Chemistry & Biology
Article

A Comparative Lipidomics Platform for Chemotaxonomic Analysis of *Mycobacterium tuberculosis*

Emilie Layre,[1] Lindsay Sweet,[1] Sunhee Hong,[1] Cressida A. Madigan,[1] Danielle Desjardins,[1] David C. Young,[1] Tan-Yun Cheng,[1] John W. Annand,[1] Keunpyo Kim,[2] Isdore C. Shamputa,[3] Matthew J. McConnell,[1] C. Anthony Debono,[1] Samuel M. Behar,[1] Adriaan J. Minnaard,[4] Megan Murray,[5] Clifton E. Barry, III,[3] Isamu Matsunaga,[6] and D. Branch Moody[1,*]

[1]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[2]Biostatistics, MedImmune, One MedImmune Way, Gaithersburg, MD 20878, USA
[3]Tuberculosis Research Section, Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[4]Stratingh Institute for Chemistry, University of Groningen, 9747 AG, Groningen, The Netherlands
[5]Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA
[6]Laboratory of Cell Regulation, Department of Viral Oncology, Institute for Virus Research Kyoto University, Kyoto 606-8507, Japan
*Correspondence: bmoody@rics.bwh.harvard.edu
DOI 10.1016/j.chembiol.2011.10.013

SUMMARY

The lipidic envelope of *Mycobacterium tuberculosis* promotes virulence in many ways, so we developed a lipidomics platform for a broad survey of cell walls. Here we report two new databases (MycoMass, MycoMap), 30 lipid fine maps, and mass spectrometry datasets that comprise a static lipidome. Further, by rapidly regenerating lipidomic datasets during biological processes, comparative lipidomics provides statistically valid, organism-wide comparisons that broadly assess lipid changes during infection or among clinical strains of mycobacteria. Using stringent data filters, we tracked more than 5,000 molecular features in parallel with few or no false-positive molecular discoveries. The low error rates allowed chemotaxonomic analyses of mycobacteria, which describe the extent of chemical change in each strain and identified particular strain-specific molecules for use as biomarkers.

INTRODUCTION

Mycobacterium tuberculosis elaborates one of nature's most complex lipid envelopes, which forms a barrier with the human host. This multilayered cell wall contains an inner phospholipid bilayer and an outer layer of α-alkyl, β-hydroxy mycolic acids and other long-chain lipids (Daffé and Draper, 1998; Hoffmann et al., 2008; Zuber et al., 2008). During a decades-long infection cycle, this unusually thick and hydrophobic barrier controls import of essential host metabolites and passage of antitubercular drugs (Adams et al., 2011), and it releases lipid antigens and adjuvants into the host (Geisel et al., 2005). During infection (Kondo et al., 1970), nutrient deprivation (Rustad et al., 2008; Singh et al., 2009), and genetic regulation (Raman et al., 2006), mycobacteria respond by broadly reorganizing their cell walls, providing a need for systems biology approaches to globally measure bacterial responses. This goal has been achieved for transcripts (Homolka et al., 2010; Rohde et al., 2007; Schnappinger et al., 2003) and proteins (Kruh et al., 2010) and has been partially achieved for the cytosolic metabolites that do not form membranes (de Carvalho et al., 2010; Marrero et al., 2010).

For lipids, sensitive mass spectrometry and nuclear magnetic resonance profiling methods are emerging that nearly simultaneously detect many types of lipids. These methods sensitively detect several previously identified lipid families whose mass/charge ratios (*m/z*) (Jain et al., 2007; Matsunaga et al., 2004; Sartain et al., 2011) or nuclear magnetic resonance (NMR) signals (Mahrous et al., 2008) match predefined values. An ideal lipidomics system would offer both broad coverage of many thousands of molecules in mycobacterial lipidomes as well as the ability to convert any unnamed compounds of known mass to named compounds through accurate mass retention time (AMRT) databases or collisional mass spectrometry. In addition, bioinformatics methods for organizing and comparing all lipids among two bacteria or two bacterial states are needed. Mycobacterial lipids, especially large uncharged lipids associated with the mycolate layer, differ from well-studied anionic phospholipids of model organisms with regard to mass and their ionization properties in mass spectrometry. Therefore, approaching this goal required development of new mycobacteria-specific mass spectral databases and software protocols for automated ion finding, as well as broadly separating chromatography optimized for unusually hydrophobic lipids associated with the mycolate membrane.

Here we report two new mycobacterial databases (MycoMass, MycoMap), an integrated set of software methods, and a universal separation method, which, when coupled with collisional mass spectrometry, meet most of these goals. Using an extract of lipids from one bacterial strain taken at one point in time, these methods provide a snapshot profile of more than 5,000 molecular features, which approach the goal of solving one static lipidome. Further, comparative lipidomics

seeks to measure lipid changes in an organism-wide basis as biological events unfold over time or to broadly characterize the molecules that differ between any two bacteria (chemotaxonomy). These kinds of applications require that many lipidomes be generated in a short period of time and bioinformatic methods for aligning mass spectrometry signals across multiple datasets to identify the subset of changed molecules. Taking advantage of stringent data filters that generated a low false-positive rate, this comparative lipidomics platform could reliably detect thousands of molecular changes after infection or among mycobacterial strains, enabling chemotaxonomic analyses of experimental and clinical strains of mycobacteria.

RESULTS

The MycoMass Database

To compile the expected *m/z* from known mycobacterial lipids, we first created an inventory of literature reports relating to *M. tuberculosis* and other medically important mycobacteria. Lipids were organized according to LIPID MAPS criteria, using a four-level classification divided into seven categories and 23 classes based on generic structures, from which 43 subclasses and 58 families were defined based on differences in the number or nature of carbohydrate or lipid moieties (Figure 1) (Fahy et al., 2005). Each lipid family contained an average of 95 alkylforms, which are individual molecules differing in the length and saturation of fatty acyl or polyketide backbones. The resulting MycoMass database catalogs more than 5,000 neutral species (M) and their deduced negative and positive ions, for a total of 32,438 entries (Figure S1 available online).

Like the independent effort of Sartain et al. (2011), MycoMass represented the first step toward a mycobacterial lipidome, because it organized and defined the scope of known lipid families requiring detection. Second, MycoMass also served as the input data that allow for mass comparison with detected molecules to support development of new automated ion annotation protocols. Third, because more than 40 of the 58 cataloged lipid families are lacking in eukaryotic cells or Gram-negative bacteria, MycoMass provides a quantitative summary of the divergence of the mycobacterial lipidome from widely studied model organisms. MycoMass lists mycoketides, phthiocerols, menaquinones, mycolactones, and hydroxyphthioceranic, mycolipanolic, phthienoic, mycolipodienic, mycosanoic, mycocerosic, and mycolic acids, as well as many other lipids found mainly or exclusively in mycobacteria and closely related actinobacteria. Therefore, MycoMass contains specific mass targets that could serve as biomarkers of infected cells. Many neutral lipids are distinct in mass and structure from anionic phospholipids that dominate in Gram-negative bacteria or eukaryotic cells. The divergence of MycoMass entries from databases for model organisms illustrate why the latter, despite their high quality, do not support studies of mycobacterial pathogenesis (Dennis et al., 2010; Quehenberger et al., 2010; van Meer et al., 2007). Also, the unusual hydrophobicity and high mass of mycobacterial neutral lipids necessitated the development of new methods of chromatography and ionization protocols.

Chromatographic Design

We harvested mycobacteria from plates, broth, or infected mice and extracted lipids with chloroform and methanol. Although more complex than simultaneous (i.e., shotgun) ionization, chromatographic separation of lipids in a mixture prior to ionization offers advantages. Column retention predicts polarity of unknown molecules, facilitating their identification. Chromatography separates molecules of similar mass in time, creating a large two-dimensional AMRT area to resolve individual components within mixtures containing thousands of ions. Perhaps the key advantage of chromatography is reducing the chemical dissimilarity of molecules entering the electrospray source at any moment. Dissimilar molecules can dramatically alter the efficiency of electrospray ionization, leading to cross-suppression, a phenomenon that particularly affects apolar lipids that dominate mycobacteria (Taylor, 2005).

Our first generation method used acetone precipitation to separate lipids into batches, followed by several reversed-phase high-performance liquid chromatography (HPLC) methods to optimize separation (Figure S2A). Multiple datasets were then reconciled into one lipidome. As expected, phospholipids and mycolyl glycolipids precipitated (Borgstrom, 1952; Takayama and Armstrong, 1976), but triglycerides, phthiocerol dimycocerosates (PDIMs), and many other lipids partitioned into both phases (Figure S2B). Any biological variable modifying the abundance of individual molecules in the mixture can change the partition coefficient of other lipids. Therefore, fluid-phase separations, despite their wide use for targeted analyses, are unsuitable for lipidomics. This development effort highlights a general difference between lipidomic and typical analytical lipid chemistry problems: rather than optimizing for any single lipid, all new methods must be validated for extremely diverse lipids in one sample.

