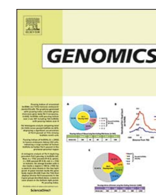
Genomics 109 (2017) 471–474

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno



Minireview

Limitations of the *Mycobacterium tuberculosis* reference genome H37Rv in the detection of virulence-related loci

Ronan F. O'Toole[a,b,*], Sanjay S. Gautam[a]

[a] School of Medicine, Faculty of Health, University of Tasmania, Hobart TAS 7000, Australia
[b] Department of Clinical Microbiology, Sir Patrick Dun Laboratory, School of Medicine, Trinity College Dublin, Dublin 8, Ireland

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis
Whole genome sequencing
H37Rv
CDC1551

ABSTRACT

The genome sequence of *Mycobacterium tuberculosis* strain H37Rv is an important and valuable reference point in the study of *M. tuberculosis* phylogeny, molecular epidemiology, and drug-resistance mutations. However, it is becoming apparent that use of H37Rv as a sole reference genome in analysing clinical isolates presents some limitations to fully investigating *M. tuberculosis* virulence. Here, we examine the presence of single locus variants and the absence of entire genes in H37Rv with respect to strains that are responsible for cases and outbreaks of tuberculosis. We discuss how these polymorphisms may affect phenotypic properties of H37Rv including pathogenicity. Based on our observations and those of other researchers, we propose that use of a single reference genome, H37Rv, is not sufficient for the detection and characterisation of *M. tuberculosis* virulence-related loci. We recommend incorporation of genome sequences of other reference strains, in particular, direct clinical isolates, in such analyses in addition to H37Rv.

1. Introduction

The first strain of *Mycobacterium tuberculosis* [1,2] to undergo whole-genome sequencing was H37Rv. The genome was sequenced by Cole and colleagues from the Pasteur Institute in Paris and the Wellcome Trust Sanger Centre in Cambridge, and published in 1998 in a landmark paper for the tuberculosis field [3]. H37Rv in itself is not a direct clinical isolate, but was derived in a laboratory from H37, the original strain collected from a patient at the Trudeau Institute, Saranac Lake, New York in 1905 [4]. Oatway and Steenken reported the dissociation of *M. tuberculosis* H37 into two variants, H37Rv and H37Ra, which exhibited high and low virulence, respectively, when inoculated through different routes into guinea pigs and rabbits [5]. H37Rv was found to exhibit a stippled, spreading colony with acid-fast rods whereas H37Ra was described as having a discrete crater-like worm-cast colony containing bacilli with granular and non-acid-fast tendencies [5]. A comparison of 39 strains of *M. tuberculosis* based on staining characteristics, cellular dimensions, colony morphology, growth rate, and animal pathogenicity, led to the deduction by Kubica and co-workers at the Trudeau Institute in 1972 that "the original tubercle bacilli of Koch are indeed those which today are referred to as *M. tuberculosis*, as exemplified by the neotype, H37Rv" [4].

2. H37Rv as a reference strain for TB research

The reported phenotypic similarity between H37Rv and the tuberculosis-causing bacteria described by Robert Koch in 1882 [6], combined with its worldwide distribution and common use, led to the adoption of H37Rv as a principal *M. tuberculosis* strain for tuberculosis vaccine [7], diagnostics [8], and drug therapy [9] research. With regard to genome-based studies, H37Rv has been used as the primary reference strain in multiple works on *M. tuberculosis* phylogeny, molecular epidemiology, and the detection of drug-resistance mutations [10–12]. Its genome is among the best curated of any bacterial species with a number of resources linking its gene annotations to transcriptomic, proteomic, and functional data such as Tuberculist (http://tuberculist.epfl.ch), KEGG (http://www.genome.jp/kegg/), and the TB Database (http://genome.tbdb.org/tbdb_sysbio/MultiHome.html). *M. tuberculosis* has been considered to be a highly monomorphic bacterium [13], hence choice of reference genome was not thought to have a significant effect on the genome analysis outputs. Indeed, based on a recent analysis of 162 closely-related Euro-American lineage 4 isolates using 7 different reference genomes, Lee and Behr reported that "the choice of reference genome, within the *M. tuberculosis* complex, has negligible influence on phylogeny and epidemiological studies of *M. tuberculosis* transmission"

