clarify the *rpoB*-associated relationships of the 3 control genera, to the groups of *Mycobacterium* species found within each MST region.

The MST-regions model is supported by linkage disequilibrium of restriction site alleles.

As shown in the results, and summarized in Tables S7 and S8, there was statistically significant linkage disequilibrium of certain *rpoB*restriction alleles with MST-regions. The results indicated that the SGM in MST-region 1 more closely resembled, for certain alleles, the RGM in MST-region 1 than the SGM in MST-region 3. This is the first time genetic support has been offered for the polar distribution of the SGM, wherein certain SGM were closely associated with the RGM, and clearly differentiated from SGM at the extreme polar opposite (that is, SGM in MST-region 3). The SGM in MST-region 2 were intermediate in behavior with respect to restriction site allele distribution.

The MST-regions model showed close association of species previously observed to be closely associated in clades of 2 or 3 species.

The RGM had a unique restriction site allele, and the *tuberculosis* complex strains (*tuberculosis, bovis, bovis BCG, and africanum*) also exhibited a unique allele [Tables 7 and 8], as previously observed [20,21]. Many previously observed *Mycobacterium* clades also appeared as closely associated species in the MST-regions model (Figure 1), for example: *neoaurum* and*vaccae* [24], *fortuitum* and *smegmatis* [1,24], *fortuitum* and *fortuitum 49403* [20], *abscessus* and *chelonae* [20,24], *marinum* and *ulcerans*[1,20,24], *avium* and *paratuberculosis* [1,20,24], *smegmatis* and *goodii*[1], *aurum* and *vaccae*, *haemophilum* and *tuberculosis*, *leprae* and *tuberculosis* [1,20], *gordonae* and *asiaticum* [1], *genavense* and *simiae*, *shimoidei* and *marinum*, *celatum type 1* and *2*, *asiaticum* and *intermedium*, intermedium and *gordonae*, *intermedium* and *intracellulare*, *gordonae* and *szulgai* [20].

The MST-regions model provided a framework for genetic elements (and corresponding phenotypes) contributing to the quantitative traits responsible for the division of genus Mycobacterium into rapidly and slowly growing mycobacteria, and possibly contributing in a cumulative way to the distribution of Mycobacterium species among MST-regions 1, 2, and 3.

Published results concerning the rRNA operons, promoters, promoter strengths, and 16S rRNA helix 18 length [24, 42,57-69], are consistent with, and thus support, the observed distribution of RGM and SGM in the MST-regions model for population structure. The published results are, by MST-region, growth rate and species (number of promoters, number of operons, 16S rRNA helix 18 type); **for MST-region 1**, **RGM**: *chelonae* (5,1,short), *abscessus* (5,1,short), *fortuitum* (5,2,short), smegmatis (4,2,short), *phlei* (4,2,short), *neoaurum* (4,2,short), *senegalense* (?,2,short), *peregrinum* (?,2,short), *wolinskyi* (?,2,short), *flavescens* (?,?,short), *thermoresistibile*(?,?,short), *fallax* (?,?,short), *aurum* (?,?,short), *chitae*(?,?,short); for **MST-region 1, SGM**: *celatum* (3,2,long), *terrae* (?,2,long), *avium* (?,1,long), *paratuberculosis* (?,1,long), *gastri* (?,?,long) *interjectum* (?,?,short), *nonchromogenicum*(?,?,long); for **MST-region 2, SGM**: *intermedium* (?,?,short), malmoense (?,?,long), gordonae (?,?,long), *intracellulare* (?,1,long), *triviale* (?,?,short), *asiaticum* (?,?,long), *szulgai* (?,?,long); for **MST-region 3, SGM**: *tuberculosis* (2,1,long), *leprae*(?,1,long), *bovis* (?,1,long), *ulcerans* (?,1,?), *marinum* (?,1,long), *simiae*(3,1,short), *genavense* (?,?,short), scrofulaceum (?,?,long), xenopi (?,?,long). It was remarkable and highly supportive of the MST-regions model, that RGM *Mycobacterium chitae*, which falls "outside" the close 16S rRNA similarity values shared by other members of the genus *Mycobacterium* [68], was also "outside" the compact cluster (though still relatively weakly connected) of RGM in MST-region 1 (Figure 1).

All mycobacteria possess at least 1 rRNA operon, with at least 2 rRNA promoters, and the rRNA promoters are of variable strength [67]. As seen in Figure 1 and summarized above, based on the MST-region, testable predictions could be made about the probable status of RGM and SGM with respect to number of operons, number of promoters, strength of promoters and length of 16S rRNA helix 18. RGM, which were all found in MST-region 1, could be predicted to have at least 4 promoters, at least one of which would be "strong," 1 or 2 rRNA operons, and a short 16S rRNA helix 18. The SGM of MST-region 1 could be predicted to have no more than 3 promoters, 1 or 2 rRNA operons, and a long 16S rRNA helix 18 except for *M. interjectum* which had a short helix 18. It is interesting that the MST-regions model (Figure 1) separated *M. interjectum* from the other SGM of MST-region 1. Furthermore, it was highly supportive of the model that the SGM *M. celatum andM. terrae*, each with 2 rRNA operons, were assigned to the halo of SGM accompanying the RGM in MST-region 1.