To avoid partitioning and errors in reconciling many datasets, our second generation replaced five HPLC systems with one single-step method for the analysis of total lipid extracts (Figure 2A, black). The key challenge was to devise a normal-phase chromatography that solubilizes, separates, and allows ionization of highly diverse molecules using one general method. We evaluated each method change on the whole dataset of unnamed molecules (Figure 2B) and four named benchmark lipids of low, intermediate, and high polarity spanning a wide range of signal intensity (10^2–10^7 counts). These benchmark lipids were PDIM, trehalose monomycolate, diacylated sulfoglycolipids, and cardiolipin, representing apolar lipids, glycolipids, sulfolipids, and phospholipids, respectively (Figure 3A). By tracking these, we successfully developed a hexanes/isopropanol/methanol solvent system for normal-phase chromatography that separates families with even density and allowed sensitive detection over a wide dynamic range (Figure 2B). This simplified system reduced lipidome generation time from several days to 45 min, allowing generation of 30 lipidomes in 1 day. This advance fulfilled a crucial performance goal for comparative lipidomics, which requires serial generation of lipidomes in triplicate under rigorously comparable conditions.

Optimizing Lipidomic Detection

Whereas metabolomics conventionally focuses on aqeous cytosolic intermediates that generate energy (de Carvalho et al.,

Category	Main Class	Subclass	Level 4 / "Family" (abbreviated name)	Alkylforms Molecules
Fatty acyls	Fatty acids and Conjugates	straight chain fatty acids	-	11
		branched fatty acids	- mycocerosic acids mycosanoic acids phthioceranic acids	8 6 9 9
		unsaturated fatty acids	- mycolipodienic acid phthienoic acids	18 1 5
		hydroxy fatty acids	hydroxyphthioceranic acid (HPA) mycolipanolic acid	18 1
		mycolic acids	mycolic acids (MA)	150
	Fatty esters	lactones	mycolactones	7
	Fatty acyl glycosides	fatty acyl glycosides of mono- and disaccharides	glucose monomycolates (GMM)	150
Glycerolipids	Monoradylglycerols	monoacylglycerols	monoacylglycerols (MAG) glycerol monomycolates (GroMM)	44 150
	Diradylglycerols	diacylglycerols	diacylglycerols (DAG)	129
	Triradylglycerols	triacylglycerols	triacylglycerols (TAG) monomeromycolyl diacylglycerol	255 81
	Glycosyldiradylglycerols	glycosyldiacylglycerols	glucuronosyl diacylglycerols diglycosylated diacylglycerols	6 5
Glycero-phospholipids	Glycerophosphoglycerols	diacylglycerophosphoglycerols	phosphatidylglycerols (PG) lysinylated diacylglycerophosphoglycerols	129 1
		monoacylglycerophosphoglycerols	lysophosphatidylglycerols	44
	Glycerophosphoinositols	diacylglycerophosphoinositols	phosphatidylinositols (PI)	129
		monoacylglycerophosphoinositols	lysophosphatidylinositols (LPI)	44
	Glycerophosphates	diacylglycerophosphates	phosphatidic acids (PA)	141
		monoacylglycerophosphates	lysophosphatidic acids (LPA)	47
	Glycerophosphoglycero-phosphoglycerols	diacylglycerophospho-glycerophosphodiradylglycerols	cardiolipins (CL)	425
	Glycero-phosphoinositolglycans	monoacylglycerophospho-inositolglycans	lysophosphatidylinositol mannosides (Lyso-PIMx)	12
		diacylglycerophospho-inositolglycans	phosphatidylinositol mannosides (PIMx)	12
		monoacylated diacylglycero-phosphoinositolglycans	monoacyl phosphatidylinositol mannosides ($PIMxAc_1$)	12
		diacylated diacylglycerophospho-inositolglycans	diacyl phosphatidylinositol mannosides ($PIMxAc_2$)	18
	Glycero-phosphoethanolamines	diacylglycerophospho-ethanolamines	phosphatidyl-ethanolamines (PE)	148
		monoacylglycerophospho-ethanolamines	lysophosphatidyl-ethanolamines (LPE)	44
Prenol lipids	Isoprenoids	C20 isoprenoids	decaprenyl phosphoribose (DPPR)	1
		C40 isoprenoids	β-carotene	1
	Quinones and hydroquinones	ubiquinones	menaquinone (MK)	35
	Polyprenols	bactoprenols	undecaprenol	1
Saccharolipids	Acylaminosugars	monoacylaminosugars	glycopeptidolipids (GPL)	720
	Acyltrehaloses	monoacyltrehaloses	trehalose monomycolates (TMM)	150
		diacyltrehaloses	diacyl trehaloses (DAT) trehalose dimycolate (TDM)	131 52
		triacyltrehaloses	triacyl trehaloses (TAT)	255
		polyacyltrehaloses	polyacyl trehalose (PAT)	19
		acylsulfotrehaloses	sulfoglycolipids (Ac_xSGL)	859
	Other saccharolipids		lipooligosaccharides	6
Polyketides	Linear polyketides		mannosylphosphomycoketides	26
			phthiocerol dimycocerosates (PDIM)	179
	Aromatic polyketides	monocyclic aromatic polyketides	monoglycosyl phenylphthiocerol dimycocerosate	157
			diglycosylated phthiocerol dimycocerosates	6
			triglycosyl phenylphthiocerol dimycocerosates (PGL)	142
			parahydroxybenzoic acid derivatives (HBAD)	2
			leprosol	1
	Non-ribosomal peptide/ polyketide hybrids		mycobactins	126
			carboxymycobactins	252
			exochelins	2
Others			lipopentapeptides (L5P)	6
			mycothiol (MSH)	1

Figure 1. MycoMass Database Content

List of the lipids cataloged in the MycoMass database (Figure S1). This database follows the LIPID MAPS organizational tree and uses lipid families' names found in the mycobacterial literature in level 4. Phosphatidylinositol mannosides (PIM_x) contain 1–6 mannosyl residues (x) and sulfoglycolipids (AcxSGL) contain 2–4 fatty acyl chains (x). Alkylforms vary by the saturation and carbon length of acyl chains and/or by the length of carbon backbones.

2010; Früh et al., 2010; Lakshmanan et al., 2011), the emerging specialty of lipidomics provides detailed information about hydrophobic molecules that form membrane barriers. We invented a lipidomics system to investigate how the mycobacterial cell wall, acting as the interface between the host and pathogen, regulates transport of drugs, antigens, and metabolites from the phagosome into the bacterium. Cytosol profiling typically uses ethyl acetate to extract aqueous-soluble compounds, emphasizing somewhat polar molecules of lower mass (50–300 atomic mass units) and rapid turnover (seconds). Lipidomic methods

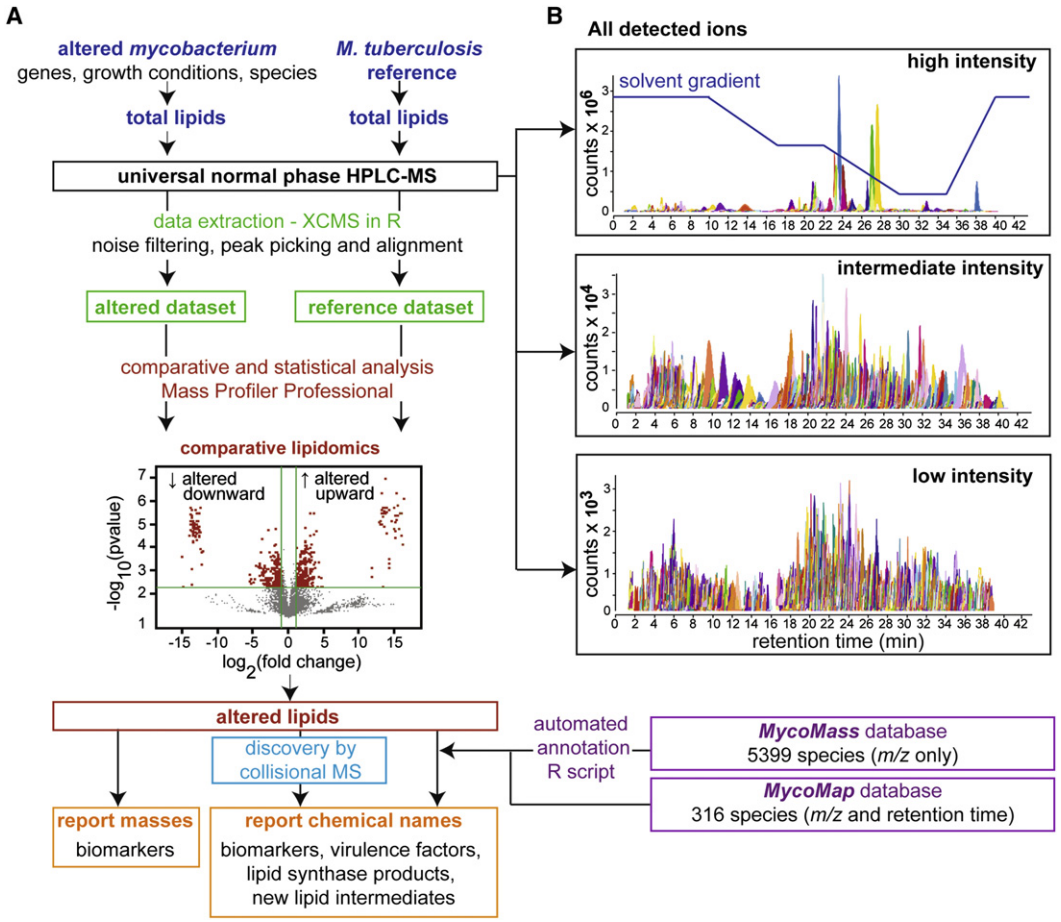

Figure 2. Lipidomics Platform

(A) Lipid extracts (dark blue) enter a workflow involving a universal normal-phase HPLC-MS system (black), software-assisted raw data extraction (green), computational comparative analysis (red), database and dataset annotation (purple), and molecular discovery through collisional mass spectrometry (light blue). This second-generation system for comparative profiling emphasizes a single-step chromatography system, in contrast to a first-generation method that uses fluid phase separation and multiple HPLC systems (Figure S2).

