Rhomboid homologs in mycobacteria: insights from phylogeny and genomic analysis

Kateete et al.
Rhomboid homologs in mycobacteria: insights from phylogeny and genomic analysis

David P Kateete, Moses Oke, Fred A Katabazi, Alfred Okeng, Jeniffer Asiimwe, Henry W Boom, Kathleen D Eisenach, Moses L Joloba

Abstract

Background: Rhomboids are ubiquitous proteins with diverse functions in all life kingdoms, and are emerging as important factors in the biology of some pathogenic apicomplexa and Providencia stuartii. Although prokaryotic genomes contain one rhomboid, actinobacteria can have two or more copies whose sequences have not been analyzed for the presence putative rhomboid catalytic signatures. We report detailed phylogenetic and genomic analyses devoted to prokaryotic rhomboids of an important genus, Mycobacterium.

Results: Many mycobacterial genomes contained two phylogenetically distinct active rhomboids orthologous to Rv0110 (rhomboid protease 1) and Rv1337 (rhomboid protease 2) of Mycobacterium tuberculosis H37Rv, which were acquired independently. There was a genome-wide conservation and organization of the orthologs of Rv1337 arranged in proximity with glutamate racemase (mur1), while the orthologs of Rv0110 appeared evolutionary unstable and were lost in Mycobacterium leprae and the Mycobacterium avium complex. The orthologs of Rv0110 clustered with eukaryotic rhomboids and contained eukaryotic motifs, suggesting a possible common lineage. A novel nonsense mutation at the Trp73 codon split the rhomboid of Mycobacterium avium subsp. Paratuberculosis into two hypothetical proteins (MAP2425c and MAP2426c) that are identical to MAV_1554 of Mycobacterium avium. Mycobacterial rhomboids contain putative rhomboid catalytic signatures, with the protease active site stabilized by Phenylalanine. The topology and transmembrane helices of the Rv0110 orthologs were similar to those of eukaryotic secretase rhomboids, while those of Rv1337 orthologs were unique. Transcription assays indicated that both mycobacterial rhomboids are possibly expressed.

Conclusions: Mycobacterial rhomboids are active rhomboid proteases with different evolutionary history. The Rv0110 (rhomboid protease 1) orthologs represent prokaryotic rhomboids whose progenitor may be the ancestors of eukaryotic rhomboids. The Rv1337 (rhomboid protease 2) orthologs appear more stable and are conserved nearly in all mycobacteria, possibly alluding to their importance in mycobacteria. MAP2425c and MAP2426c provide the first evidence for a split homologous rhomboid, contrasting whole orthologs of genetically related species. Although valuable insights to the roles of rhomboids are provided, the data herein only lays a foundation for future investigations for the roles of rhomboids in mycobacteria.

Background

The genus Mycobacterium consists of ~148 species [1], of which some are leading human and animal pathogens. Tuberculosis (TB), the most important mycobacterial disease, is caused by genetically related species commonly referred to as "the Mycobacterium tuberculosis Complex" (MTC: Mycobacterium tuberculosis; M. bovis, also the causative agent of bovine TB; M. bovis BCG; M. africanum; M. carnetti and M. microti [2]). M. leprae and M. ulcerans are respectively the causative agents for two other important diseases, Leprosy and Buruli ulcer [3,4]. Besides the three major diseases, M. avium subsp. Paratuberculosis causes John’s disease (a fatal disease of dairy cattle [5]) and is also suspected to cause Crohn’s disease in humans [5]. In addition, M. avium and other non-tuberculuous mycobacteria (NTM) have become important opportunistic pathogens of immunocompromised humans and animals [6,7].
Mycobacteria have versatile lifestyles and habitats, complexities also mirrored by their physiology. While some can be obligate intracellular pathogens (i.e. the MTC species) [8], others are aquatic inhabitants, which can utilize polycyclic aromatic hydrocarbons (i.e. M. vanbaalenii) [9]. The biology of pathogenic mycobacteria remains an enigma, despite their importance in human and veterinary medicine. Except for the mycolactone of M. ulcerans, glycolipids (such as PDIMs) and proteins (such as ESAT-6) of MTC species [10,11], largely, in contrast to most bacterial pathogens, pathogenic mycobacteria lack obvious virulence factors and the mechanisms in which they cause diseases are still obscure [12]. Genome sequencing projects have provided invaluable tools that are accelerating the understanding of the biology of pathogenic mycobacteria. As such, genome sequencing data has guided the characterization of genes/pathways for microbial pathogens, accelerating discovery of novel control methods for the intractable mycobacterial diseases [5,13-16].

The rhomboid protein family exists in all life kingdoms and has rapidly progressed to represent a ubiquitous family of novel proteins. The knowledge and the universal distribution of rhomboids was engendered and accelerated by functional genomics [17]. The first rhomboid gene was discovered in Drosophila melanogaster as a mutation with an abnormally rhomboid-shaped head skeleton [17,18]. Genome sequencing data later revealed that rhomboids occur widely in both eukaryotes and prokaryotes [17]. Many eukaryotic genomes contain several copies of rhomboid-like genes (seven to fifteen) [19], while most bacteria contain one homolog [19].

Despite biochemical similarity in mechanism and specificity, rhomboid proteins function in diverse processes including mitochondrial membrane fusion, apoptosis and stem cell differentiation in eukaryotes [20]. Rhomboid proteases are also involved in life cycles of some apicomplexan parasites, where they participate in red blood cell invasion [21-25]. Rhomboids are now linked to general human diseases such as early-onset blindness, diabetes and pathways of cancerous cells [20,26,27]. In bacteria, aarA of Providencia stuartii was the first rhomboid homolog to be characterized, which was shown to mediate a non-canonical type of quorum sensing in this gram negative species [28-30]. Since then, bacterial rhomboids are being characterized, albeit at low rate; glpU of Bacillus subtilis is involved in cell division and glucose transport [31], while glpG of Escherichia coli [17,32] was the first rhomboid to be crystallized, paving way for delineation of the mechanisms of action for rhomboid proteases [33,34].

Although universally present in all kingdoms, not all rhomboids are active proteases [19,35]. Lemberg and Freeman [35] defined the rhomboid family as genes identified by sequence homology alone, and the rhomboid proteases as a subset that includes only genes with all necessary features for predicted proteolytic activity. As such, rhomboid-like genes in eukaryotic genomes are classified into the active rhomboids, inactive rhomboids (known as the iRhoms) and a diverse group of other proteins related in sequence but predicted to be catalytically inert. The eukaryotic active rhomboids are further divided into two subfamilies: the secretase rhomboids that reside in the secretory pathway or plasma membrane, and the PARL subfamily, which are mitochondrial [35].

Despite their presence in virtually all eubacteria, there is a paucity of information about the functions of bacterial rhomboids. Hitherto, full phylogenetic analysis of rhomboids from the complex and populous prokaryotes has not been done; although it can provide important functional and evolutionary insights [17,35], it is a huge and difficult task to perform at once. Many species of mycobacteria contain two copies of rhomboid homologs whose sequences have not been investigated for the presence of functional signatures. Furthermore, actinobacteria can have up to five copies of rhomboids, the significance of which is currently not known. This study aimed at determining the distribution, evolutionary trends and bioinformatic analysis of rhomboids from an important genus -Mycobacterium.