The SGM of MST-region 2 were not characterized enough for predictions, except with respect to type of 16S rRNA helix 18 which could be long or short. It was interesting and perhaps significant that both *intermedium* and *triviale* were located at junctions between the MST-regions (Figure 1), and both had short 16S rRNA helix 18 lengths. The SGM of MST-region 3 could be predicted to have no more than 3 promoters, only 1 operon, and long or short 16S rRNA helix 18 length. It was remarkable that the MST-regions model of *Mycobacterium* population structure recognized that even though the RGM *M. chelonae* and *M. abscessus* each had only 1 rRNA operon, they belonged in MST-region 1 in a tight cluster with the RGM.

Rhodococcus equi joined MST-region 1 via *Mycobacterium abscessus* of the RGM. *Corynebacterium diphtheria* joined MST-region 2 via *Mycobacterium intermedium* of the SGM. *Nocardia nova* joined MST-region 3 via *Mycobacterium simiae* of the SGM. The 3 "connecter" *Mycobacterium* species, *abscessus, intermedium, and simiae*, all had short 16S rRNA helix 18 lengths.

The 6 different SGM *Mycobacterium kansasii* species [21,55,56,70] should provide a particularly interesting test of the MST-regions model for *Mycobacterium* population structure. *Mycobacterium kansasii* species *I, III, and V* were found in MST-region 3 along with pathogenic SGM like *Mycobacterium tuberculosis* and *Mycobacterium leprae* (Figure 1). In contrast, *Mycobacterium kansasii* species *II, IV, and VI* were found in MST-region 1 along with the halo of SGM accompanying the RGM. The *M. kansasii* species in MST-region 1 were separated from the *M. kansasii* species of MST-region 3, in accordance with the linkage disequilibrium of certain *rpoB* restriction enzyme site alleles (Tables S7 and S8). It would be interesting to know whether these site alleles' genetic differences were paralleled with other genetic and phenotypic differences, for example, number of rRNA promoters, operons, or length of the 16S rRNA helix 18. Genomic comparisons should be especially informative, particularly if *M. kansasii* sub-species members within their MST-region resemble each other significantly more than they resemble those subspecies members found in the other MST-region (that is, comparing *M. kansasii* subspecies within and between MST-regions).

The MST-regions model of Mycobacterium population structure provided a framework for the differential association of MST-region associated Mycobacterium species with related genera.

As already noted, *Rhodococcus equi* was associated with MST-region 1 RGM (Figure 1). This aspect of the MST-regions model is supported by recent findings that biodegradation (and associated genes) of high molecular weight polycyclic aromatic hydrocarbons (PAHs) is only observed in a few genera, including the RGM of genus *Mycobacterium*, and *Rhodococcus* [71]. It would be very interesting now to check the SGM in MST-region 1 for the ability to biodegrade PAHs, or the presence of the necessary genomic regions.

The "gold-standard" phylogenetic evidence connecting *Corynebacterium*, *Mycobacterium* and *Nocardia*genera as a subgroup within the bacterial phylum *Actinobacteria*, are comparisons of 16S rRNA gene sequences. These 16S rRNA sequence comparisons are supplemented with information from comparative genomics [72,73]. Based on the 16S rRNA gene sequence comparisons, *Rhodococcus* is a separate, but closely related genus, with respect to the genus *Nocardia* [74]. Additional recently published comparative genome studies reinforce the close relationship among *Corynebacterium, Mycobacterium* and *Nocardia* [75,76]. Therefore, it was expected that the *rpoB* gene, including its 342bp fragment, would share certain restriction site alleles in common with all 3 genera, while other alleles would differentiate the control genera (the control species) from *Mycobacterium* species (Table S6).

Gao et al. [72] study reports 13 proteins that are specific for the *Corynebacterineae* or the *Corynebacterium-Mycobacterium-Nocardia* subgroup. This study refers to the presence of mycolic acids as a defining feature of the *Corynebacterium-Mycobacterium-Nocardia* subgroup, and mycolic acid analysis is one of the original methods for identifying members of the genus *Mycobacterium* [77]. Thus it is interesting that 3 of the 13 proteins play an important role in the biosynthesis of the cell envelope and their products are the sites of resistance to the first-line anti-tuberculosis drug ethambutol (EMB). The same study of Gao et al. [72] found 14 proteins shared only by *Mycobacterium* and *Nocardia*, and specifically not shared by *Corynebacterium*. Among these 14 proteins were certain PE and PPE proteins. Although not mentioned as one of the 14 proteins, PPE34 (gene RV1917c) would warrant a closer look, as comparisons were based on the reduced genome of *M. leprae*, which may have lacked this protein. PPE34 is apparently being subject to host "pressure" in several African countries [15] and countries in Asia [14], and has been implicated in manipulation of immunological responses [16]. Another interesting finding is that in *M. bovis* BCG, PPE34 mutants exhibit reduced fitness at slow growth rate in a carbon limited chemostat [78].

Cholesterol acquisition is required for bacterial persistence during the chronic stage of *Mycobacterium tuberculosis* infection in mice [79], and genes involved in cholesterol utilization comprise a major part of the gene set required for bacterial survival in animal models of tuberculosis [79]. An interesting question is do *Mycobacterium* species in different MST regions have correspondingly different genotypic and phenotypic patterns associated with cholesterol acquisition and utilization?