(B) Extracted ion chromatograms of the overall features detected with high, intermediate, and low intensity by analyzing *M. tuberculosis* H37Rv total lipids.

extract higher-mass (300–3,000 atomic mass units), low-turnover (days) compounds into chloroform-containing solvents, which, for mycobacteria, are unusually diverse long-chain lipids.

Comparative lipidomics requires precise normalization of input lipids as well as reproducible and sensitive detection of aligned, replicate lipidomes. After initially adopting published ionization conditions for anionic membrane phospholipids (Kruve et al., 2010), reduction of countercurrent gas flow and increasing voltage produced greater than 100-fold increases in sensitivity and revealed high mass neutral lipids that were otherwise undetectable, such as trehalose monomycolate (Figures 3B and 3C). These gains resulted from generating greater force toward the detector and were seen for many ions in the mycobacterial lipidome. This dramatic and broad-based improvement in sensitivity resulted from adjustments of ionization conditions that are better suited to the larger, uncharged lipids that populate the mycolate layer.

Interlipidome comparisons rely on precise normalization of input lipids, which was accomplished using cultures harvested at a similar optical density of 0.6 (±0.1) and weights determined with less than 1% SD. Using this method, the SD of triplicate intensity measurements of benchmarks was about 2% for technical replicates and 10% for biological replicates (Figure 3D). Thus, experimental error was low in absolute terms and derived mainly from non-uniform bacterial culture rather than LC-MS detection. Normalization at the detection level was further confirmed by continuous detection of calibrants, tracking total ion current generated by all lipids, and by monitoring abundant structural lipids that serve as housekeeping controls. For example, cardiolipin showed highly reproducible intensity among biological replicates (Figure 3D) and did not change significantly between diverse samples subject to biological variables (data not shown). Serial dilution of input lipids determined that 500 μg/ml provided nonsaturating and near linear detection of benchmarks (Figure 3E) and other lipids over a 10-fold or higher change in concentration. Thus, a 20 μl injection from scant in vivo specimens yet produced allowed sensitive and broad lipidomic coverage with benchmark lipids detected below the picogram range (Figure S3). Therefore, the platform met the criteria for sensitivity, reproducibility, and relative quantitation.

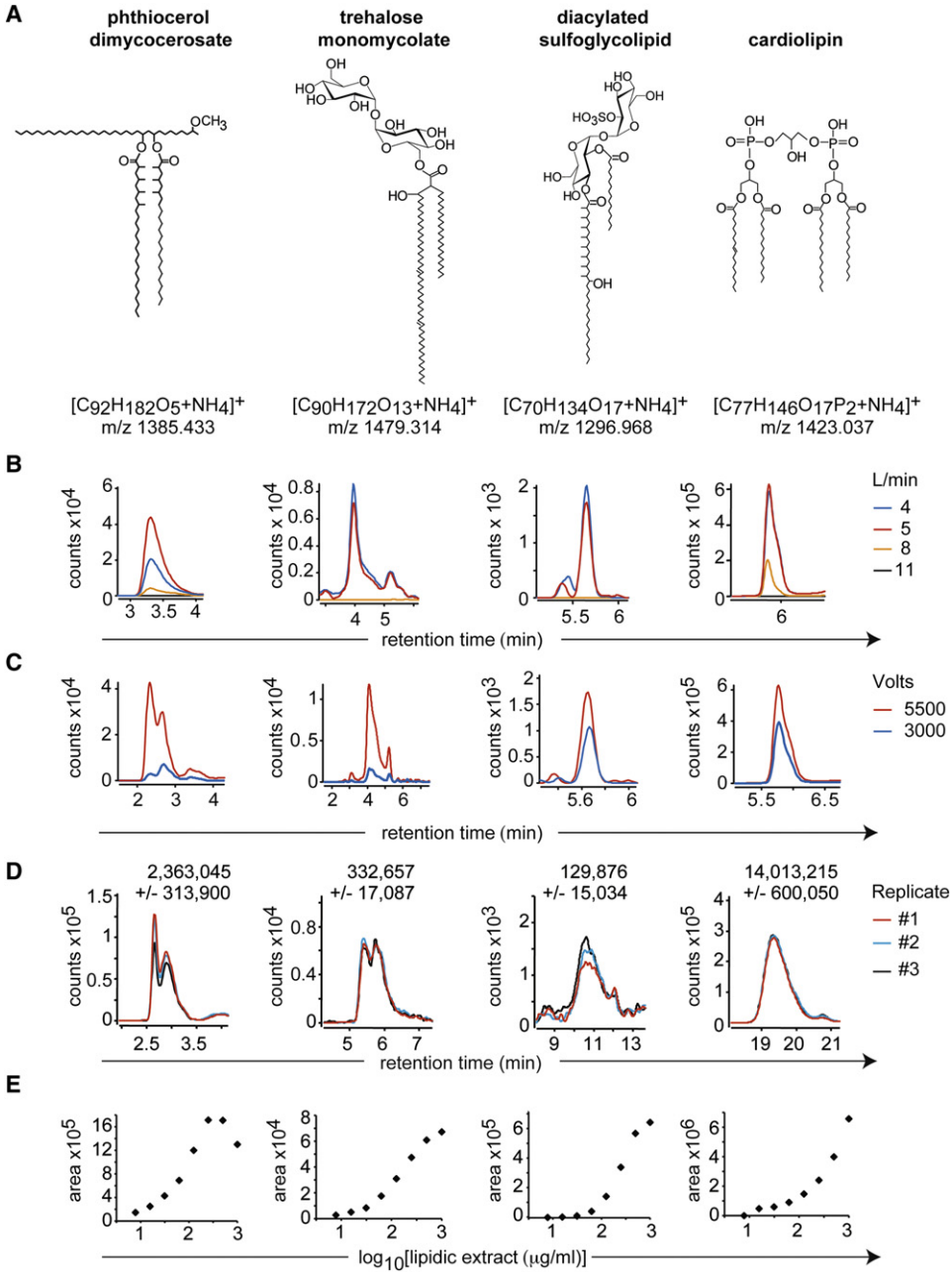

Figure 3. Validation of a Universal Normal-Phase HPLC-MS Detection

(A–C) Structures (A) of diverse benchmark lipids for HPLC-MS method optimization related to countercurrent gas (B) and source voltage (C) as measured in biological replicates (D).

(E) Relationship of signal intensity derived from areas under the curves of ion chromatograms to input mass of total lipid. Data are representative of three or more experiments.

Mapping the Lipidome of *M. tuberculosis*

This single-step universal chromatography method was implemented for assigning the retention time (RT) of each lipid family as an organizing principle for mapping of the lipidome. Large raw datasets were processed by XCMS for noise filtering, peak picking, and deconvolution to resolve coeluting ions and peak alignment across replicates so that features with equivalent AMRT values are aligned across biological conditions and their intensities reported in a final data matrix (Smith et al., 2006) (Figure 2A, green). A feature was defined as a three-dimensional value of *m/z*, RT, and intensity detected in triplicate. Typical analyses of *M. tuberculosis* lipid extract yielded a data matrix of ~6,000 and ~5,000 features acquired in the positive and negative ion mode, respectively (Figure S4). Even when lacking chemical names, features with high fold change have value as markers of the bacterial response (Figure 2A). Nevertheless, we sought to