Herein we report that mycobacterial rhomboids are active proteases with different evolutionary history, with Rv0110 orthologs representing a group of prokaryotic rhomboids whose progenitor may be the ancestor for eukaryotic rhomboids.

**Results and discussion**

A quest for the role(s) of rhomboids in mycobacteria is overshadowed by their diverse functions across kingdoms and even within species. Their presence across kingdoms implies that rhomboids are unusual useful factors that originated early in the evolution of life and have been conserved [20]. However, neither the reason for their implied significance nor the path of their evolution are understood; the key to answering these questions is rooted in understanding not only the sequence distribution of these genes, but more importantly, their functions across evolution [17,20]. This study reports that mycobacterial rhomboids are active rhomboid-serine-proteases with different evolutionary history. Reverse Transcriptase-PCRs on mycobacterial mRNA indicate that both copies of rhomboids are transcribed.

The distribution of rhomboids in mycobacteria: a nearly conserved rhomboid with unique genome organization across the genus

In determining the distribution of rhomboid homologs in mycobacteria, we used the two rhomboids of *M.*
H37Rv, Rv0110 (rhomboid protease 1) and Rv1337 (rhomboid protease 2) as reference and query sequences. Many mycobacterial genomes contained two rhomboids, which were orthologous either to Rv0110 or Rv1337. However, there was only one homolog in the genomes of the MAC (Mycobacterium avium complex) species, \textit{M. leprae} and \textit{M. ulcerans}, which were orthologous either to Rv1337 (MAC and \textit{M. leprae} rhomboids) or Rv0110 (\textit{M. ulcerans} rhomboid). \textit{M. ulcerans} was the only mycobacterial species with an ortholog of Rv0110 as a sole rhomboid. Thus, with the exception of \textit{M. ulcerans} which had a rhomboid-like element (MUL_3926, pseudogene), there is a genome-wide conservation of the rhomboids orthologs to Rv1337 (rhomboid protease 2) in mycobacteria (figure 1).

Despite evolutionary differences across the genus, the Rv1337 mycobacterial orthologs shared a unique genome organization at the rhomboid locus, with many of the rhomboid surrounding genes conserved (figure 1). Typically, upstream and downstream of the rhomboid were \textit{cysM} (cysteine synthetase) and \textit{mur1} (glutamate racemase) encoding genes. Since Rv1337 orthologs are almost inseparable from \textit{mur1} and \textit{cysM}, it is likely that they are co-transcribed (polycistronic) or functional partners. As such, we may consider the cluster containing mycobacterial Rv1337 orthologs as a putative operon. According to Sassetti et al. [36,37], many of the rhomboid surrounding genes are essential while others (including rhomboid protease 2, Rv1337) are required for the survival of the tubercle bacillus in macrophages [38].

Despite massive gene decay in \textit{M. leprae}, ML1171 rhomboid had similar genome arrangement observed for mycobacterial species. Upstream of ML1171 were gene elements (pseudogenes) ML1168, ML1169 and ML1170 (the homolog of \textit{cysM} which is conserved downstream most Rv1337 orthologs). Similar to \textit{M. lepae}, the MAC species also had an ortholog of Rv1337 as a sole rhomboid; perhaps the ortholog of Rv0110 was lost in the progenitor for MAC and \textit{M. leprae} (these species are phylogenetically related and appear more ancient in comparison to \textit{M. marinum}, \textit{M. ulcerans} and MTC species [39]). In contrast to most mycobacterial genomes, \textit{cysM} was further upstream the \textit{M. marinum} rhomboid (MMAR\_4059); and despite being genetically related to MTC species [40], MMAR\_4059 does not share much of the genome organization observed for Rv1337 MTC orthologs (figure 1).

The rhomboid-like element of \textit{M. ulcerans} (MUL\_3926, pseudogene) was identical to MMAR\_4059 (~96% similarity to MMAR\_4059) with a 42 bp insertion at the beginning and eight single nucleotide polymorphisms (SNPs). Perhaps the insertion disrupted the open reading frame (ORF) of MUL\_3926, converting it into a pseudogene. Interestingly, MUL\_3926 nearly assumed the unique organization observed for mycobacterial orthologs of Rv1337, in which the rhomboid element was upstream of \textit{mur1}.

The functional and evolutionary significance for the unique organization of the Rv1337 orthologs in mycobacteria is not clear. Since physiological roles are not yet ascribed to mycobacterial rhomboids, it is not certain whether MUL\_3926 (pseudogene) would mimic similar roles in that it almost assumed similar genomic organization (note: functions have been ascribed to certain pseudogenes [41-43]). However, the fact that \textit{M. ulcerans} is a new species (recently evolved from \textit{M. marinum} [40]) that has undergone reductive evolution, MUL\_3926 could be a consequence of these recent phenomena [44]. Interestingly, MUL\_3926 was the only rhomboid-like element in mycobacteria.

In contrast, the genome organization for Rv0110 orthologs was not conserved, and mirrored the genetic relatedness of mycobacteria (figure 2). As such, the orthologs from MTC species, \textit{M. marinum} and \textit{M. ulcerans}, which are genetically related and are assumed to have the same \textit{M. marinum}-like progenitor [39,40,45,46] had similar organization for Rv0110 ortholog. Downstream and upstream of the rhomboid were respectively, the transmembrane acyltransferase and the Proline-Glutamate polymorphic GC rich-repetitive sequence (PE-PGRS) encoding genes. PE-PGRS occurs widely in \textit{M. marinum} and MTC genomes [39] but it was a pseudogene upstream MUL\_4822 of \textit{M. ulcerans}. The distances between MTC Rv0110 orthologs and the neighboring genes were long, in contrast to the short distances between Rv1337 rhomboids and their neighboring genes.

Similarly, the genome organization for the Rv0110 orthologs of \textit{M. gilvum}, \textit{M. vanbaalenii} and \textit{Mycobacterium} species M.Jls, Mkmcs and Mmcs was also similar. Upstream and downstream the rhomboid was, respectively, the glyoxalase/bleomycin resistance protein/dioxygenase encoding gene and a gene that encodes a hypothetical protein. In contrast to MTC species, the Rv0110 orthologs in these species were close or contiguous with the neighboring genes (figure 2). The genome organization of MAB\_0026 of \textit{M. abscessus} and MSMEG\_5036 of \textit{M. smegmatis} were unique to these species (not shown).

Many bacterial genomes contain a single copy of rhomboid. However, filamentous actinobacteria such as \textit{Streptomyces coelator} and \textit{Streptomyces scabiei} have as many as four or five copies of rhomboid-like genes. Since multi-copy rhomboids in prokaryotic genomes are not yet characterized, it is not certain whether prokaryotic rhomboids can also have diverse functions, similar to multi-copy rhomboids in eukaryotic genomes. Mycobacteria and actinobacteria at large exhibit diverse
physiological and metabolic properties. It remains to be determined whether the diversity in number, nature and functions of rhomboids can contribute to the complex lifestyles of these organisms [8].