Comparative genomics of the dormancy regulons in *Mycobacterium* and related genera, reveals that the genera *Mycobacterium* (with one exception, *M. leprae*), *Nocardia, and Rhodococcus*contain these regulons [80]. *M. leprae*, which has lost a considerable amount of its genome, does not have these regulons. Also, the genus *Corynebacterium* does not have these dormancy regulons [80]. Nevertheless, both *M. leprae* and *Corynebacterium diptheriae* have their different places in the MST-regions model.

In summary, from 16S rRNA gene sequence comparisons, and from genomic comparisons, there is convincing evidence that the genera *Corynebacterium, Mycobacterium, Nocardia* and *Rhodococcus* are distinct, but closely related, and it is logical that a conserved gene fragment such as the 342bp *rpoB* gene segment would contain restriction site alleles shared among these genera. The significance of the observations (Figure 1) that *Corynebacterium diptheriae* joined at MST-region 2, *Nocardia nova* joined at MST-region 3, and *Rhodococcus equi* joined at MST-region 1, remains to be determined. However, as already mentioned, *R. equi* shares with RGM the ability to biodegrade high molecular weight PAHs [71], and so joining MST-region 1 via the RGM was "reasonable".

The origin of restriction site allele polymorphism

Most restriction site allele polymorphism in the 342bp *rpoB* fragment is attributed to random synonymous single nucleotide polymorphism within the respective restriction site [20,21]. Since restriction sites are mostly palindromes with potential regulatory functions, including the potential to be methylated, the number and location of these sites may have physiological import, even within this short fragment, and even though the change was synonymous leaving the protein amino acid primary structure intact [81,82]. Therefore, it cannot be ruled out that rather than being passive DNA markers of the quantitative trait "genus *Mycobacterium* species identity," the *rpoB* restriction site alleles themselves may have contributed to that trait, and been subject to the same "pressures" that pushed the development of the new species.

Methylation events could affect *rpoB* DNA or RNA directly or indirectly. The effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases has been reviewed [83]. A more recent study describes a modification-dependent endonuclease in *Mycobacterium* that is a member of the Mrr family, and that may have endonuclease activity toward methylated DNA, that is, may restrict certain epigenetic modifications [84]. Bacterial host-mediated modification of PvuII restriction in *Mycobacterium tuberculosis* is partly responsible for IS6110 restriction fragment length polymorphism [85].

Methylated modified nucleosides in tRNAs, including tRNAs of the genus *Mycobacterium*, play important roles in the structure and function of tRNAs [86]. Mutations in 16Sand 23S rRNAmethylases in *Mycobacterium tuberculosis*confer resistance to capreomycin, viomycin, and streptomycin [87,88]. In summary, methylation changes (loss or gain) to RNA in the genus *Mycobacterium* can affect the physiology of the bacterium, possibly including transcription and translation for the *rpoB* subunit, and can be selected for by various "pressures" including antibiotic therapy.

Horizontal (aka lateral) gene transfer (HGT) contributing to the genus Mycobacterium genome

Comparative genomic and phylogenetic approaches define a core and pan genome for the genus *Mycobacterium* [89,90]. Studies find little recombination in core regions (which include the *rpoB* gene), whereas recombination outside these regions indicates significant acquisition of genes from non-mycobacterial sources [89,91], or between *Mycobacterium* strains [92]. Various vehicles are described for the acquisition of non-mycobacterial genes, or the HGT between mycobacteria. The general findings include HGT between mycobacteria mediated by a conjugative plasmid [92], HGT between different strains of the same *Mycobacterium* species facilitated by mycobacterial biofilms

[93], HGT in free living protozoa with and between the *Mycobacterium* species they host [94], and HGT via mycobacteriophages [95].

Focus on Mycobacterium tuberculosis

There are several reports of extensive HGT in the ancient progenitor of *Mycobacterium tuberculosis* [96-98], before its change into the *Mycobacterium tuberculosis* complex [99,100]. Gene reorganization through gene rearrangements as revealed through comparative genomics is suggested as an important contributor to speciation within the genus *Mycobacterium*, and the absence of rearrangements in the *Mycobacterium tuberculosis* complex (MTBC) supports consideration of the complex members as a single genospecies [101]. This is consistent with the observation that members of the MBTC examined in this study had identical 342bp *rpoB* restriction site alleles (Table S1) and thus an identical position in the MST-regions model of *Mycobacterium* population structure (Figure 1). A recent study finds that rpoB is a core gene not only for the genus *Mycobacterium*, but for the whole subclass *Actinobacteridae* [102]. Other recent studies (beyond the scope of this article) extend the use of *rpoB* to genetic analysis of microorganisms outside the subclass *Actinobacteridae*. Snyder and Champness [103] review various strategies used by bacteria to exchange genetic material between species, and even between kingdoms. Also, various mechanisms are considered for the internal generation of genetic polymorphism within a bacterium [103]. The *rpoB* gene helps generate part of the translation apparatus, the most highly conserved of all the cellular components [103,104], and thus while the large amount of information related to genus *Mycobacterium* population structure in this one small 342bp *rpoB* gene fragment was remarkable, its informative nature was also reasonable [105,106].