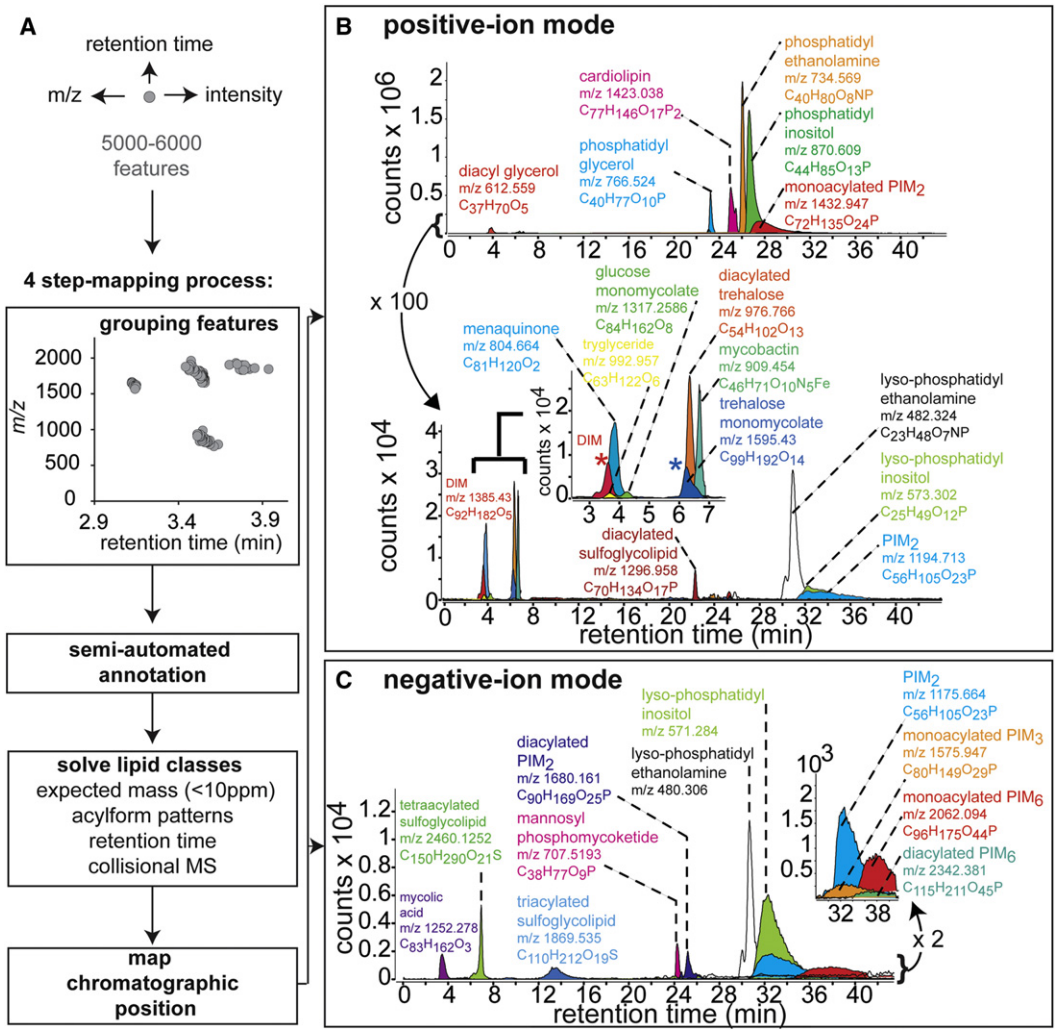

Figure 4. Mapping the Lipidome of *M. tuberculosis* H37Rv

(A) HPLC-MS dataset of *M. tuberculosis* H37Rv of ~6,000 features, which are three-dimensional coordinates of linked *m/z*, RT, and intensity. One lead compound in each cluster was tentatively identified by automated annotation using MycoMass, confirmed by four analytical criteria, and mapped to the chromatographic system in positive (B) and negative (C) ion mode.

(B and C) Extracted ion chromatograms of mapped lipid families. Neutral formulas of the studied alkylforms and detected *m/z* of the respective $[M+NH_4]^+$ or $[M+H]^+$ adducts (B) and $[M-H]^-$ forms (C) are indicated. RTs of lipids typically varied by less than 5 s in one experiment, but varied up to 60 s among users with differing columns. Phosphatidylinositol mannosides (PIM_x) are listed according to the number (x) of mannosyl residues. Thirty lipid families mapped in this way comprise the MycoMap. Features annotation and collisional MS are shown in Figure S4.

"presolve" many key features of *M. tuberculosis* H37Rv as named compounds to allow broad monitoring of known molecules.

We created fine maps of 30 families of lipids by repeatedly applying a four-step process (Figure 4A). First, when plotted as RT versus *m/z*, features self-organize into clusters comprising many alkylforms with the same head group. This clustering results from normal-phase chromatography, which resolves all molecules into separate families, but not alkylforms within a family. Tight clustering facilitates identification of families and limits their overlap in RT, reducing molecular heterogeneity and cross-suppression between disparate lipid families. Second, clustered features were tentatively assigned to a known lipid using an in-house designed script using R software (see Supplemental Information), allowing automatic naming of features whose *m/z* matched within 10 ppm to any entry in MycoMass. This software achieved tentative annotation of 624 and 366 features in positive and negative ion mode datasets, respectively (Figure S4).

Step 3 tests the assignments by comparing alkylform diversity, RT, and collisional MS of feature groups to known molecules (Figure S4). For example, features initially annotated as PDIM by software were found to match the RT of a synthetic PDIM standard (3.6 min, data not shown) and appeared as an alkane series of the expected length (C87-C102) of *M. tuberculosis* PDIMs (Figure S4; Figure 5A). One ion of *m/z* 1371.413 matching the expected *m/z* for the ammoniated adduct of $C_{91}H_{180}O_5$ PDIM (*m/z* 1371.412) showed the fragments expected of phthiocerol

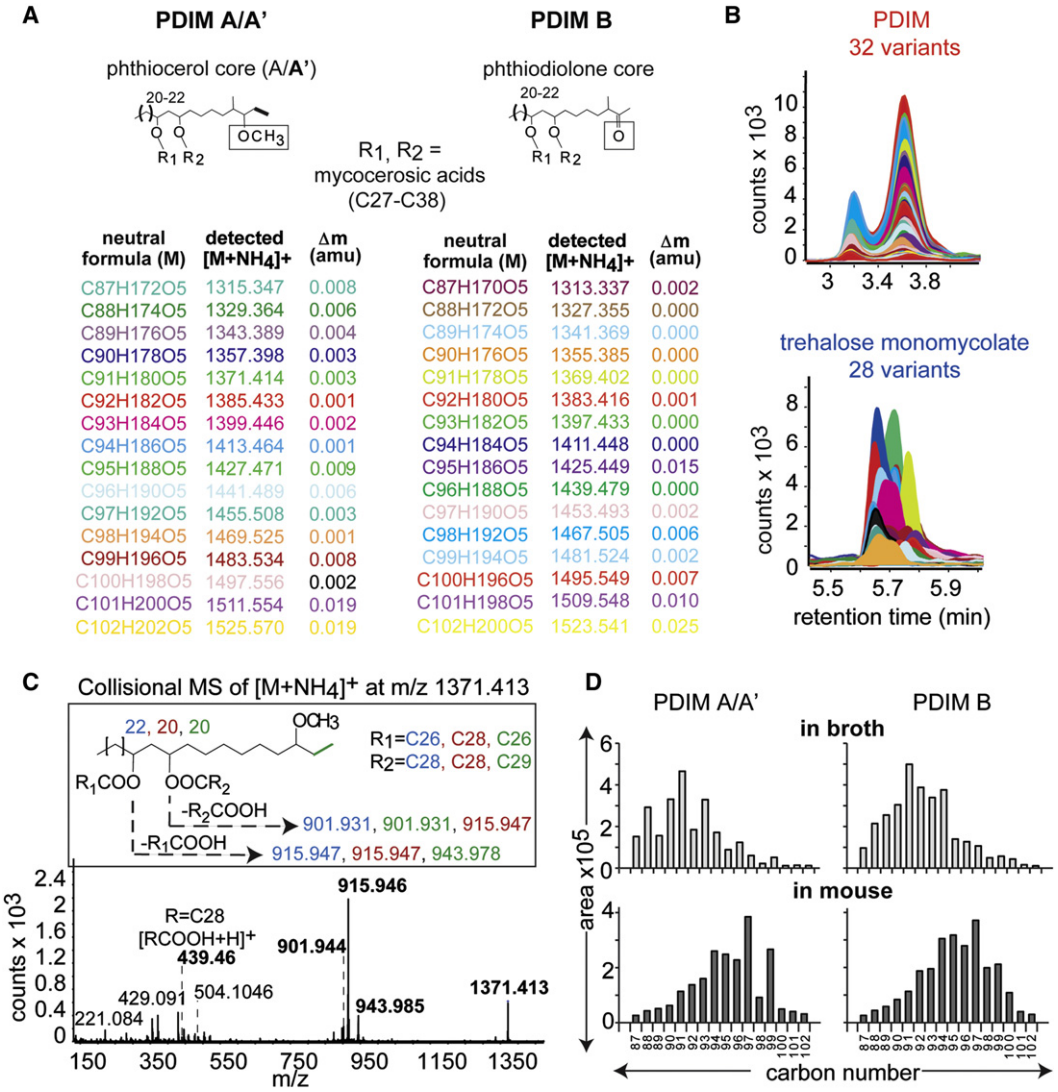

Figure 5. In Vitro and In Vivo Fine Mapping of *M. tuberculosis* H37Rv Lipid Families

(A–C) Extracts of *M. tuberculosis* H37 Rv grown in vitro were subject to detection of PDIM, trehalose monomycolate (A and B), and 28 other lipids (Figure S5) illustrating the PDIM detected alkylforms (A) and RT profiles (B) that match by color and were confirmed by collisional mass spectrometry (C).

(D) *M. tuberculosis* Erdman 2.5 grown in broth media or in one infected mouse were similarly analyzed for the PDIM A/A′ and B alkylforms with the indicated overall carbon number. Similar results were found in separate analysis of three mice (Figure S6).

and mycocerosyl substructures (Figure 5C). We repeated this process for 30 annotated lipid families (Figures 4B and 4C; Figure S4). Because most of the mycobacterial lipidome's diversity is found in neutral lipids (Figure S1), we obtained larger datasets in the positive ion mode than in the negative ion mode (Figures 4B and 4C). This result contrasts with conventional lipidomic studies that emphasize detection of negatively charged membrane phospholipids of mammalian cells. Both modes provided a high dynamic range that spanned four log orders of magnitude, so diacylglycerides, monoacyl PIM, and other strongly detected lipids ($\sim 10^6$ counts), as well as PDIM, triglycerides, menaquinone, sulfoglycolipids, mannosyl phosphomycoketide, lysophospholipids, and lower-intensity PIM families ($\sim 10^4$ counts), could be tracked in parallel with one injection.