**Similarity between the two mycobacterial rhomboid paralogs**

Across the genus, the similarity between the two mycobacterial rhomboid paralogs was as low as that between prokaryotic and eukaryotic rhomboids (~10-20% identity) [19]. Since paralogs perform biologically distinct functions [47], the two mycobacterial rhomboids may have distinct roles. Eukaryotic rhomboid paralogs are also dissimilar and differ in functions in a particular species [17]. In contrast, the orthologs had significantly high homology (see table 1), with an average identity of 74%. Rv0110 orthologs within the MTC and MAC species had an identity of ~100% while those from other mycobacterial species had identities ranging from 61 to 78% (table 1). The exception was MAB_0026 of *M. abscessus*, which shared a significantly low homology with Rv0110 (38% identity at 214 amino acid overlap).
This could be due to the large evolutionary distance between M. abscessus and other mycobacteria. Since proteins of ~70% identity or more are likely to have similar functions [48], MAB_0026 may have unique roles.

The two mycobacterial rhomboids were acquired independently

To determine evolutionary relationship between the two rhomboid paralogs, phylogenetic analysis was done and included distant eukaryotic and prokaryotic rhomboids. The mycobacterial rhomboids clustered into two distinct clades with high Bootstrap values (99-100%), indicating that the rhomboids could have been acquired independently (figure 3A). Each clade consisted of rhomboids orthologous either to Rv0110 or Rv1337, grouped according to genetic relatedness of mycobacteria [39], with MAB_0026 of M. abscessus appearing the most distant. The phylogenetic analysis confirmed that the two mycobacterial rhomboids are paralogs, but their progenitor could not be determined. Thus, the mycobacterial rhomboid paralogs may be “outparalogs” (i.e. they could have resulted from duplication(s) preceding a speciation event [47]), while the orthologs could have originated from a single ancestral gene in the last common ancestor [47]). The Neighbor-Joining and Minimum Evolution phylogenetic trees were compared and gave almost comparable results.

The Rv0110 (rhomboid protease 1) mycobacterial orthologs (boxed blue) clustered with eukaryotic secre- 
tase and PARI rhomboids with a high Bootstrap value (85%, figure 3A). When grouped with eukaryotic
rhoms, the Bootstrap value for this clade increased to 90%, with rhoms forming a distinct clade (not shown). The Rv0110 mycobacterial orthologs may represent pro-
karyotic rhomboids with similar lineage or progenitor for eukaryotic active rhomboids. This was previously noted by Koonin et al [19], who hinted on a subfamily of eukaryotic rhomboids that clustered with rhomboids of Gram positive bacteria. Indeed, the Rv0110
Table 1 The distribution and similarity of mycobacterial rhomboids

| Orthologs of Rv0110 (rhomboid protease 1) |
|------------------------------------------|
| Species/strain | Rhomboid | Length | % Identity | E-value | % Identity | E-value |
| H37Rv | Rv0110 | 284 | 100 | 5e-143 | 26 | 3e-06 |
| BCG Tokyo | JTY_0114 | 284 | 100 | 3e-143 | 26 | 3e-06 |
| M. bovis | Mb0114 | 284 | 100 | 3e-143 | 26 | 3e-06 |
| M. ulcerans | MUL_4822 | 254 | 78 | 5e-104 | 27 | 1e-04 |
| M. marinum | MMAR_0300 | 289 | 77 | 1e-103 | 26 | 2e-06 |
| M. avium | MAV_1554 | 223 | NS | 1e-04 |
| M. leprae | MLEP_1554 | 223 | NS | 1e-04 |
| MAP | MAP2425c | 223 | NS | 1e-04 |
| M. smegmatis | MSMEG_4904 | 219 | NS | 1e-05 |
| M. vanbaalenii | Mvan_5753 | 290 | 61 | 6e-77 | NS | 6e-08 |
| M. gilvum | Mflv_1071 | 279 | 61 | 7e-73 | NS | 2e-06 |
| M. abscessus | MAB_0026 | 287 | 38 | 7e-38 | NS | 1e-04 |
| Orthologs of Rv1337 (rhomboid protease 2) |
| Species/strain | Rhomboid | Length | % Identity | E-value | % Identity | E-value |
| H37Rv | Rv1337 | 240 | 27 | 7e-06 | 100 | 7e-137 |
| BCG Tokyo | JTY_1373 | 240 | 27 | 7e-06 | 100 | 7e-137 |
| M. bovis | Mb1372 | 240 | 27 | 7e-06 | 100 | 7e-137 |
| M. marinum | MMAR_4059 | 222 | 26 | 8e-07 | 83 | 2e-106 |
| M. avium | MAV_1554 | 223 | 28 | 9e-05 | 75 | 7e-95 |
| M. leprae | MLEP_1554 | 238 | 27 | 1e-04 | 73 | 7e-94 |
| MAP | MAP2425c | 223 | NS | 1e-04 |
| M. smegmatis | MSMEG_4904 | 219 | NS | 1e-05 |
| M. vanbaalenii | Mvan_4290 | 225 | NS | 4e-05 |
| M. gilvum | Mflv_2355 | 225 | 27 | 7e-04 | 66 | 9e-68 |
| M. abscessus | MAB_1481 | 225 | 27 | 8e-05 | 61 | 4e-07 |

* In comparison to Rv0110 and Rv1337 of M. tuberculosis H37Rv; lengths refer to number of amino acids.
* **Mycobacterium tuberculosis.**
* **Mycobacterium bovis.**
* **Mycobacterium species JLS.**
* **Mycobacterium species Kms.**
* **Mycobacterium avium subspecies Paratuberculosis.**
* Species with one rhomboid.
NS: Not Significant (< 10% identity).

Mycobacterial rhomboids are active rhomboid-serine-proteases

Multiple sequence alignment revealed that all mycobacterial rhomboids contain the putative rhomboid catalytic residues Gly199, Ser201 and His254. The mycobacterial rhomboids also contained two additional C-terminal Histidins (His145 and His150, which together with His254 are universally conserved in the rhomboid proteins [19]) and five invariant transmembrane residues (Gly202, Gly257, Gly261, Asn154 and Ala200) that are also conserved in many rhomboid proteins [33]. However in mycobacteria, Ala252 which occurs in many eukaryotic and prokaryotic rhomboids was substituted by Gly (figure 4). Furthermore, Tyr205 which stabilizes the rhomboid protease active site of glpG [17,33] and of many rhomboid proteases was only conserved in MAB_0026 of M. abscessus, being substituted by Phe in mycobacterial rhomboids (figure 4). Thus, Phe is the stabilizing residue in the protease active site for majority of mycobacterial rhomboids (Phe is an additional stabilizing residue for rhomboid proteases [17]).

The nature of the transmembrane helices (TMHs) formed by mycobacterial rhomboids was analyzed to determine whether they conform to those of active mycobacterial orthologs contained extra eukaryotic motifs and have topologies similar to that of rho-1 of drosophila. Koonin et al [19] alluded that rhomboids could have emerged in a bacterial lineage and were eventually widely disseminated (to other life kingdoms) by horizontal transfer [19]. Conversely, the Rv1337 mycobacterial orthologs (boxed red) formed a distinct clade, different from Rv0110 mycobacterial orthologs. These rhomboids appeared evolutionary stable and did not cluster with eukaryotic rhomboids.