Perhaps the ancient event in the *Mycobacterium tuberculosis* progenitor, of HGT acquisition of the DR region [61,107], a clustered, regularly interspaced, short palindromic repeat (CRISPR) region (on which the spoligotyping test [14] is based) helped achieve the relative stability of the MTBC genome, and helps explain the absence of rearrangements in the *Mycobacterium tuberculosis* complex (MTBC) [101]. CRISPR loci provide sequence-directed immunity against phages and plasmids thereby limiting HGT [108,109]. This could help explain and support the identical position showed by MTBC members in the MST-regions model of *Mycobacterium* population structure. One recent study by Namouchi et al. [110] finds that comparative genomic examination of 24 *Mycobacterium tuberculosis* genomes reveals that despite the relative stability observed in the MTBC members, *Mycobacterium tuberculosiss* is rapidly diversifying by the actions of mutation, recombination, and natural selection. Thus the MST-regions model of *Mycobacterium* population structure, even for *Mycobacterium tuberculosis* complex members, is possibly subject to change as species change their characteristics, and new species emerge [33,111].

Concluding remarks

Pigmentation was one of the original phenotypic characteristics used as an aid to differentiate and clarify *Mycobacterium* species [7,43,69,77] and is still an important diagnostic attribute. However, the pigmentation trait's "molecular clock (rate of genetic change)" is not similar enough to the molecular clock of the quantitative trait "prokaryotic species" to be employed as a sole criterion to differentiate among *Mycobacterium* species and visualize the genus *Mycobacterium* population structure for relationships among *Mycobacterium* species. This is clearly seen in Figure 1 (the MST model) of the article, where each strain is labeled with its growth and pigmentation characteristics. This problem has been noted for phenotypic characteristics in general [43] and was a major impetus for the development of molecular techniques for diagnostic, epidemiologic, and taxonomic purposes.

What is the "molecular clock" of the quantitative trait "prokaryotic speciation?"Since the definition of prokaryotic species is still developing, currently being stretched by comparisons of newly accessible sequences of whole genomes [33,101,105,106], this question has no answer yet. A major issue for discussion is whether a definition of prokaryotic species can be developed which considers the ecology (including clinical significance) of the strains in addition to their evolutionary distance [33,106]. However, there are standard measurements required for taxonomic recognition of new prokaryotic, including *Mycobacterium*, species [43,77]. DNA-DNA hybridization (DDH) with a 70% binding criterion (or temperature of melting difference of less than 6 degrees C) for species identity is the gold standard. Stahl and Urbance [68], found that phylogenetic analysis of 16S rRNA sequences (an already accepted technology by the time of their work) yielded *Mycobacterium* species' differentiation and relationships roughly compatible with gold standard DDH and traditional phenotypic assignations. 16S rRNA similarity of 97% or more is the common cutoff for species identity. However, later work has shown that 16S rRNA is sometimes too conserved to discriminate between species recognized by phenotypic differences and confirmed as separate species by DDH. Other sequences, including *rpoB* [20,21], were found to improve the discrimination of species in a manner complementary to 16S rRNA (or rDNA) analysis. Recent work has shown that in addition to its ability to differentiate among bacteria at the species level, the *rpoB* gene has a strong "phylogenetic signal" that allows it to help determine ranks higher than the species in prokaryotic taxonomy [105].

In summary, in a very approximate view of the matter, the "molecular clock" for prokaryotic speciation is the time it takes between evolving species to accumulate enough genetic changes to exhibit about a 30% difference in DNA content from the type or reference strain, at which time by current standard measurements (DNA-DNA hybridization, DDH) they will be recognized as separate species. Obviously, pigmentation changes reflect differences in DNA content, but do not occur at a rate appropriate for use as a marker of speciation, and are not enough "change" to be necessary or sufficient to alone cause "speciation". Each of the 3 pigmentation types, nonchromogen, scotochromogen, and photochromogen were found in each of the MST-regions 1, 2, and 3, although the RGM did not include any photochromogens (Figure 1). On the other h and, various conserved or so-called house-keeping genes (including *rpoB*) have been found to accumulate genetic change at a rate that roughly reflects the rate of accumulated 30% change in DNA content of an emerging species, and thus can be analyzed for polymorphisms that are in linkage disequilibrium with "speciation". This is the rationale and justification for the construction and correctness of the 342bp *rpoB* fragment restriction alleles-based MST-regions model of genus *Mycobacterium* population structure. The model should be consistent with the classical identification (differentiation) of species, which includes pigmentation as one of the attributes on which to discriminate among species (it is consistent with this), and should be consistent with already observed and confirmed relationships (clades) among species (it is consistent with this). Furthermore, one should try to place important biological attributes in the framework of the model, even if on their own they are insufficient to cause "speciation," and even if their "molecular clocks" make them unsuitable for markers of prokaryotic speciation.

Finally, a new model such as presented in the proposed article, may fortuitously reveal new relationships which need to be explained. This is the case with the MST region 1 RGM and associated MST region 1 SGM, in which number and strength of rRNA promoters, type of 16S rRNA helix 18 (long or short), and number of rRNA operons (that is, growth rate attributes) may together help explain some of the differences that contributed to the divergence between the SGM of RGM-containing MST-region 1, and the SGM of MST region 3.