Fine Mapping

Fine mapping is a process whereby each alkylform within a family is separately assigned. For example, PDIM naturally occurs with between 87 and 102 total carbon atoms within A and B families, distinguished by methoxy or keto substitutions (Constant et al., 2002) (Figure 5A). Sixteen length variants in two families predict 32 alkylforms. We detected and mapped all 32 alkylforms as nearly overlapping chromatograms (Figures 5A and 5B). Similarly, we mapped 28 trehalose mycolates (Figure 5B) and the alkylforms of 28 other lipid families, covering 318 compounds to create MycoMap (Figure S5). Fine mapping supports applications that take advantage of species- or strain-specific patterns in actinobacteria for clinical diagnosis or chemotaxonomic assignment (Song et al., 2009). Furthermore, alkylforms within a family can change differentially in response to a biological

variable, so fine mapping can describe chemical remodeling. For example, *M. tuberculosis* harvested from mouse lung produces longer PDIM alkylforms compared to bacteria from in vitro culture, resulting from increased methylmalonate availability in tissues (Jain et al., 2007). We detected 32 alkylforms of PDIM produced in infected mouse lungs, which confirmed previously reported in vivo lengthening (Figure 5D; Figure S6). Thus, fine mapping illustrates that the broadly comparative method described here can meet requirements previously accomplished with targeted ion finding. Detection is adequately sensitive for lipids extracted from infected tissues, where mammalian lipids greatly predominate.

Comparative Lipidomics

Comparative lipidomics requires efficient algorithms to process raw LC-MS data, measure the intensity of individual chromatograms, and align thousands of features across many lipidomes to generate a data matrix. Our first-generation (Figure S2) and other (Sartain et al., 2011) data extraction methods pool isotopes and adducts of deduced neutral molecules (M) as one intensity value. This approach causes quantitative errors related to adduct assignment and pooling of many intensity values, which leads to quantitative errors when count values detected in the nonlinear range of counts to mass are summed. When comparing two large datasets, these kinds of errors caused many ions to be incorrectly assessed as changed molecules, leading to unacceptably high false-positive molecular finding rates. Manual inspection of chromatograms improved reliability but were too cumbersome to evaluate more than 100,000 peaks in one experiment. Therefore, we implemented XCMS ion finding algorithm, which treats all ions as separate features and bypasses ion batching errors. Implementation of reliable automated ion finding algorithms was the key advance allowing comparison of datasets with more than 100,000 features.

Next, Mass Profiler Professional software was implemented for comparisons and statistical analysis of XCMS-generated data matrices to report changed features and their p values corrected with the Benjamini-Hochberg multiple comparisons test (Figure 2A, red). To quickly highlight significant changes, results were displayed in two-dimensional scatter plots of fold change versus corrected p value, also known as volcano plots (Figures 6A–6D). Because a typical comparative lipidomics experiment generates hundreds of changed molecules, which collectively exceed any capacity for detailed biological validation, the overriding design objective is limiting false-positive molecular discovery. Therefore, we used stringent filters to remove ions absent in any replicate and those with intensity values showing high variance (corrected $p > 0.05$). We considered a feature to be changed when its intensity value changed at least two-fold (Figures 6C and 6D, red dots), which exceeds the sum of the typical variation observed among biological triplicates (Figure 3D). Despite these stringent filters, the biofinformatic pipeline permitted broad coverage, typically 4,000 to 10,000 comparisons per experiment. By comparing two triplicate analyses of the same bacterial culture, the percentage of features that are described as changed represented the false-positive molecular discovery rate (Figure 6A). Remarkably, 6,498 pairwise comparisons yielded no false-positive results from cumulative errors in extraction, separation, detection, and software analysis. Similar analysis for biological replicate cultures showed an error rate of 0.7% (Figure 6B). Thus, errors derive mainly from culture rather than LC-MS detection. The near zero rate of false molecular detection provided a blank canvas against which any molecules changing after introducing biological variables would likely be caused by the biological variable, setting the stage for chemotaxonomy analyses of mycobacteria.

Chemotaxonomy

The goals of chemotaxonomy are two-fold: (1) measure the number of changed molecules as a descriptor of chemical relatedness of two bacteria and (2) provide a list of changed molecules to discover biomarkers. We compared virulent *M. tuberculosis* H37Rv with avirulent *M. smegmatis*, and with a reference strain of the W Beijing clade of *M. tuberculosis* (HN878) (Figures 6C and 6D). Molecular features showing intensity changes that met variance criteria for genuine differences showed two patterns. For any feature that shows background signal in one dataset, its intensity is assigned as 1 rather than 0. Therefore, all or nothing changes in molecules, which represent the best biomarkers, appear as high ($>2^{10}$) but not infinite change values. Features with a 2- to 2^{10}-fold change represented features present in both bacteria with altered concentrations. These features might represent regulated lipids that define the physiological state of the bacterium. The intraspecies and intragenus comparisons detected 648 changes (11%) and 4,339 changes (47%), respectively (Figures 6C and 6D). Thus, the scope of chemical change correlates with genetic relatedness, validating the discriminatory potential of this lipidomics method.

Because each feature contains embedded AMRT information and can be subjected to automated annotation, this comparison can rapidly identify strain specific biomarkers without further experimentation. For example, the W Beijing lineage of *M. tuberculosis* is hypervirulent in mice (Dormans et al., 2004; Manca et al., 2005; Reed et al., 2004) and has emerged worldwide as a human pathogen with distinct transmission features (Glynn et al., 2002). Among 5,886 pairwise comparisons between *M. tuberculosis* H37Rv and the clinical reference strain for Beijing (HN878), we identified 303 features upregulated in the Beijing strain, of which 69 represented all or nothing changes (Figure 6C, enlarged). Automated and manual annotation showed that 38 are alternative adducts and isotopes of the same alkane series with nominal neutral mass values (M) between 1785 and 1911 (Figure 6C, green; Figure S7A). These ions and their key fragments (Figure 6E) correspond to the expected mass of triglycosylated phenolic glycolipids (PGLs). This virulence-associated glycolipid has been previously identified on a genetic basis by intact polyketide synthase 15/1 in the W Beijing lineage, but not H37Rv, which has a frameshift in this locus (Constant et al., 2002). Thus, we identified a known strain-specific molecule using a rapid and unbiased cell wall screen. Identification of PGLs in 38 distinct molecular forms represents a redundant and convincing form of detection that is not possible using bioinformatic methods that batch isotopes and adducts.

Lipidomic versus Targeted Scanning for PGLs

The relationship between Beijing lineage and PGLs was previously known, but there were no clinical tests for the screening

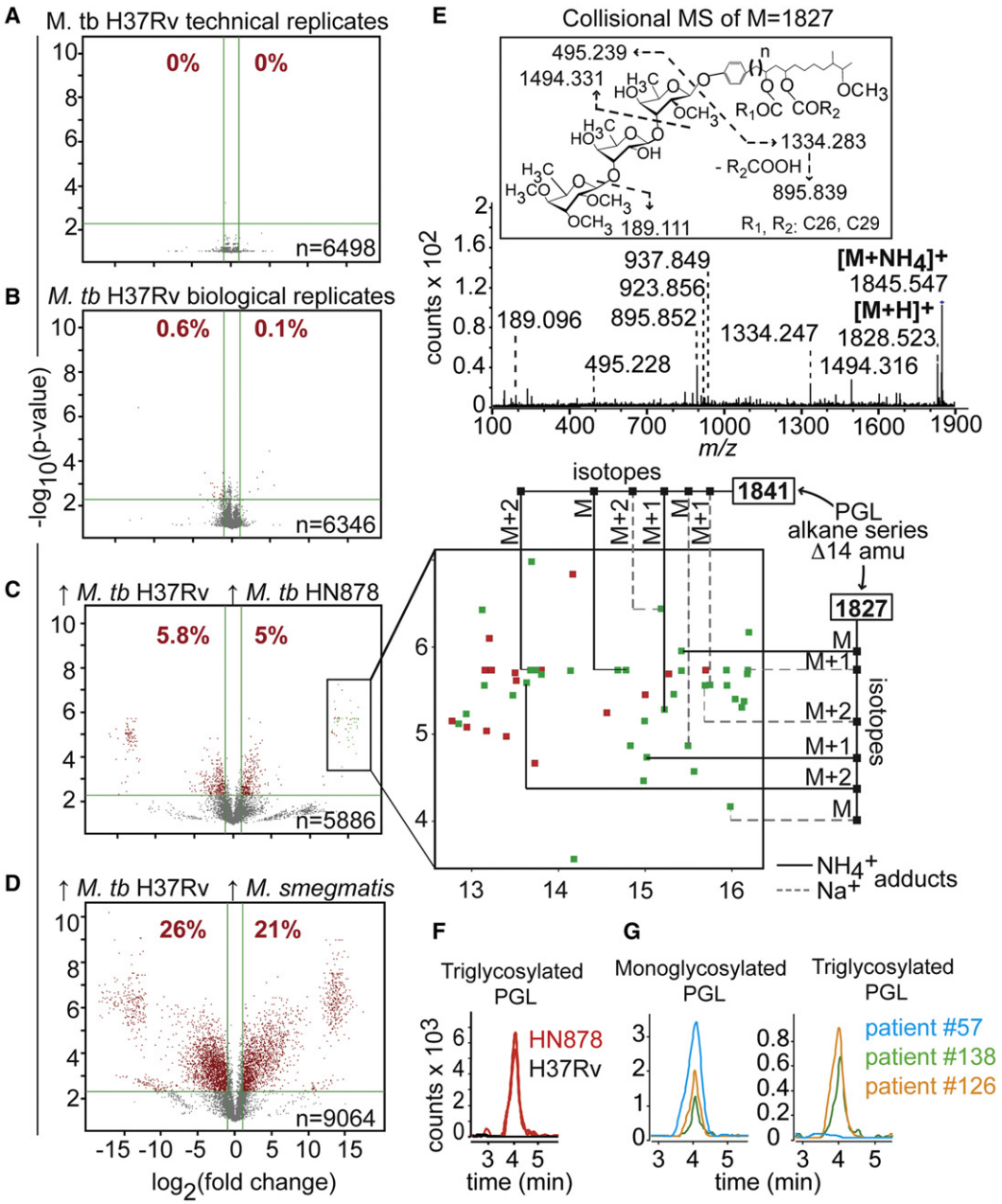

Figure 6. Comparative Chemotaxonomy

(A–D) Pairwise comparison of extractable lipids represented as volcano plots, showing in red the features meeting criteria for two-fold change and significance ($p < 0.05$, corrected for multiple comparison) also indicated as a percentage of all features (n). *M. tuberculosis* H37Rv lipid extract from one (A) or two (B) liquid cultures were analyzed in triplicate and compared with *M. tuberculosis* Beijing HN878 (C) or *M. smegmatis* (D). Among features uniquely present in the W Beijing strain (C, inset, and listed Figure S7A), 38 (green) corresponded to isotopes (M, M+1, M+2) and adducts (NH_4^+ or Na^+) of a triglycosylated PGL alkane series, as illustrated for two alkylforms of nominal masses of 1827 and 1841.