MAB_0026 of M. abscessus which had low homology with Rv0110 also appeared distant and clustered poorly with mycobacterial orthologs, in contrast with its paralog MAB_1481 (figure 3A). Since orthologs have an ancestral gene in the last common ancestor [47], MAB_0026 could be a “pseudoortholog” (i.e. it is a distant paralog that appears orthologous due to differential, lineage-specific gene loss [47]). In phylogenetic analysis of mycobacterial rhomboids orthologous to Rv0110, MAB_0026 was also distant from rhomboids of other actinobacteria (figure 3B). Since M. abscessus is one of the earliest species to diverge of all mycobacterial species [39], the low homology could reflect evolutionary distance or stability of this rhomboid. However, the high homology of MAB_1481 (62% identity with Rv1337) contrasts the low homology of MAB_0026 (38% identity with Rv0110), negating the notion of evolutionary distance and instead favors evolutionary stability of MAB_0026.
Figure 3 Mycobacterial rhomboids have different evolutionary history

A: Mycobacterial rhomboids clustered into two distinct clades (boxed blue and red). The Rv0110 mycobacterial orthologs (boxed blue) clustered with eukaryotic active rhomboids (unboxed). The Rv1337 mycobacterial orthologs could have been acquired at the same time, and the orthologs of Rv0110 were eventually lost in the MAC species and M. leprae. Mouse-protein farnesyl transferase, FT, [GenBank: AAI38303] was the outgroup.

B: MAB0026 of M. abscessus (underlined blue) is conspicuously distant from its mycobacterial orthologs (boxed blue).
rhomboid proteases. Mycobacterial orthologs of Rv0110 formed seven TMHs and topologies similar to those of eukaryotic rhomboid rho-1 of *Drosophila* (see figure 5). As in rho-1, the rhomboid catalytic residues GxSx & H (Gly199, Ser201 and His254, × being any residue) were localized respectively, in TMH4 and TMH6 (see figure 5 and details in additional file 1). In mycobacterial orthologs of Rv0110, the two C-terminal histidine and asparagine (His145, His150 and Asn154) were localized in TMH2, in contrast to eukaryotic rhomboid proteases which have these residues in TMH3 [17,19,23]. However, in our analyses, we found His145, His150 and...
Asn154 in TMH2 in rho-1, similar to Rv0110 (see additional file 2). Despite the proteins being evolutionary diverse, other studies found the overall structure of TMHs of rhomboid proteases conserved, with eukaryotic rhomboid proteases containing seven TMHs while archaea and eubacteria contain six [23,49]. It is not clear whether these similarities infer evolutionary or functional significance; similar topologies with eukaryotic rhomboids could imply occurrence of a common bacterial universal progenitor for the eukaryotic rhomboids [19]. Nevertheless, prokaryotic and eukaryotic integral transmembrane proteins can have similar architecture, with striking similarity in the amino acid frequency distribution in their TMHs [50].

In contrast, the mycobacterial orthologs of Rv1337 formed either six or five TMHs, as observed in most bacterial and archaeal rhomboids [19]. The orthologs of pathogenic mycobacteria formed six TMHs, while those of non-pathogenic mycobacteria formed five (see figure 5). The Gx5x and H catalytic residues were found respectively, either in TMH4 and TMH6 (for Rv1337 orthologs of pathogenic mycobacterial with six TMHs -see details in additional file 3) or in TMH3 and TMH5 (for Rv1337 orthologs of non-pathogenic mycobacterial with five TMHs, see additional file 4). The mycobacterial orthologs with six TMHs had the two C-terminal His and Asn residues in TMH2, as in the Rv0110 orthologs; however, in the orthologs with five TMHs, these residues were outside the TMHs (see additional file 4). Although His145, His150 and Asn154 are not essential for catalytic activity [33], it is not clear whether their absence in TMHs can affect functionality. This seems unlikely in that functions have been ascribed to the catalytically inert eukaryotic iRhoms lacking the minimum catalytic sites [26,27]. Alternatively, the observed differences may imply functional divergence, with rhomboids of pathogenic mycobacteria being functionally different from those of non-pathogenic mycobacteria. Indeed, Rv1337 was essential for the survival of the tubercle bacilli in macrophages [38]. Nevertheless, experimental evidence will be necessary for validation of these assertions.

Figure 5 The topology of mycobacterial rhomboids. Boxed (yellow) are the transmembrane domains containing the rhomboid catalytic residues and locations for the C-termini conserved residues. The Rv0110 mycobacterial orthologs formed topologies similar to those of the secretase eukaryotic rhomboid rho-1. The Rv1337 mycobacterial orthologs formed either six or five TMHs. The orthologs of pathogenic mycobacteria formed six TMHs while the orthologs of non-pathogenic mycobacteria formed five TMHs.
Extra protein domains in mycobacterial rhomboids

Mycobacterial rhomboids contained extra protein motifs, many of which were eukaryotic. The orthologs of Rv0110 contained diverse eukaryotic motifs, while the Rv1337 orthologs maintained a fairly constant number and type of motifs, either fungal cellulose binding domain or bacterial putative redox-active protein domains (table 2). It is difficult to account for the origin of eukaryotic motifs in mycobacterial rhomboids; nevertheless, extra protein motifs are common in eukaryotic rhomboids where their significance is also not known [17]. Since eukaryotic rhomboids are presumed to have been acquired from bacteria through horizontal gene transfer mechanisms [19], the extra protein motifs may

| Species/strain | Rhomboid | Number of aTMHs | TMH with active Site | Extra motif | E-value | Target signal |
|---------------|----------|-----------------|----------------------|-------------|---------|--------------|
| H37Rv         | Rv0110   | 7               | 4 & 6                | DUF1751     | 0.27    | Mitochondrial |
|               |          |                 |                      | Sva³        | 0.68    |              |
|               |          |                 |                      | ZF-B_box³   | 0.00021 |              |
| M. marinum    | MMAR_0300| 7               | 4 & 6                | ZF-B_box    | 0.00012 | Other        |
|               |          |                 |                      | FixQ⁴       | 0.016   |              |
| M. ulcerans   | MUL_4822 | 7               | 4 & 6                | EcsB⁵       | 0.17    | Mitochondrial |
| M. sp Jls     | Mjls_5528| 7               | 4 & 6                | IBR⁶        | 0.301   | Other        |
|               |          |                 |                      | ZF-B_box    | 0.013   |              |
|               |          |                 |                      | Dynactin p62²| 0.24    |              |
|               |          |                 |                      | Tim1⁷⁸      | 0.36    |              |
| M. vanbaalenii| Mvan_5753 | 7           | 4 & 6                | ZF-B_box    | 0.0044  | Other        |
|               |          |                 |                      | Dynactin p62| 0.11    |              |
|               |          |                 |                      | DUF1751     | 0.028   |              |
| M. gilvum     | Mflv_1071| 7               | 4 & 6                | ZF-B_box    | 0.015   | Other        |
|               |          |                 |                      | DUF1751     | 0.02    |              |
| M. smegmatis  | MSMEG_5036| 7           | 4 & 6                | -           |         | Mitochondrial |
| M. abscessus  | MAB_0026 | 7               | 4 & 6                | ZF-B_box    | 0.0064  | Other        |
| H37Rv         | Rv1337   | 6               | 4 & 6                | C_BB3     | 0.17    | Mitochondrial |
| M. marinum    | MMAR_4059| 6               | 4 & 6                | C_GCAxxG_C_C | 0.0626 | Secretory    |
| M. avium      | MAV_1554 | 6               | 4 & 6                | C_GCAxxG_C_C | 0.0099 | Secretory    |
| M. leprae     | ML1171   | 6               | 4 & 6                | C_GCAxxG_C_C | 0.031  | Other        |
| M. abscessus  | MAB_1481 | 6               | 4 & 6                | -           |         | Other        |
| M. smegmatis  | MSMEG_4904| 5           | 3 & 5                | C_GCAxxG_C_C | 0.025  | Secretory    |
| M. sp Jls     | Mjls_3833| 5               | 3 & 5                | DUF2154     | 0.6     | Secretory    |
| M. vanbaalenii| Mvan_4290| 5               | 3 & 5                | -           |         | Secretory    |
| M. gilvum     | Mflv_2355| 5               | 3 & 5                | -           |         | Secretory    |