A new paradigm for genus *Mycobacterium* population structure was presented. The MST-regions model of genus *Mycobacterium* population structure was robust in identifying and clustering species and subspecies, unambiguous, transparent for alleles, consistent with genotypes and phenotypes, and statistically testable. In a wider context, the method may prove useful to elucidate DNA markers and genes of quantitative traits in general, including those traits essential for *Mycobacterium*-associated diseases, which could then be targeted for preventive and therapeutic interventions.

References

1. Dai J, Chen Y, Lauzardo M (2011) Web-accessible database of hsp65 sequences from Mycobacterium reference strains. J Clin Microbiol 49: 2296-2303.
2. http://www.who.int/mediacentre/factsheets/fs104/en/
3. http://who.int/tb/publications/global_report/gtbr12_executivesummary.pdf
4. Billinger ME, Olivier KN, Viboud C, de Oca RM, Steiner C, et al. (2009) Nontuberculous mycobacteria-associated lung disease in hospitalized persons, United States, 1998-2005. Emerg Infect Dis 15: 1562-1569.
5. Butler WR, Crawford JT (1999) Nontuberculous mycobacteria reported to the Public Health Laboratory Information System by state public health laboratories. Centers for Disease Control and Prevention, Atlanta, GA.
6. Behr MA, Falkinham JO 3rd (2009) Molecular epidemiology of nontuberculous mycobacteria. Future Microbiol 4: 1009-1020.
7. Brown-Elliott BA, Wallace RJ Jr (2002) Clinical and taxonomic status of pathogenic nonpigmented or late-pigmenting rapidly growing mycobacteria. Clin Microbiol Rev 15: 716-746.
8. Van Der Zanden AG, TeKoppele-Vije EM, Vijaya Bhanu N, Van Soolingen D, Schouls LM (2003) Use of DNA extracts from Ziehl-Neelsen-stained slides for molecular detection of rifampin resistance and spoligotyping of Mycobacterium tuberculosis. J Clin Microbiol 41: 1101-1108.
9. Zaczek A, Brzostek A, Augustynowicz-Kopec E, Zwolska Z, Dziadek J (2009) Genetic evaluation of relationship between mutations in rpoB and resistance of Mycobacterium tuberculosis to rifampin. BMC Microbiol 9: 10.
10. Jou R, Chen HY, Chiang CY, Yu MC, Su IJ (2005) Genetic diversity of multidrug-resistant Mycobacterium tuberculosis isolates and identification of 11 novel rpoB alleles in Taiwan. J Clin Microbiol 43: 1390-1394.
11. Sreenu VB, Kumar P, Nagaraju J, Nagarajaram HA (2006) Microsatellite polymorphism across the M. tuberculosis and M. bovis genomes: implications on genome evolution and plasticity. BMC Genomics 7: 78.
12. Patterson RJ, Youmans GP (1970) Multiplication of Mycobacterium tuberculosis Within Normal and "Immune" Mouse Macrophages Cultivated With and Without Streptomycin. Infect Immun 1: 30-40.
13. Patterson RJ, Youmans GP (1970) Demonstration in tissue culture of lymphocyte-mediated immunity to tuberculosis. Infect Immun 1: 600-603.
14. Freidlin PJ, Goldblatt D,Kaidar-Shwartz H, Dveyrin Z, Rorman E (2011) Quality assurance for molecular epidemiology of tuberculosis methods in the Mycobacterium reference laboratory. Accred Qual Assur 16: 623-635.
15. Freidlin PJ, Goldblatt D, Kaidar-Shwartz H, Rorman E (2009) Polymorphic exact tandem repeat A (PETRA): a newly defined lineage of mycobacterium tuberculosis in israel originating predominantly in Sub-Saharan Africa. J Clin Microbiol 47: 4006-4020.
16. Bansal K, Sinha AY, Ghorpade DS, Togarsimalemath SK, Patil SA, et al. (2010) Src homology 3-interacting domain of Rv1917c of Mycobacterium tuberculosis induces selective maturation of human dendritic cells by regulating PI3K-MAPK-NF-kappaB signaling and drives Th2 immune responses. J Biol Chem 285: 36511-36522.
17. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, et al. (2006) Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103: 2869-2873.
18. McEvoy CR, van Helden PD, Warren RM, Gey van Pittius NC (2009) Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic Mycobacterium tuberculosis PPE38 gene region. BMC Evol Biol 9: 237.
19. vanIngen J, Boeree MJ, van Soolingen D, Iseman MD, Heifets LB, et al. (2012) Are phylogenetic position, virulence, drug susceptibility and in vivo response to treatment in mycobacteria interrelated? Infect Genet Evol 12: 832-837.
20. Kim BJ, Lee SH, Lyu MA, Kim SJ, Bai GH, et al. (1999) Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB). J ClinMicrobiol 37:1714-1720.
21. Kim BJ, Lee KH, Park BN, Kim SJ, Bai GH, et al. (2001) Differentiation of mycobacterial species by PCR-restriction analysis of DNA (342 base pairs) of the RNA polymerase gene (rpoB). J Clin Microbiol 39: 2102-2109.
22. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN (2006) Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev 19: 658-685.
23. Tortoli E (2012) Phylogeny of the genus Mycobacterium: many doubts, few certainties. Infect Genet Evol 12: 827-831.
24. Dai J, Chen Y, Dean S, Morris JG, Salfinger M, et al. (2011) Multiple-genome comparison reveals new loci for Mycobacterium species identification. J Clin Microbiol 49: 144-153.
25. Felsenstein, J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.
26. Takezaki N, Gojobori T (1999) Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences. Mol Biol Evol 16: 590-601.
27. Templeton AR (2010) The diverse applications of cladistic analysis of molecular evolution, with special reference to nested clade analysis. Int J Mol Sci 11: 124-139.
28. Gutacker MM, Smoot JC, Migliaccio CAL, Ricklefs SM, Hua S, et al (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.
29. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, et al. (2012) High resolution discrimination of clinical Mycobacterium tuberculosis complex strains based on single nucleotide polymorphisms. PLoS One 7: e39855.
30. Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genet 6 Suppl 1: S26.
31. Dickerson RE, Geis I (1969) The structure and action of proteins. WA Benjamin,Inc. Publishers. Menlo Park, California,USA:120
32. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41: e1.
33. Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, et al. (2005) Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol 3: 733-739.
34. Schadt, EE (2013) The coming revolution: microbes and multiscale biology. Microbe 8:70-73.
35. http://www.ncbi.nlm.nih.gov/entrez/
36. JavaScript, 3rd edition, The Definitive Guide, by Flanagan, D. copyright 1998,1997,1996 O'Reillyand Associates, Inc. 101 Morris Street, Sebastopol, CA.
37. Roberts RJ, Vincze T, Posfai J, Macelis D (2010) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 38: D234-236.
38. http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html

39. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) Version 3.6(alpha) (Department of Genetics, Univ. Washington, Seattle).

40. http://faculty.vassar.edu/lowry/fisher.html

41. Noether GE (1971) Introduction to statistics – A fresh approach. Houghton Mifflin Company, Boston.

42. http://math.usask.ca/~laverty/S245/Tables/wmw.pdf

43. Rastogi N, Legrand E, Sola C (2001) The mycobacteria: an introduction to nomenclature and pathogenesis. Rev Sci Tech 20: 21-54.

44. Adékambi T, Colson P, Drancourt M (2003) rpoB-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. J ClinMicrobiol 41: 5699-5708.

45. Adékambi T, Drancourt M (2004) Dissection of phylogenetic relationships among 19 rapidly growing Mycobacterium species by 16S rRNA, hsp65, sodA, recA and rpoB gene sequencing. Int J SystEvolMicrobiol 54: 2095-2105.

46. Jeong JH, Seo YH, Kim KH, Ahn JY, Park PH, et al. (2012) Mycobacterium wolinskyi infection confirmed by rpoB gene sequencing. J Clin Lab Anal 26: 325-327.

47. Pfyffer GE (2007) Mycobacterium: general characteristics, laboratory detection, and staining procedures. In: Manual of Clinical Microbiology 1: 543-572.

48. Vincent V, Gutierrez MC (2007) Mycobacterium:laboratory characteristics of slowly growing mycobacteria. In: Manual of Clinical Microbiology (9th edn.) 1: 573-588.

49. Brown-Elliot BA, Wallace Jr RJ (2007) Mycobacterium: clinical and laboratory characteristics of rapidly growing mycobacteria. In: Manual of Clinical Microbiology 9th edition, volume 1, editor Murray PR, ASM Press, Washington, DC 589-600.

50. Mollet C, Drancourt M, Raoult D (1997) rpoB sequence analysis as a novel basis for bacterial identification. MolMicrobiol 26: 1005-1011.

51. Lee H, Park HJ, Cho SN, Bai GH, Kim SJ (2000) Species identification of mycobacteria by PCR-restriction fragment length polymorphism of the rpoB gene. J Clin Microbiol 38: 2966-2971.

52. Lee H, Bang HE, Bai GH, Cho SN (2003) Novel polymorphic region of the rpoB gene containing Mycobacterium species-specific sequences and its use in identification of mycobacteria. J Clin Microbiol 41: 2213-2218.

53. Itoh S, Kazumi Y, Abe C, Takahashi M (2003) Heterogeneity of RNA polymerase gene (rpoB) sequences of Mycobacterium gordonae clinical isolates identified with a DNA probe kit and by conventional methods. J Clin Microbiol 41: 1656-1663.

54. Kim BJ, Hong SH, Kook YH, Kim BJ (2013) Molecular evidence of lateral gene transfer in rpoB gene of Mycobacterium yongonense strains via multilocus sequence analysis. PLoS One 8: e51846.

55. Picardeau M, Prod'Hom G, Raskine L, LePennec MP, Vincent V (1997) Genotypic characterization of five subspecies of Mycobacterium kansasii. J Clin Microbiol 35: 25-32.

56. Zhang Y, Mann LB, Wilson RW, Brown-Elliott BA, Vincent V, et al. (2004) Molecular analysis of Mycobacterium kansasii isolates from the United States. J Clin Microbiol 42: 119-125.

57. Bercovier H, Kafri O, Sela S (1986) Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. BiochemBiophys Res Commun 136: 1136-1141.