(E) Collisional mass spectrometry of $[M+NH_4]^+$ adduct of PGLs confirmed structure composed of a phthiocerol core esterified by C27 and C30 mycocerosic acids (R_1COOH, R_2COOH).

(F and G) Extracted ion chromatograms of a representative alkylform of the monoglycosylated (m/z 1553.442) or triglycosylated (*m/z* 1845.554) form of PGLs for laboratory and patient isolates show sensitive detection that is not confounded by other lipids and separate detection of the two PGL glycoforms. The mass spectrum of triglycosylated PGLs is shown in Figure S7B.

of triglycosylated PGLs found in virulent strains, or for comparing this virulence-associated molecule with the monoglycosylated form found in Bacillus Calmette-Guérin (BCG) (Daffé and Laneelle, 1988). Genetic tests of the pks15/1 locus generally rule out PGL production when abnormal, but are not sufficient to rule in PGL production, because many other genes are

needed (Malaga et al., 2008; Pérez et al., 2004). Because the best available chemical test, radio-thin layer chromatography, requires biosynthetic labeling in biosafety level 3, it is not feasible in clinical laboratories (Reed et al., 2004). Therefore, despite considerable interest in the dispersion of Beijing strains worldwide and direct evidence for PGLs as a virulence factor in mice, studies of PGLs in human isolates have been limited. Using AMRT (RT = 4 min, *m/z* 1845.547) and a diagnostic fine map from the lipidomics platform (Figure 6C), we converted from a broad scanning mode to a simplified, specific analysis of ions corresponding to monoglycosylated (*m/z* 1553.442) and triglycosylated (*m/z* 1845.554) PGLs. Signal intensity is more than 100-fold above baseline levels and is not confounded by any other ion (Figure 6F; Figure S7B). We applied this test to detect PGLs in patient isolates from South Korea (Figure 6G) and identified both mono- and triglycosylated PGLs in isolates genotyped as Beijing strains with an intact pks15/1 locus. Further, we identified an isolate with discordant production of the mono- and triglycosylated PGLs. These studies illustrate the transition from lipidomic scanning to a focused analysis. In contrast with the current gold standard test requiring radioactive labeling, this test uses standard media and is rapid, sensitive, and chemically specific, because collision-induced dissociation (CID)-MS provides chemical detail. More generally, all 303 events that meet statistical and fold change criteria become candidate targets for strain-specific biomarkers or determinants of W Beijing physiology.

DISCUSSION

Based on its sensitivity, this comparative lipidomics system can be used to evaluate any genetic or biological perturbation, even within infected cells. Based on the low rate of false-positive molecular discovery, it is possible to embark on unbiased discovery for all molecules regulated by any single gene deletion or metabolic perturbation. Therefore, this method is currently being applied to determine cell wall changes induced by antitubercular drugs, evolution of multidrug resistance, dormancy, cellular infection, and iron deprivation. The development of a new clinically useful test for the PGL virulence determinant provides a glimpse of the high value of extending general lipidomic maps of model organisms to pathogens with unusual lipid repertories. A simple test to monitor PGLs in clinical *M. tuberculosis* strains allows new investigation to determine whether the virulence-inducing effects seen in mice might occur in humans or are outweighed by fitness costs or decreased transmission from altered immune response (Comas et al., 2010).

These studies also provide a quantitative estimate of the scope of current knowledge of the mycobacterial lipidome. Even using MycoMass, the largest mycobacterial database available, we annotate only up to 20% of the detected molecules in any lipidome. The events corresponding to unnamed molecules might derive from fragmentation or redundant detection of unexpected adducts or might simply be molecules produced by *M. tuberculosis* that are not known in the literature. Based on the low rates of source fragmentation and redundant detection of molecules in altered ionization states observed during the mapping process, it appears that knowledge of the mycobacterial lipidome is far from complete. Indeed, the mycobacterial genome has an unusually large number of lipid synthases, and for many of these, their products remain unknown. These facts are surprising given the worldwide scope of the tuberculosis epidemic, with an estimated 1.6 million deaths per year (World Health Organization, 2009). Subtraction of all entries comprising MycoMass from those in any routinely generated lipidome shown here provides a tangible list of unnamed molecules that represents a map for solving the molecular toolkit of the world's most devastating bacterial pathogen.

SIGNIFICANCE

Mycobacterium tuberculosis remains one of the world's most deadly bacterial pathogens and survives within human cells using a protective lipid envelope comprised of distinct layers. This lipidic cell wall regulates uptake of nutrients and antitubercular drugs while shedding lipid adjuvants, antigens, and pathogen-specific markers of infection. To profile mycobacteria on an organism-wide basis, we first solved a static _M. tuberculosis_ lipidomic dataset comprised of mass spectrometry datasets, a lipid database containing more than 5,000 neutral masses from medically relevant mycobacteria, and an accurate mass-retention time map of more than 300 lipids with 30 fine maps of alkyl chain variants. Among 58 lipid types in the MycoMass and MycoMap databases, more than 40 are lacking in eukaryotic or Gram-negative organisms, illustrating the need to move beyond model organisms for direct study of the specialized molecules in pathogens. We implemented a broadly separating, single-step chromatography system together with automated ion finding and statistical and annotation software to create a platform for comparative lipidomics. This platform iteratively solves replicate lipidomes before and after infection or among various clinical isolates to provide broad measurements of pathogen response and chemotaxonomic information. Pairwise comparison of ~6,000 aligned features describes chemical relatedness of mycobacteria with low false-positive molecular discovery rates. Broad chemotaxonomic analyses of mycobacteria measured the extent of chemical change associated with species and strain-specific variants and provided detailed lists of the molecules changed. Unbiased scanning of a W Beijing strain of _M. tuberculosis_ identified the known biomarker phenolic glycolipid and provided the basis for a new clinically applicable test for forms of this glycolipid that have or have not been associated with virulence.

EXPERIMENTAL PROCEDURES

MycoMass Database

Lipids for *M. tuberculosis*, *M. smegmatis*, *M. bovis* BCG, *M. avium*, *M. leprae*, and *M. marinum* were reported according to LIPID MAPS conventions. Alkyl-forms are variations in length and unsaturation of lipids based on all possible lipid substitutions, except for PIMs and trehalose dimycolates in which a smaller number of specific combinations are known to occur (Fujita et al., 2005; Gilleron et al., 2006). From calculated neutral mass values (M), the expected $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$, $[M-H]^-$, $[M+HCOO]^-$, and $[M+CH_3COO]^-$ are shown to five significant figures. Due to their distinctive appearance in MS, $[M+Fe^{54/56}-2H]^+$ mycobactins and phosphatidyl ethanolamine $[2M+H]^+$ were listed for a total of 32,438 entries to the MycoMass

database (Figure S1) (http://www.brighamandwomens.org/research/depts/medicine/rheumatology/labs/moody).

Mycobacterial Culture

M. smegmatis mc²155, *M. bovis* BCG, and *M. tuberculosis* H37Rv (Trudeau Institute) and *M. tuberculosis* HN878 (Robert N. Husson) were cultured in 6 ml Middlebrook 7H9 broth supplemented with 10% oleic acid/albumin/dextrose/catalase (Becton Dickinson) in 50 ml polystyrene tubes (Corning) and shaken at 100 rpm at 37°C until visible growth appeared, up to 2 weeks depending on the species. One mL of the starter culture was transferred in triplicate to 45 ml fresh media in 250 ml sterile polystyrene containers with vented caps and in singlicate to 45 ml of fresh media supplemented by 0.05% Tween 80 for growth monitoring by OD_{600} measurement. Triplicate Tween-free cultures were harvested when the Tween culture reached 0.6 (±0.1) OD_{600}. W Beijing family clinical isolates were obtained from retreatment (subjects #126 and #138) and one newly diagnosed case (subject #57) was obtained from a TB natural history study (ClinicalTrials.gov identifier: NCT00341601) at the National Masan Tuberculosis Hospital in the Republic of Korea. Cultures were obtained following sputum processing, microscopic examination for acid-fast bacilli and the BacT liquid culture system (Biomerieux) or from Ogawa slants (ShinYang Chemicals, Korea) incubated at 37°C in ambient air for a maximum of 8 weeks. Primary cultures were identified using classical methods (Lévy-Frébault and Portaels, 1992) and stored at −80°C. Drug susceptibility testing was performed by the proportion method on Lowenstein-Jensen medium using previously described methods (Canetti et al., 1969; Wayne, 1974). For mass spectrometry, isolates were cultured in 20 ml of 7H9 medium supplemented with 10% albumin/dextrose/catalase (EMD Chemicals, San Diego, CA) containing glycine-alanine salts and incubated in 250 ml bottles in a rolling incubator at 37°C.