The rhomboid family domain was excluded.

- Extra domain not detected.

Other: cellular localization target other than secretory and mitochondrial.

1: Transmembrane helices.
2: Mycobacterium tuberculosis.
3: Mycobacterium species Jls.
4: Eukaryotic integral membrane protein.
5: Cd27 binding protein.
6: B-box zinc finger.
7: Cbb3-type cytochrome c oxidase component.
8: Bacterial ABC transporter protein.
9: In Between Ring IBR fingers.
10: Dynactin p62 family.
11: Tim17/Tim22/Tim23 family.
12: Fungal cellulose binding domain.
13: Putative redox-active protein.
14: Predicted membrane protein.
have originated from prokaryotic progenitors. Mycobacterial rhomboids also contained N-signal peptides and eukaryotic subcellular localization target signals which were either mitochondrial or secretory (see table 2), with scores higher than or comparable to those of rho-7 and PARL. These observations further allude to a common ancestor for mycobacterial and eukaryotic active rhomboids [17].

A novel nonsense mutation at the Trp73 codon split the MAP rhomboid into two hypothetical proteins

The annotated rhomboid of *M. avium* subsp. *Paratuberculosis* (MAP) in the genome databases appeared truncated; MAP_2425c (hypothetical protein) was significantly shorter than MAV_1554 of genetically related *M. avium* (147 vs. 223 residues, respectively). Upstream of MAP_2425c was MAP_2426c (74 residues), similar to the amino-terminal portion of MAV_1554 (100% identity) while the former (MAP_2425c) was similar to the carboxyl-terminal portion of MAV_1554 (100% identity). MAP_2425c and MAP_2426c were separated by 10 bp that translate into three residues (Gln, His and Lys, present in similar location in MAV_1554) and a stop codon TGA, at nucleotide position 217, which split the homolog into two ORFs. Because MAP and *M. avium* are genetically related, initially, we thought MAP2425c and MAP2426c are truncated portions (resulting from genome annotation errors) and should have been a whole rhomboid of MAP. Thus, we aimed to determine the correct annotation for the MAP rhomboid. Using MAV_1554 specific primers, we PCR-amplified and sequenced homologs of MAP2425c and MAP2426c (954 bp) from a cattle isolate of MAP (strain 27, see table 3); the amplicon was similar to MAP2426c and MAP2425c (containing an internal stop codon TGA at nucleotide positions 217-219, and 10 bp translating into residues Gln, His and Lys, in similar location as those of MAV1554). Thus, we confirmed the annotations for MAP2425c (hypothetical protein) and MAP2426c (hypothetical protein). It was later revealed that a nonsense mutation at nucleotide positions 217-219 (formerly TGG, the codon for Trp73), substituted guanine at the wobble position for adenine, creating a stop codon (i.e. TGG[Trp73]® TGA[stop

Table 3 Features of PCR-amplified mycobacterial rhomboids

| Primer | Species/Strain | Amplicon size (bp) | ORF (bp) | Amino acids | Accession number | Protein ID |
|--------|----------------|--------------------|----------|-------------|-----------------|------------|
| 0110F  | aH37Rv         | 967                | 855      | 284         | HM453890        | ADO17908   |
| 0110R  | bBCG           | 967                | 855      | 284         | HM453894        | ADO17912   |
|        | cJN55          | 967                | 855      | 284         | HM453896        | ADO17914   |
|        | dBN44          | 967                | 855      | 284         | HM453892        | ADO17910   |
| 5036F  | eSMR5          | 1000               | 891      | 296         | HM453900        | ADO17919   |
| SMR5R  | aH37Rv         | 967                | 855      | 284         | HM453890        | ADO17908   |
| 1337F  | bH37Rv         | 869                | 723      | 240         | HM453891        | ADO17909   |
| 1337R  | cBCG           | 869                | 723      | 240         | HM453895        | ADO17913   |
|        | dJN55          | 869                | 723      | 240         | HM453897        | ADO17915   |
|        | eBN44          | 869                | 723      | 240         | HM453897        | ADO17911   |
| 1554F  | fSU-36800      | 954                | 672      | 223         | HM453898        | ADO17916   |
| 1554R  | g27            | 954                | 291      | 72          | HM453899        | ADO17917   |
|        | (MAP2425c)     |                    |          |             |                 | ADO17917   |
|        |                | 444                | 147      |             | HM453899        | ADO17917   |
| 4904F  | hSMR5          | 845                | 738      | 245         | HM453901        | ADO17920   |
| 4904R  |                |                    |          |             |                 | ADO17920   |

a: *M. tuberculosis*.  
b: *M. bovis*.  
c: *M. bovis* (cattle isolate).  
d: *M. tuberculosis* (patient isolate).  
e: *M. smegmatis* (streptomycin resistant derivative of MC*strain 155*).  
f: *M. avium* (patient isolate).  
g: *M. avium* subsp. *Paratuberculosis* (cattle isolate).  
h: Accession numbers are for GenBank.  

Primer sequences are described in methods.
codon]. Usually, nonsense mutations disrupt ORFs resulting in truncated and non-functional proteins; however, this rare scenario resulted into two unique ORFs of MAP, providing the first evidence of a split rhomboid, contrasting whole orthologs of genetically related species. Although the significance of this is currently not known, cDNA was amplified from both ORFs, implying that both hypothetical proteins may be expressed (see figure 6).

**What are the lengths of MTC rhomboids?**

In genome databases, the lengths for annotated sequences of rhomboids from genetically related mycobacteria vary, and initially we thought this reflected strain diversity. For instance, lengths for Rv0110 orthologs of MTC species are either 249 or 284 residues, while Rv1337 orthologs from the same species are 240 residues. In contrast, MT1378 (ortholog of Rv1337) of *M. tuberculosis* CDC 1551 is 227 amino acids, 13 residues shorter at the NH2-terminus. Thus, we aimed to validate the sizes of rhomboids from related strains/species. Genomic analyses at the rhomboid loci for the sequenced MTC genomes revealed that MTC rhomboid orthologs are 100% identical and are of equal length. Rhomboids were PCR-amplified from MTC with common primer sets for each ortholog (see methods), and sequencing data confirmed that MTC rhomboid orthologs are identical and are of the same size (284 residues for Rv0110 orthologs and 240 residues for Rv1337 orthologs). Rhomboid sequences were deposited in GenBank and accession numbers were assigned (see table 3).