Abbreviations: CFP-10, 10-kDa culture filtrate protein; ESAT-6, 6-kDa early secreted antigen target; ESX-1, ESAT-6 secretion system 1; HGC, homologous gene cluster; MoCo, molybdopterin cofactor; NADP, nicotinamide adenine dinucleotide phosphate; ORF, open-reading frame; PDIM, phthiocerol dimycocerosate; PE, proline-glutamate motif protein; PPE, proline-proline-glutamate motif protein

* Corresponding author at: School of Medicine, Faculty of Health, University of Tasmania, Medical Science 1 Building, 17 Liverpool Street, Hobart TAS 7000, Australia.
E-mail address: ronan.otoole@utas.edu.au (R.F. O'Toole).

http://dx.doi.org/10.1016/j.ygeno.2017.07.004
Received 25 June 2017; Received in revised form 17 July 2017; Accepted 18 July 2017
Available online 22 July 2017
0888-7543/ © 2017 Elsevier Inc. All rights reserved.

[14]. H37Rv is a therefore a key tool in the study of *M. tuberculosis* and the preferred strain of choice for phylogenetic and epidemiological studies.

However, it is becoming more apparent from comparisons of the genomes of clinical isolates of *M. tuberculosis* that features of the H37Rv strain that can limit its robustness in the study of *M. tuberculosis* pathogenicity. For example, a previous study reported that H37Rv did not induce caseous necrosis in mice or the formation of multinucleate giant cells in contrast to clinical isolates of *M. tuberculosis* [15]. Other researchers have reported that low doses of H37Rv failed to produce clinically-recognisable disease in rhesus monkeys within 18 weeks of intra-bronchial inoculation, in contrast to the Erdman strain, although animals were tuberculin-skin test positive after 4 weeks [16]. In an *in vitro* infection assay using human monocyte-derived macrophages, H37Rv was found to exhibit a slower growth rate with respect to a patient isolate [17]. Due to the observed differences between strains, De Groote and co-workers have highlighted the need to use other strains of *M. tuberculosis*, in addition to H37Rv, when measuring the *in vivo* efficacy of anti-tubercular drug regimens [18]. Following a comparison of the virulence, immunopathology, and transmissibility of eight different clinical isolates of *M. tuberculosis* in mice, Marquina-Castillo and colleagues proposed that "that the current use of H37Rv as the standard for animal models may be flawed because there were important differences in pathology caused by H37Rv" [19].

3. Single-locus variations between H37Rv and other strains of *M. tuberculosis*

Prolonged laboratory culture of H37Rv, since isolation of H37 in 1905, could have potentially contributed to changes in this strain over time. Solans and co-investigators identified a + G insertion at the − 74 position upstream of the *whiB6* (Rv3862c) start codon in the genomes of H37Rv and H37Ra that correlated with the decreased production of the 6-kDa early-secreted antigen target (ESAT-6) in these strains [20]. ESAT-6 is exported along with the 10-kDa culture filtrate protein (CFP-10) through the cell envelope of *M. tuberculosis* by the type VII ESAT-6 secretion system 1 (ESX-1) [21]. ESAT-6 mediates mycobacterial translocation from the phagosome to the cytosol which is required for the pathogenicity of *M. tuberculosis* [22]. The + G insertion upstream of the *whiB6* gene was not found in the genomes of 76 clinical isolates of *M. tuberculosis*, including CDC1551 which exhibits higher levels of ESAT-6 production with respect to H37Rv [20]. Introduction of a wild-type copy of the *whiB6* gene into H37Rv re-established ESAT-6 expression and secretion to the level of clinical strains highlighting the importance of the + G insertion [20].

The PhoP/PhoR two-component system is required for the growth of *M. tuberculosis* during infection of macrophages and mice [23]. The system regulates the biosynthesis of complex cell wall lipids *i.e.* sulphatides, diacyltrehalose, and polyacyltrehalose [23]. A single nucleotide polymorphism has been identified in H37Rv, with respect to CDC1551, which causes a leucine to proline substitution at position 152 of the sensor kinase PhoR (Rv0758) and is associated with elevated cell wall hydrophobicity [24]. Introduction of the Leu152Pro substitution into the CDC1551 background resulted in an increase in hydrophobicity to the level of the H37Rv strain indicating that the SNP has functional consequences for *M. tuberculosis* [24].