58. Menéndez Mdel C, Rebollo MJ, NúñezMdel C, Cox RA, García MJ (2005) Analysis of the precursor rRNA fractions of rapidly growing mycobacteria: quantification by methods that include the use of a promoter (rrnA P1) as a novel standard. J Bacteriol 187: 534-543.

59. Menendez MC, Garcia MJ, Navarro MC, Gonzalez-y-Merchand JA, Rivera-Gutierrez S, et al (2002) Characterization of an rRNA operon (rrnB) of Mycobacterium fortuitum and other mycobacterial species: implications for the classification of mycobacteria. J Bacteriol 184: 1078-1088.

60. Arnvig KB, Gopal B, Papavinasasundaram KG, Cox RA, Colston MJ (2005) The mechanism of upstream activation in the rrnB operon of Mycobacterium smegmatis is different from the Escherichia coli paradigm. Microbiology 151: 467-473.

61. Ji YE, Colston MJ, Cox RA (1994) The ribosomal RNA (rrn) operons of fast-growing mycobacteria: primary and secondary structures and their relation to rrn operons of pathogenic slow-growers. Microbiology 140 : 2829-2840.

62. Stadthagen-Gomez G, Helguera-Repetto AC, Cerna-Cortes JF, Goldstein RA, Cox RA, et al. (2008) The organization of two rRNA (rrn) operons of the slow-growing pathogen Mycobacterium celatum provides key insights into mycobacterial evolution. FEMS MicrobiolLett 280: 102-112.

63. Stinear TP, Seemann T, Harrison PF, Jenkin GA, Davies JK, et al. (2008) Insights from the complete genome sequence of Mycobacterium marinum on the evolution of Mycobacterium tuberculosis. Genome Res 18: 729-741.

64. Helguera-Repetto C, Cox RA, Muñoz-Sànchez JL, Gonzalez-y-Merchand JA (2004) The pathogen Mycobacterium marinum, a faster growing close relative of Mycobacterium tuberculosis, has a single rRNA operon per genome. FEMS Microbiol Lett 235: 281-288.

65. Rivera-Gutiérrez S, Montoro-Cardoso E, Valdivia JA, Cox RA, Gonzalez-y-Merchand JA (2003) The number and organization of the rRNA genes of several strains of Mycobacterium simiae. FEMS Microbiol Lett 227: 133-139.

66. Arnvig KB, Zeng S, Quan S, Papageorge A, Zhang N, et al. (2008) Evolutionary comparison of ribosomal operon antitermination function. J Bacteriol 190: 7251-7257.

67. Gonzalez-y-Merchand JA, Garcia MJ, Gonzalez-Rico S, Colston MJ, Cox RA (1997) Strategies used by pathogenic and nonpathogenic mycobacteria to synthesize rRNA. J Bacteriol 179: 6949-6958.

68. Stahl DA, Urbance JW (1990) The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. J Bacteriol 172: 116-124.

69. Tortoli E (2003) Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. Clin Microbiol Rev 16: 319-354.

70. Han SH, Kim KM, Chin BS, Choi SH, Lee HS, et al. (2010) Disseminated Mycobacterium kansasii infection associated with skin lesions: a case report and comprehensive review of the literature. J Korean Med Sci 25: 304-308.

71. DeBruyn JM, Mead TJ, Sayler GS (2012) Horizontal transfer of PAH catabolism genes in Mycobacterium: evidence from comparative genomics and isolated pyrene-degrading bacteria. Environ SciTechnol 46: 99-106.

72. Gao B, Paramanathan R, Gupta RS (2006) Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. Antonie Van Leeuwenhoek 90: 69-91.

73. Gao B, Gupta RS (2012) Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. Microbiol Mol Biol Rev 76: 66-112.

74. Chun J, Goodfellow M (1995) A phylogenetic analysis of the genus Nocardia with 16S rRNA gene sequences. Int J Syst Bacteriol 45: 240-245.

75. McGuire AM, Weiner B, Park ST, Wapinski I, Raman S, et al. (2012) Comparative analysis of Mycobacterium and related Actinomycetes yields insight into the evolution of Mycobacterium tuberculosis pathogenesis. BMC Genomics 13: 120.

76. Doerks T, van Noort V, Minguez P, Bork P (2012) Annotation of the M. tuberculosis hypothetical orfeome: adding functional information to more than half of the uncharacterized proteins. PLoS One 7: e34302.

77. Lévy-Frébault VV, Portaels F (1992) Proposed minimal standards for the genus Mycobacterium and for description of new slowly growing Mycobacterium species. Int J Syst Bacteriol 42: 315-323.

78. Beste DJ, Espasa M, Bonde B, Kierzek AM, Stewart GR, et al. (2009) The genetic requirements for fast and slow growth in mycobacteria. PLoS One 4: e5349.

79. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, et al. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PloS Pathog 7: e1002251.

80. Gerasimova A, Kazakov AE, Arkin AP, Dubchak I, Gelfand MS (2011) Comparative genomics of the dormancy regulons in mycobacteria. J Bacteriol 193: 3446-3452.

81. Rocha EP, Danchin A, Viari A (2001) Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. Genome Res 11: 946-958.

82. Karlin S, Burge C, Campbell AM (1992) Statistical analyses of counts and

distributions of restriction sites in DNA sequences. Nucleic Acids Res 20: 1363-1370.