**Putative gene clusters for mycobacterial rhomboids**

To determine putative functional coupling between mycobacterial rhomboids and other genes, genes in clusters formed by mycobacterial rhomboids at the KEGG database [51] were analyzed. The gene cluster formed by Rv1337 was conserved across the genus and extended to other actinobacteria such as *Norcardia* and *Corynebacteria*. This cluster included 58 genes (Rv1311 to Rv1366, see additional file 5) of which some are essential and others are required for the growth of *M. tuberculosis* in macrophages [38], a necessary step during pathogenesis of the tubercle bacillus. Conversely, the Rv0110 orthologs formed clusters reflecting the genetic relatedness of mycobacteria. Thus, the orthologs from MTC species and *M. marinum* formed similar clusters consisting of 61 genes (Rv0080 to Rv0140, see additional file 6). These clusters also included essential genes and those required for survival of the tubercle bacillus in macrophage. However, MUL_4822 of genetically related *M. ulcerans* was not included in the MTC/*M. marinum* cluster, and formed a unique cluster consisting of only 19 genes (MUL_4791 to MUL_4824) with two genes upstream of the rhomboid (MUL_4823 and MUL_4824, see additional file 7). It is not certain whether this
reflects functional divergence of MUL_4822 from Rv0110, in spite of evolutionary relatedness of M. ulcerans and MTC species.

The gene cluster of Rv0110 orthologs of M. vanbaalenii, M. gilvum and Mycobacterium species Jls, Kms and Mcs were also similar, and consisted of 48 genes (Mjls_5512 to Mjls_5559, see additional file 8), whose orthologs in MTC species are required for the growth of the tubercle bacillus in macrophages [38]. Conversely, the cluster for MAB_0026 of M. abscessus consisted of only three genes (MAB_0024, MAB_0025 and MAB_0026), shared with actinobacteria other than mycobacteria. Many MTC orthologs in the gene clusters of MUL_4822, Mjls_5529 and MAB_0026 are required for the growth of the bacillus in macrophages, the implication of which requires further study. There was no gene cluster formed by MSMEG_5036 of M. smegmatis. The essential genes in mycobacterial rhomboid gene clusters are described in additional file 9.

Transcription analysis
Due to their ubiquity in eubacteria, we aimed to determine the expression of mycobacterial rhomboids in a preliminary fashion by screening for in vivo transcription. RT- (Reverse Transcriptase) PCRs amplified rhomboid cDNAs from mycobacterial mRNA, indicating that both copies of mycobacterial rhomboids are transcribed, and possibly expressed (see figure 6).

Functional insights
Signal transduction and Metabolite transport
Since mycobacterial rhomboids contain rhomboid catalytic signatures, they may be functionally similar to aarA and rho-1, rescuing phenotypes associated with deletion of these genes in P. stuartii and D. melanogaster rhomboid mutants [52]. Due to their diverse functions, rhomboids appear good candidates for investigation in studies elucidating inter/intra-species/kingdom signaling mechanisms [29,53-55].

Furthermore, gluP (contains a rhomboid domain) of B. subtilis is involved in sugar transport [17,32], while aarA activates the TatA protein transporter in P. stuartii [31]. As such, the putative gene clusters for mycobacterial rhomboids contained putative metabolite transporters and transcriptional regulators. Since genes in clusters for transport and signal transduction genes tend to have similar roles [56], mycobacterial rhomboids may have such roles.

Roles in pathogenesis?
In a TraSH analysis by Rengarajan et al, Rv1337 was required for the survival of M. tuberculosis H37Rv in macrophages [38], a necessary step during the development of TB. The genome wide conservation of Rv1337 alludes to a possibly important protein. The pathogenesis of M. ulcerans, (the only mycobacterium lacking the Rv1337 ortholog) is known and it culminates in skin ulcerations caused by the plasmid encoded polyketide toxin -mycolactone [4,40,44,57]. Buruli ulcer contrasts with the tuberculous nature of lesions formed by many pathogenic mycobacteria, whose pathogenesis is not well understood and remains a vast field of study.

Moonlighting properties?
It is possible to predict functional coupling between genes based on conservation of gene clusters among genomes [56,58]. Since proteins encoded by conserved gene pairs appear to interact physically [58], the evolutionary conservation of the Rv1337 genome arrangement might have functional implications. mur1 is a moonlighting protein (ability to perform multiple independent functions [59]) that exhibits both racemization and DNA gyrase activities [59]. Since rhomboids are known for diverse functions, the proximity of Rv1337 orthologs with a moonlighting protein makes them suspects for moonlighting properties.

Conclusions
Mycobacterial rhomboids have different evolutionary history
The two mycobacterial rhomboids are phylogenetically distinct and could have been acquired independently. The mycobacterial orthologs of Rv0110 (rhomboid protease 1) appear to be under evolutionary pressure; hence they were lost in the MAC species and M. leprae. These orthologs represent prokaryotic rhomboids whose progenitor may be the ancestor for euukaryotic rhomboids. The Rv1337 (rhomboid protease 2) mycobacterial orthologs appear more stable and are conserved nearly in all mycobacteria, possibly alluding to their importance in mycobacteria.

MAP2425c and MAP2426c provide the first evidence of a split rhomboid contrasting whole orthologs of genetically related species.

Mycobacterial rhomboids are active rhomboid proteases
Mycobacterial rhomboids are active rhomboid proteases, with the active site being stabilized by Phe. Although valuable insights to the roles of rhomboids are provided, the data herein only lays a foundation for future investigations for the roles of rhomboids in mycobacteria.

Methods
Strains and cultures
Mycobacterium smegmatis SMR5 (streptomycin resistant derivative of MC2155) and M. avium (patient isolate SU-36800) were obtained from the Joint Clinical Research Center (JCRC), Kampala, Uganda. The streptomycin resistant derivatives of M. tuberculosis H37Rv and M. bovis BCG were provided by Dr. Peter Sander,
University of Zurich, Switzerland. *M. tuberculosis* BN44 and *M. bovis* JN55 are characterized clinical isolates [60,61]. *M. avium* subsp. *Paratuberculosis* was provided by Dr. Julius B. Okuni, Faculty of Veterinary Medicine, Makerere University. *M. smegmatis* was cultured in LB/0.05% Tween 80 containing 200 μg/ml streptomycin. MTC and MAC strains were cultured in middlebrook 7H9 or 7H10 (supplemented with mycobactin for MAP cultures).