Lipid Extraction

LC-MS grade solvents (Fisher) and clean borosilicate glassware (Fisher), amber vials (Supelco), and Teflon-lined caps (Fisher) were used. Laboratory strains were centrifuged (4,000 rpm,10 min) to clarify culture supernatants, which were passed twice through a 0.22 μm filter to detect secreted compounds. Cell pellets were washed twice in 10 ml Optima water, resuspended in 1 ml of CH_3OH, transferred to a 50 ml amber glass bottle, and contacted with 25 ml $CHCl_3/CH_3OH$ (2:1, v/v) overnight to sterilize bacteria. $CHCl_3/CH_3OH$ suspensions were transferred in 50 ml conical glass tubes and shaken on an Orbitron rotator for at least 1 hr. After centrifugation, lipid extracts were decanted, and bacteria pellets were subjected to two additional extractions using $CHCl_3$:CH_3OH (1:1, v/v) and $CHCl_3$:CH_3OH (1:2, v/v) with pooling of extracts and evaporation with GeneVac EZ-2 (SP Scientific) using the low boiling point mixture setting. Dried lipids were resuspended in a minimum volume of $CHCl_3$:CH_3OH (1:1, v/v) and dried under nitrogen in preweighed vials, then reweighed in triplicate on microbalance (Mettler Toledo, XP205), and values were reported when fully dried as shown by replicate measurements showing less than 1% variance. Using 2 mg of lipid extract, replicate measures showed variance of 20 μg, providing mass errors below 1% for in vitro-derived samples. Extracts were then redissolved in $CHCl_3$:CH_3OH (1:1, v/v) at 1 mg/mL. For clinical isolates, mid-log phase cultures (OD = 0.5±1) were centrifuged at 3,000 rpm for 15 min and extracted as described (Reed et al., 2004).

HPLC-Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry

Using an Agilent Technologies 6520 Accurate-Mass Q-Tof and a 1200 series HPLC system with a Varian Monochrom diol column (3 μm × 150 mm × 2 mm) and a Varian Monochrom diol guard column (3 μm × 4.6 mm), up to 50 μg of lipid extract was dried under nitrogen and resuspended at 0.5 mg/mL in solvent A (hexanes:isopropanol, 70:30 [v: v], 0.02% [m/v] formic acid, 0.01% [m/v] ammonium hydroxide), filtered or centrifuged at 1,500 rpm for 5 min to remove trace nonlipidic materials prior to transfer to a glass autosampler vial (Agilent). Ten μg were injected, and the column (20 C) was eluted at 0.15 ml/min with a binary gradient from 0% to 100% solvent B (isopropanol:methanol, 70:30 [v/v], 0.02% [m/v] formic acid, 0.01% [m/v] ammonium hydroxide): 0–10 min, 0% B; 17–22 min, 50% B; 30–35 min, 100% B; 40–44 min, 0% B, followed by additional 6 min 0% B postrun. Ionization was maintained at 325°C with a 5 l/min drying gas flow, a 30 psig nebulizer pressure, and 5,500 V. Spectra were collected in positive and negative ion mode from *m/z* 100 to 3,000 at 1 spectrum/s. Continuous infusion calibrants included *m/z* 121.050873 and 922.009798 in positive ion mode and *m/z* 112.985587 and 1033.98810 in negative ion mode. CID-MS was performed with an energy of 30–60 V. For any large-scale comparative analysis, the column is conditioned by three succesive 10 min elutions with B, B/A 1:1 (v/v) and A solvents followed by three mock injections with solvent cycling before mycobacterial samples are analyzed.

HPLC-MS Data Extraction and Alignment

Raw data files were converted to mzData using MassHunter and processed in R using the XCMS (version 1.24) (Smith et al., 2006) centWave peak finder method designed for high mass accuracy data (Tautenhahn et al., 2008). XCMS was downloaded from http://metlin.scripps.edu/xcms/index.php. Briefly, peaks were deconvoluted and aligned across samples using an s/n threshold of 5, a maximum tolerated *m/z* deviation of 10 ppm, a frame width of *mzdiff* = 0.001, a peak width of 20–120 s, and a band width of 5. The aligned output consisted of accurate mass, retention time (RT = 1–2,640 s), and intensity of each peak and was exported as .csv files for analysis in Mass Profiler Professional or for automatic annotation in R. Although mass accuracy of <2 ppm was achieved with optimized conditions, trace compounds from biological sources can provide lower mass accuracy, leading to a 10 ppm mass window, which was validated to efficiently capture data points that describe a chromatographic peak when detected in complex mixtures with varying peak intensity.

Semi-automated Annotation

Features, median *m/z*, and median RT information from XCMS output were exported and compared to the MycoMass database formatted as .csv files and performed in R (version 2.11.1) using an in-house designed script. Initial matches for a reference database yielded 1,020 and 768 matches for the positive and negative ion mode, respectively, which were subsequently vetted with AMRT data (MycoMap), reducing the number of annotations to 624 and 366, respectively (Figure S4).

MycoMap Database

The dataset exported as an Excel file was displayed as retention time versus *m/z* on an Excel scatter plot to identify clusters. One lead compound in each group was assigned a chemical formula when it passed all tests: matching the mass of a known lipid family in MycoMass within 10 ppm, matching the alkylform patterns in MycoMass, matching the RT of standards when available, and showing expected fragmentation patterns (Figure S4). Retention times were matched for triglycerides (trioleyl, Avanti), phthiocerol dimycocerosate and mannosyl phosphomycoketide (synthetic, A. J. Minnaard) (Casas-Arce et al., 2008; van Summeren et al., 2006), sulfoglycolipids (purified, M. Gilleron), phosphatidylinositol and monoacyl PIM_2 (in-house purified), mycobactin and carboxymycobactin (from *M. bovis* BCG, C. Ratledge), and glucose monomycolate and trehalose monomycolate (Moody et al., 2002; Moody et al., 2000). Finally, fine mapping assigned a molecular formula, *m/z* and retention time to each alkylform to create MycoMap.

Comparative Lipidomics

XCMS data matrices listing detected features, median *m/z*, and median RT of triplicate lipidic extracts were imported into Mass Profiler Professional (B.02.00) for pairwise comparison (two strains in triplicate) using the Student's paired t test with multiple testing correction (Benjamini and Hochberg, 1995). Features were identified by database matching or collisional mass spectrometry.

Mouse Infection

After aerosol inoculation of C3H mice with *M. tuberculosis* Erdman 2.5 (200 organisms/mouse), mice were sacrificed after 8 weeks. Lung pairs were homogenized in PBS with a bead-beater as previously described (Kamath and Behar, 2005). Homogenates were washed twice with 5 ml of PBS (2,000 rpm, 10 min). The pelleted material was extracted in 3 ml of methanol, vortexed, and transferred to 6 ml of chloroform and methanol and

extracted as above. *M. tuberculosis* Erdman culture used for mouse infection was maintained in parallel in triplicate and extracted as described above.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.chembiol.2011.10.013.

ACKNOWLEDGMENTS

This work is supported by National Institutes of Health Grants U19 AI076217, R01 AI071155, and AR048632; the Broad Institute; and the Burroughs Wellcome Fund for Translational Research. The authors thank M. Gilleron and C. Ratledge for providing lipid standards, R.N. Husson for providing *M. tuberculosis* Beijing strains, and S. Fortune for reading the manuscript.

Received: July 25, 2011
Revised: September 18, 2011
Accepted: October 3, 2011
Published: December 22, 2011

REFERENCES

Adams, K.N., Takaki, K., Connolly, L.E., Wiedenhoft, H., Winglee, K., Humbert, O., Edelstein, P.H., Cosma, C.L., and Ramakrishnan, L. (2011). Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. Cell *145*, 39–53.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B *57*, 289–300.

Borgstrom, B. (1952). Investigation on lipid separation methods. Separation of phospholipids from neutral fat and fatty acids. Acta Physiol. Scand. *25*, 101–110.

Canetti, G., Fox, W., Khomenko, A., Mahler, H.T., Menon, N.K., Mitchison, D.A., Rist, N., and Smelev, N.A. (1969). Advances in techniques of testing mycobacterial drug sensitivity, and the use of sensitivity tests in tuberculosis control programmes. Bull. World Health Organ. *41*, 21–43.

Casas-Arce, E., ter Horst, B., Feringa, B.L., and Minnaard, A.J. (2008). Asymmetric total synthesis of PDIM A: a virulence factor of Mycobacterium tuberculosis. Chemistry *14*, 4157–4159.

Comas, I., Chakravartti, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., and Gagneux, S. (2010). Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat. Genet. *42*, 498–503.

Constant, P., Perez, E., Malaga, W., Lanéelle, M.A., Saurel, O., Daffé, M., and Guilhot, C. (2002). Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the Mycobacterium tuberculosis complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. J. Biol. Chem. *277*, 38148–38158.