Comparisons have also been made between the genome sequences of H37Rv and its attenuated relative, H37Ra. These analyses have been useful for detecting possible H37Ra-specific variations that may contribute to its reduced virulence and include SNPs in genes that encode the two-component response regulator, PhoP, and the phospholipase C, PlcD, as well as members of the PE and PPE multi-protein families [25–27]. Applying the more recent single-molecule real time (SMRT) sequencing technology to re-sequencing of the H37Ra genome, Elghraoui et al. noted that the number of H37Ra-specific SNPs with respect to H37Rv was less than half that of the 76 previously-reported for the H37Ra genome sequence generated by Sanger-based shotgun sequencing of clones libraries [26,28]. The authors also reported that they identified an error at position 459399 of the H37Rv reference genome sequence upstream of Rv0383c. While this issue is not unique to H37Rv or H37Ra, the transition to newer sequencing platforms may involve revisiting some reference genomes obtained using earlier technologies.

4. Single-locus variations in different laboratory stocks of H37Rv

In addition to inter-strain single locus variations, Ioerger and colleagues performed whole-genome sequencing on isolates of H37Rv obtained from 6 laboratories in the USA and South Africa and detected genetic differences between different stocks of the strain [29]. Excluding differences shared with the genome sequence of strain H37Ra, the authors found 24 single nucleotide polymorphisms or indels among the six isolates of H37Rv, of which 22 occurred in coding regions. The variants are believed to have arisen from so-called *in vitro* evolution of the isolates following cultivation over time at the different sites [29]. Although the number of differences identified was not high (5–10 polymorphisms per strain), some of them may be functionally relevant to *M. tuberculosis*. For example, two of the isolates, H37RvLP and H37RvJO, were found to contain a + GC insertional frame-shift in the mycocerosic acid synthase gene, *mas* (Rv2940c), causing them to be deficient in biosynthesis of the cell wall glycolipid phthiocerol dimycocerosate (PDIM) [29]. An earlier signature-tagged mutagenesis study found that synthesis of PDIM is required for growth of *M. tuberculosis* in the lungs of mice [30]. Furthermore, Domenech and Reed reported that H37Rv is prone to losing its ability to produce PDIM following *in vitro* passage and recommended that researchers conducting *in vivo* virulence studies in this background strain confirm the presence of PDIM [31]. Owing to these differences among different stocks of H37Rv, Ioerger and co-workers concluded that "H37Rv as a standard reference strain should be used with some caution, as experimental results derived with 'H37Rv' may depend on the laboratory in which it is maintained and the associated unique genetic characteristics" [29].

5. Differences in gene content between H37Rv and *M. tuberculosis* clinical isolates

Fleischmann and colleagues were among the first to detect the absence of corresponding genes of strain *M. tuberculosis* CDC1551 in the H37Rv genome [32]. They noted the insertion of 17 complete open-reading frames (ORF) in CDC1551 that are not present in H37Rv. Among these ORFs, 9 have functional assignments that include an adenylate cyclase (MT1360), a glycosyl-transferase (MT1800), an oxidoreductase (MT1801), a 12 transmembrane-domain protein (MT1802), a membrane lipoprotein (MT2619), a proline-proline-glutamate motif (PPE) family protein (MT3248), paralogs of *moaB* (MT3426) and *moaA* (MT3427), and a putative transcription regulator (MT3428).

MoaA and MoaB are required for the biosynthesis of molybdopterin cofactor (MoCo), an essential co-factor for redox reaction enzymes such as the *narGHI*-encoded nitrate reductase [33,34]. Nitrate reductase has been shown to enable *M. tuberculosis'* adaptation to hypoxic conditions and its persistence in the lungs of guinea pigs [35]. Situated adjacent to the *moaA* and *moaB* genes, MT3428 encodes EmbR2, an orthologue of the transcriptional regulator EmbR (MT1305) [36]. EmbR2 inhibits the autokinase activity of PknH and the subsequent phosphoryl transfer to EmbR [36]. An *M. tuberculosis pknH* mutant has been shown to survive and replicate to a higher level in the lungs and spleens of mice than its parental strain [37]. Through its modulation of the kinase activity of PknH, EmbR2 is believed to participate in the physiology and virulence of *M. tuberculosis* [36].

From their analysis of 100 strains belonging to the *M. tuberculosis* complex, Brosch, Gordon and co-workers determined that the *moa* gene containing region, RvD5, was absent from strain H37Rv but present in

Download English Version:

https://daneshyari.com/en/article/8646364

Download Persian Version:

https://daneshyari.com/article/8646364

Daneshyari.com