**PCR conditions**

Chromosomal DNA was extracted from mycobacteria by boiling heat-killed cells for 10 min and centrifuging briefly at 5000 g to obtain the supernatant containing DNA. Amplification reactions contained 20 pmol each of the rhomboid specific forward and reverse primers (see below), 1.5 U of high fidelity *Taq* polymerase (Roche Applied Science, Mannheim, Germany), Custom PCR Master Mix (Thermo Scientific, Surry, UK), ~200 ng template DNA and nuclease-free water in a reaction volume of 10 μL. The reactions were performed in a Peltier thermocycler (MJ Research, Waterman, MA, USA) at the following conditions: initial denaturation at 94°C for 5 min, followed by 30 cycles each consisting of 94°C, 0.5 min; 60°C, 0.3 min & 72°C, 1 min, with a final extension at 72°C for 10 min. Following amplification, the amplicons were purified with QIAquick PCR purification kit (Qiagen, Hilden, Germany) and sequenced at ACGT (Wheeling, IL, USA). After analyzing with BioEdit software and BLAST algorithm for similarity searches, rhomboid sequences were deposited in the GenBank database (see table 3 for accession numbers).

The following primers were used: 0110F, 5'-ATA TTCGGCTTTGGCCGGAAC-3' (forward) and 0110R, 5'-ACGCCGAAAGACGCGGTATCTAC-3' (reverse) for MTC Rv0110 orthologs; 1337F, 5'-CCGACGCCGGAATCAAAGACTC-3' (forward) and 1337R, 5'-CCGACGCCGGGATGACAAAGACTC-3' (reverse) for MTC Rv1337 orthologs. For MAC species, primer pair 1554F, 5'-ACGGCCGGATGACAAAGACTC-3' (forward) and 1554R, 5'-TGCCGAGCTCATGTCTTGGG-3' (reverse) for MTC Rv1554 orthologs; 4904F, 5'-ACGCCGGGATGACAAAGACTC-3' (forward) and 4904R, 5'-ACGCCGGATGACAAAGACTC-3' (reverse) for homolog MSMEG_4904 were used. Primers were synthesized by IDT (Leuven, Belgium).

**Transcription assays**

mRNA was purified from mycobacteria with the Oligotex mRNA mini kit (Qiagen, Hilden, Germany) and ~60 ng/μl (in 15 μl) mRNA used as template for cDNA synthesis. Reverse Transcriptase-PCRs were performed with the Titan One Tube RT-PCR System (Roche Applied Science, Mannheim, Germany) to amplify Rv0110 and Rv1337 cDNAs in separate reactions. Except for the initial cDNA synthesis step (50°C for 30 min), PCR conditions were similar to those described above. RT-PCRs were repeated with primers (1337int1: TGGACGTCAACGCGCATCAG, forward, and 1337int2: CCAGCCCAATGACACTGATCCC, reverse) that amplify an internal fragment (~350 bp) of Rv1337 orthologs.

**Bioinformatic analyses**

**Identification of rhomboids in mycobacteria**

Rhomboid sequences for rho-7 [GenBank: NP_523704.1] of *D. melanogaster*, PARL [GenBank: NP_061092.3] of human, glpG [GenBank: AAA23890] of *E. coli* and aarA [GenBank: L28755] of *P. stuartii* were obtained from GenBank [62]. These sequences were used as queries in BLAST-searches for rhomboid homologs from an array of mycobacterial genome databases: “tuberculist” [63], GIB-DDBJ [64] and J. Craig Venter institute [65].

**Sequence analysis**

The similarity between mycobacterial rhomboids was determined using specialized BLAST b2seq for comparing two or more sequences [66]. Multiple sequence alignments were performed with ClustalW [67] or MUSCLE [68]. Mycobacterial rhomboids were examined for the presence of rhomboid family domains and catalytic signatures (GxSx). The TMH predictions were done using the TMHMM Server v. 2.0 [69]. The data generated was fed into the TMRPres2D [70] database to generate high resolution images. Cellular localization signals were predicted using TargetP 1.1 server [71].

**Phylogenetic analysis**

Phylogenetic analysis was conducted using MEGA4 software [72]. The evolutionary history of mycobacterial rhomboids was determined using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together was determined using the Bootstrap test (1000 replicates). The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). For comparison of evolutionary history, trees were also constructed using “Minimum Evolution” and “Maximum Parsimony”.

**Functional predictions**

To predict possible roles for mycobacterial rhomboids, sequences were analyzed at the KEGG database [51] for the genome arrangement, presence of extra protein domains, nature of gene clusters, orthologs and paralogs. Other parameters used to glean functions from mycobacterial rhomboid sequences included analyzing their
topologies. To predict functional relatedness among genes within mycobacterial rhomboid clusters, sequences in the clusters were aligned by ClustalW, and Neighbor-Joining trees deduced using default settings.

**Additional material**

Additional file 1: The topology and location of catalytic residues in mycobacterial rhomboid protease 1 (Rv0110 orthologs). As in rho-1, the catalytic residues are located in TMH4 (Gly199 and Ser201) and TMH6 (His254), while His145, His150 and Asn154 are in TMH2.

Additional file 2: The topology and location of catalytic residues in rho-1 of Drosophila. As in mycobacterial rhomboid protease 1, the catalytic residues are located in TMH4 (Gly199 and Ser201) and TMH6 (His254), while His145, His150 and Asn154 are in TMH2.

Additional file 3: The topology and location of catalytic residues in mycobacterial rhomboid protease 2 (Rv1337 orthologs). The orthologs of pathogenic mycobacteria formed six TMHs, with catalytic residues in TMH4 (Gly199 and Ser201) and TMH6 (His254), His145, His150 and Asn154 are located in TMH2 as in rhomboid protease-1 (Rv0110 orthologs).

Additional file 4: The topology and location of catalytic residues in mycobacterial rhomboid protease 2 (Rv1337 orthologs) of nonpathogenic mycobacteria. These rhomboids formed five TMHs, with catalytic residues in TMH4 (Gly199 and Ser201) and TMH5 (His254), while His145, His150 and Asn154 are outside the TMHs (boxed).

Additional file 5: ClustalW-Neighbor Joining analysis of the genes in Rv1337 cluster. Boxed (blue) are the genes that grouped with Rv1337. Essential genes in this clade are Rv1327c, Rv1327c, Rv1331, Rv1340 and Rv1344.

Additional file 6: ClustalW-Neighbor Joining analysis of the genes in Rv0110 cluster. Boxed (blue) are the essential genes in that grouped with Rv0110 (Rv0118c, Rv0127, Rv0107c, Rv0116c, Rv0121c, Rv0121c, Rv0133 and Rv0139).

Additional file 7: ClustalW-Neighbor Joining analysis of the genes in MUL4822 cluster. Boxed (blue) are the genes that grouped with MUL4822. Several of the MTC orthologs in this clade are essential for the growth of M. tuberculosis in macrophages.

Additional file 8: ClustalW-Neighbor Joining analysis of the genes in MJs5529 cluster. Boxed (blue) are the genes that grouped with MJs5529, whose homologs are essential in M. tuberculosis. Several of the MTC orthologs in this clade are essential for the growth of M. tuberculosis in macrophages.

Additional file 9: The essential genes in mycobacterial rhomboid gene clusters (doc). 1 According to Sassetti et al [37] and Renganaran et al [38]. 1 Essential (for optimal growth). 2 Required for growth in macrophage. 3 Mutation slows growth.

**References**

1. Ezureby JP. List of Prokaryotic names with Standing in Nomenclature. [http://www.bacterio.cict.fr/my/mycobacterium.html].

2. Cole ST, Boshch R, Parkhill J, Generit T, Chamber C, Harris D, Gordon SV, Ewigmar K, Gas S, Barry CE, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 1998, 393(6685):537-544.