83. McClelland M, Nelson M, Raschke E (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Res 22: 3640-3659.

84. Zheng Y, Cohen-Karni D, Xu D, Chin HG, Wilson G, et al. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. Nucleic Acids Res 38: 5527-5534.

85. vanSoolingen D, de Haas PE, Blumenthal RM, Kremer K, Sluijter M, et al. (1996) Host-mediated modification of PvuII restriction in Mycobacterium tuberculosis. J Bacteriol 178: 78-84.

86. Varshney U, Ramesh V, Madabushi A, Gaur R, Subramanya HS, et al. (2004) Mycobacterium tuberculosis Rv2118c codes for a single-component homotetrameric m1A58 tRNA methyltransferase. Nucleic Acids Res 32: 1018-1027.

87. Kumar A, Saigal K, Malhotra K, Sinha KM, Taneja B (2011) Structural and functional characterization of Rv2966c protein reveals an RsmD-like methyltransferase from Mycobacterium tuberculosis and the role of its N-terminal domain in target recognition. J Biol Chem 286: 19652-19661.

88. Georghiou SB, Magana M, Garfein RS, Catanzaro DG, Catanzaro A, et al. (2012) Evaluation of genetic mutations associated with Mycobacterium tuberculosis resistance to amikacin, kanamycin and capreomycin: a systematic review. PLoS One 7: e33275.

89. Smith SE, Showers-Corneli P, Dardenne CN, Harpending HH, Martin DP, et al. (2012) Comparative genomic and phylogenetic approaches to characterize the role of genetic recombination in mycobacterial evolution. PLoS One 7: e50070.

90. Zakham F, Belayachi L, Ussery D, Akrim M, Benjouad A, et al. (2011) Mycobacterial species as case-study of comparative genome analysis. Cell MolBiol (Noisy-le-grand) 57 Suppl: OL1462-1469.

91. Krzywinska E, Krzywinski J, Schorey JS (2004) Naturally occurring horizontal gene transfer and homologous recombination in Mycobacterium. Microbiology 150: 1707-1712.

92. Rabello MC, Matsumoto CK, Almeida LG, Menendez MC, Oliveira RS, et al. (2012) First description of natural and experimental conjugation between Mycobacteria mediated by a linear plasmid. PLoS One 7: e29884.

93. Nguyen KT, Piastro K, Gray TA, Derbyshire KM (2010) Mycobacterial biofilms facilitate horizontal DNA transfer between strains of Mycobacterium smegmatis. J Bacteriol 192: 5134-5142.

94. Lamrabet O, Merhej V, Pontarotti P, Raoult D, Drancourt M (2012) The genealogic tree of mycobacteria reveals a long-standing sympatric life into free-living protozoa. PLoS One 7: e34754.

95. Pope WH, Jacobs-Sera D, Russell DA, Peebles CL, Al-Atrache Z, et al. (2011) Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PLoS One 6: e16329.

96. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, et al. (2005) Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog 1: e5.

97. Rosas-Magallanes V, Deschavanne P, Quintana-Murci L, Brosch R, Gicquel B, et al. (2006) Horizontal transfer of a virulence operon to the ancestor of Mycobacterium tuberculosis. MolBiolEvol 23: 1129-1135.

98. Veyrier F, Pletzer D, Turenne C, Behr MA (2009) Phylogenetic detection of horizontal gene transfer during the step-wise genesis of Mycobacterium tuberculosis. BMC Evol Biol 9: 196.

99. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, et al. (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99: 3684-3689.

100. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, et al. (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat Genet 45: 172-179.

101. Garcia-Betancur JC, Menendez MC, Del Portillo P, Garcia MJ (2012) Alignment of multiple complete genomes suggests that gene rearrangements may contribute towards the speciation of Mycobacteria. Infect Genet Evol 12: 819-826.

102. Adékambi T, Butler RW, Hanrahan F, Delcher AL, Drancourt M, et al. (2011) Core gene set as the basis of multilocus sequence analysis of the subclass Actinobacteridae. PLoS One 6: e14792.

103. Snyder L, Champness W (2007) Molecular genetics of bacteria. ASM Press, Washington, DC USA.

104. Ganoza MC, Kiel MC, Aoki H (2002) Evolutionary conservation of reactions in translation. MicrobiolMolBiol Rev 66: 460-485, table of contents.

105. Konstantinidis KT, Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187: 6258-6264.

106. Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. ProcNatlAcadSci U S A 102: 2567-2572.

107. van Embden JD, van Gorkom T, Kremer K, Jansen R, van Der Zeijst BA, et al. (2000) Genetic variation and evolutionary origin of the direct repeat locus of Mycobacterium tuberculosis complex bacteria. J Bacteriol 182: 2393-2401.

108. Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322: 1843-1845.

109. Maniv I, Hatoum-Aslan A, Marraffini LA (2013) CRISPR decoys: Competitive inhibitors of CRISPR immunity. RNA Biol 10.

110. Namouchi A, Didelot X, Schöck U, Gicquel B, Rocha EP (2012) After the bottleneck: Genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. Genome Res 22: 721-734.

111. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. PLoS Genet 8: e1002453.