3. Cole ST, Ewigmar K, Parkhill J, Johnson KD, Thompson NV, Wheeler PR, Online N, Generit T, Chamber C, Harris D, et al. Massive gene decay in the legioso bacillus. Nature 2001, 409(6823):1007-1011.

4. Demangel C, Sinner TP, Cole ST. Buruli ulcer: reductive evolution enhances pathogenicity of Mycobacterium ulcerans. Nat Rev Microbiol 2009, 7(15):60.

5. Bannantine JP, Barletta RG, Stabel JR, Paustian ML, Kapur V. Application of the Genome Sequence to Address Concerns That Mycobacterium avium subspecies paratuberculosis Might Be a Foodborne Pathogen. Foodborne Pathogens and Disease 2004, 1(1):3-15.

6. Rubin DS, Rahal JJ. Mycobacterium-avium complex. Infect Clin North Am 1994, 8(2):413-426.

7. Valentijn-Weigand P, Goether P. Pathogenesis of Mycobacterium avium subspecies paratuberculosis infections in ruminants: still more questions than answers. Microbes Infect 1999, 1(3):1121-1127.

8. Ventura M, Chanchay A, Tauch A, Chandra G, Fitzgerald GF, Chater KF, van Sinderen D. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev 2007, 71(3):495-548.

9. Khan AA, Kim SJ, Paine DD, Cargnello CE. Classification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, Mycobacterium sp. strain PYR-1, as Mycobacterium vanbaalenii sp. nov. Int J Syst Evol Microbiol 2002, 52(1):699-2007.

10. Bordin P, Rosenkinds D, Andersen P, Cole ST, Boshch R: ESAT-6 proteins: protective antigens and virulence factors? Trends in microbiology 2004, 12(1):50-58.

11. Chen JM, Islam ST, Ren H, Liu J. Differential productions of lipid virulence factors among BCG vaccine strains and implications on BCG safety. Vaccine 2007, 25(48):8114-8122.
25. Brossier F, Jewett TJ, Sibley LD, Urban S: Proteolysis and Toxoplasma invasion. *Int J Parasitol Pathog* 2009, 8(2):595-600.

26. Dowse TJ, Pascall JC, Brown KD, Soldati D: Membrane-embedded protease poses for the rhomboids: a nearly ubiquitous family of intramembrane serine proteases. *Nat Rev Genet* 2003, 4(11):912-922.

27. Rather PN, Orosz E: Functional characterization of a novel gene, gluP, is preferentially cleaved different adhesins implicated in all invasive stages of malaria. *PLoS Pathog* 2006, 2(10):e113.

28. Caruthers VB: Proteolysis and Toxoplasma invasion. *Int J Parasitol* 2006, 36(5):595-600.

29. Dowse TJ, Pascall JC, Brown KD, Soldati D: Apicomplexan rhomboids have a potential role in microneme protein cleavage during host cell invasion. *Int J Parasitol* 2005, 35(7):747-756.

30. Srinivasan P, Coppins I, Jacobs-Lorena M: Distinct roles of Plasmodium rhomboid 1 in parasite development and malaria pathogenesis. *PLoS Pathog* 2009, 5(1):e1000262.

31. Brossier F, Jezewski T, Sibley LD, Urban S: A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. *Nat Protoc* 2005, 10(2):414-415.

32. Yan Z, Zou H, Tian F, Grands JR, Mxwson AJ, Lu PY, Li LY: Human rhomboid family-1 gene silencing causes apoptosis or autophagy to epithelial cancer cells and inhibits xenograft tumor growth. *Mol Cancer Ther* 2008, 7(6):1355-1364.

33. Zou H, Thomas SM, Yan ZW, Grands JR, Vogt A, Li LY: Human rhomboid family-1 gene RHBDF1 participates in GPCR-mediated transactivation of EGFR growth signals in head and neck squamous cancer cells. *FASEB J* 2009, 23(3):1243-1252.

34. Waters CM, Basler BL: Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* 2003, 19:219-240.

35. Fedorova NJ, Basler BL: Interspecies communication in bacteria. *J Clin Invest* 2003, 112(9):1291-1299.

36. Rather PN, Orosz E: Characterization of arA, a pleiotropic negative regulator of the 2′-N-acetyltransferase in Providencia stuartii. *J Bacteriol* 1994, 176(16):5140-5144.

37. Mesarik J, Mesarik FM, Dahl MK: Expression of a novel gene, ghdp, is essential for normal Bacillus subtilis cell division and contributes to growth export. *BMC Microbiol* 2004, 4:13.

38. Clemmer KM, Sturgill GM, Veenstra A, Rather PN: Functional characterization of Escherichia coli GlpG and additional rhomboid proteases using an arA mutant of Providencia stuartii. *J Bacteriol* 2006, 188(9):3415-3419.

39. Wu Z, Yan N, Feng L, Obertstein A, Yan H, Baker RP, Gu L, Jeffrey PD: Urban S, Shi Y: Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* 2006, 13(12):1084-1091.

40. Liebermann RL, Wolfe MS: Membrane-embedded protease poses for phoshoot. *Proc Natl Acad Sci USA* 2007, 104(21):8737-8742.

41. Lemberg MK, Freeman M: Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res* 2007, 17(11):1634-1646.

42. Sassetti CM, Boyd DH, Rubin EJ: Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci USA* 2001, 98(22):12712-12717.
from cattle carcases at a city slaughterhouse in Uganda. *Vet Rec* 2009, 164(21):655-658.

61. Asiimwe BB, Kaivula T, Illienius G, Huard RC, Ghebremichael S, Asiimwe J, Joloba ML. *Mycobacterium tuberculosis* Uganda genotype is the predominant cause of TB in Kampala, Uganda. *The International Journal of Tuberculosis and Lung Disease* 2008, 12:386-391.

62. NCBI: National Center for Biotechnological Information. [http://www.ncbi.nlm.nih.gov/]

63. TuberculList-GenoList. [http://genolist.pasteur.fr/TuberculList/]

64. GB: Genome Information Broker. [http://gb.genes.nig.ac.jp/]

65. JCVI: J Craig Venter Institute. [http://cmr.jcvi.org/cgi-bin/CMR/CmrHomePage.cgi]

66. Specialized BLAST: Align two or more sequences. [http://blast.ncbi.nlm.nih.gov/blast.cgi]

67. ClustalW. [http://www.ebi.ac.uk/clustalw/]

68. MUSCLE: Multiple Sequence Comparison by Log-Expectation. [http://www.ebi.ac.uk/Tools/muscle/index.html]

69. ExPASy Proteomics Server. [http://www.cbs.dtu.dk/services/TMHMM-2.0/]

70. TMRRRes2D Tool. [http://bioinformatics.bioc.uoa.gr/TMRRes2D/]

71. ExPASY Tools. [http://www.cbs.dtu.dk/services/TargetP/]

72. MEGA 4: Molecular Evolutionary Genetics Analysis. [http://www.megasoftware.net/]

Cite this article as: Kateete et al. Rhomboid homologs in mycobacteria: insights from phylogeny and genomic analysis. *BMC Microbiology* 2010 10:272.