Daffé, M., and Laneelle, M.A. (1988). Distribution of phthiocerol diester, phenolic mycosides and related compounds in mycobacteria. J. Gen. Microbiol. *134*, 2049–2055.

Daffé, M., and Draper, P. (1998). The envelope layers of mycobacteria with reference to their pathogenicity. Adv. Microb. Physiol. *39*, 131–203.

de Carvalho, L.P., Fischer, S.M., Marrero, J., Nathan, C., Ehrt, S., and Rhee, K.Y. (2010). Metabolomics of Mycobacterium tuberculosis reveals compartmentalized co-catabolism of carbon substrates. Chem. Biol. *17*, 1122–1131.

Dennis, E.A., Deems, R.A., Harkewicz, R., Quehenberger, O., Brown, H.A., Milne, S.B., Myers, D.S., Glass, C.K., Hardiman, G., Reichart, D., et al. (2010). A mouse macrophage lipidome. J. Biol. Chem. *285*, 39976–39985.

Dormans, J., Burger, M., Aguilar, D., Hernandez-Pando, R., Kremer, K., Roholl, P., Arend, S.M., and van Soolingen, D. (2004). Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different Mycobacterium tuberculosis genotypes in a BALB/c mouse model. Clin. Exp. Immunol. *137*, 460–468.

Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Jr., Murphy, R.C., Raetz, C.R., Russell, D.W., Seyama, Y., Shaw, W., et al. (2005). A comprehensive classification system for lipids. J. Lipid Res. *46*, 839–861.

Früh, K., Finlay, B., and McFadden, G. (2010). On the road to systems biology of host-pathogen interactions. Future Microbiol. *5*, 131–133.

Fujita, Y., Naka, T., McNeil, M.R., and Yano, I. (2005). Intact molecular characterization of cord factor (trehalose 6,6′-dimycolate) from nine species of mycobacteria by MALDI-TOF mass spectrometry. Microbiology *151*, 3403–3416.

Geisel, R.E., Sakamoto, K., Russell, D.G., and Rhoades, E.R. (2005). In vivo activity of released cell wall lipids of Mycobacterium bovis bacillus Calmette-Guérin is due principally to trehalose mycolates. J. Immunol. *174*, 5007–5015.

Gilleron, M., Lindner, B., and Puzo, G. (2006). MS/MS approach for characterization of the fatty acid distribution on mycobacterial phosphatidyl-myo-inositol mannosides. Anal. Chem. *78*, 8543–8548.

Glynn, J.R., Whiteley, J., Bifani, P.J., Kremer, K., and van Soolingen, D. (2002). Worldwide occurrence of Beijing/W strains of Mycobacterium tuberculosis: a systematic review. Emerg. Infect. Dis. *8*, 843–849.

Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J.M., and Engelhardt, H. (2008). Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. Proc. Natl. Acad. Sci. USA *105*, 3963–3967.

Homolka, S., Niemann, S., Russell, D.G., and Rohde, K.H. (2010). Functional genetic diversity among Mycobacterium tuberculosis complex clinical isolates: delineation of conserved core and lineage-specific transcriptomes during intracellular survival. PLoS Pathog. *6*, e1000988.

Jain, M., Petzold, C.J., Schelle, M.W., Leavell, M.D., Mougous, J.D., Bertozzi, C.R., Leary, J.A., and Cox, J.S. (2007). Lipidomics reveals control of Mycobacterium tuberculosis virulence lipids via metabolic coupling. Proc. Natl. Acad. Sci. USA *104*, 5133–5138.

Kamath, A.B., and Behar, S.M. (2005). Anamnestic responses of mice following Mycobacterium tuberculosis infection. Infect. Immun. *73*, 6110–6118.

Kondo, E., Kanai, K., Nishimura, K., and Tsumita, T. (1970). Analysis of host-originated lipids associated with "in vivo grown tubercle bacilli". Jpn. J. Med. Sci. Biol. *23*, 315–326.

Kruh, N.A., Troudt, J., Izzo, A., Prenni, J., and Dobos, K.M. (2010). Portrait of a pathogen: the Mycobacterium tuberculosis proteome in vivo. PLoS ONE *5*, e13938.

Kruve, A., Herodes, K., and Leito, I. (2010). Electrospray ionization matrix effect as an uncertainty source in HPLC/ESI-MS pesticide residue analysis. J. AOAC Int. *93*, 306–314.

Lakshmanan, V., Rhee, K.Y., and Daily, J.P. (2011). Metabolomics and malaria biology. Mol. Biochem. Parasitol. *175*, 104–111.

Lévy-Frébault, V.V., and 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.

Mahrous, E.A., Lee, R.B., and Lee, R.E. (2008). A rapid approach to lipid profiling of mycobacteria using 2D HSQC NMR maps. J. Lipid Res. *49*, 455–463.

Malaga, W., Constant, P., Euphrasie, D., Cataldi, A., Daffé, M., Reyrat, J.M., and Guilhot, C. (2008). Deciphering the genetic bases of the structural diversity of phenolic glycolipids in strains of the Mycobacterium tuberculosis complex. J. Biol. Chem. *283*, 15177–15184.

Manca, C., Tsenova, L., Freeman, S., Barczak, A.K., Tovey, M., Murray, P.J., Barry, C., and Kaplan, G. (2005). Hypervirulent M. tuberculosis W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. J. Interferon Cytokine Res. *25*, 694–701.

Marrero, J., Rhee, K.Y., Schnappinger, D., Pethe, K., and Ehrt, S. (2010). Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for Mycobacterium tuberculosis to establish and maintain infection. Proc. Natl. Acad. Sci. USA *107*, 9819–9824.

Matsunaga, I., Bhatt, A., Young, D.C., Cheng, T.Y., Eyles, S.J., Besra, G.S., Briken, V., Porcelli, S.A., Costello, C.E., Jacobs, W.R., Jr., and Moody, D.B. (2004). Mycobacterium tuberculosis pks12 produces a novel polyketide presented by CD1c to T cells. J. Exp. Med. *200*, 1559–1569.

Moody, D.B., Guy, M.R., Grant, E., Cheng, T.Y., Brenner, M.B., Besra, G.S., and Porcelli, S.A. (2000). CD1b-mediated T cell recognition of a glycolipid antigen generated from mycobacterial lipid and host carbohydrate during infection. J. Exp. Med. *192*, 965–976.

Moody, D.B., Briken, V., Cheng, T.Y., Roura-Mir, C., Guy, M.R., Geho, D.H., Tykocinski, M.L., Besra, G.S., and Porcelli, S.A. (2002). Lipid length controls antigen entry into endosomal and nonendosomal pathways for CD1b presentation. Nat. Immunol. *3*, 435–442.

Pérez, E., Constant, P., Laval, F., Lemassu, A., Lanéelle, M.A., Daffé, M., and Guilhot, C. (2004). Molecular dissection of the role of two methyltransferases in the biosynthesis of phenolglycolipids and phthiocerol dimycoserosate in the Mycobacterium tuberculosis complex. J. Biol. Chem. *279*, 42584–42592.

Quehenberger, O., Armando, A.M., Brown, A.H., Milne, S.B., Myers, D.S., Merrill, A.H., Bandyopadhyay, S., Jones, K.N., Kelly, S., Shaner, R.L., et al. (2010). Lipidomics reveals a remarkable diversity of lipids in human plasma. J. Lipid Res. *51*, 3299–3305.

Raman, S., Puyang, X., Cheng, T.Y., Young, D.C., Moody, D.B., and Husson, R.N. (2006). Mycobacterium tuberculosis SigM positively regulates Esx secreted protein and nonribosomal peptide synthetase genes and down regulates virulence-associated surface lipid synthesis. J. Bacteriol. *188*, 8460–8468.

Reed, M.B., Domenech, P., Manca, C., Su, H., Barczak, A.K., Kreiswirth, B.N., Kaplan, G., and Barry, C.E., 3rd. (2004). A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature *431*, 84–87.

Rohde, K.H., Abramovitch, R.B., and Russell, D.G. (2007). Mycobacterium tuberculosis invasion of macrophages: linking bacterial gene expression to environmental cues. Cell Host Microbe *2*, 352–364.

Rustad, T.R., Harrell, M.I., Liao, R., and Sherman, D.R. (2008). The enduring hypoxic response of Mycobacterium tuberculosis. PLoS One *3*, e1502.

Sartain, M.J., Dick, D.L., Rithner, C.D., Crick, D.C., and Belisle, J.T. (2011). Lipidomic analyses of Mycobacterium tuberculosis based on accurate mass measurements and the novel "Mtb LipidDB". J. Lipid. Res. *52*, 861–872.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C., and Schoolnik, G.K. (2003). Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. J. Exp. Med. *198*, 693–704.

Singh, A., Crossman, D.K., Mai, D., Guidry, L., Voskuil, M.I., Renfrow, M.B., and Steyn, A.J. (2009). Mycobacterium tuberculosis WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response. PLoS Pathog. *5*, e1000545.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem. *78*, 779–787.

Song, S.H., Park, K.U., Lee, J.H., Kim, E.C., Kim, J.Q., and Song, J. (2009). Electrospray ionization-tandem mass spectrometry analysis of the mycolic acid profiles for the identification of common clinical isolates of mycobacterial species. J. Microbiol. Methods *77*, 165–177.

Takayama, K., and Armstrong, E.L. (1976). Isolation, characterization, and function of 6-mycolyl-6′-acetyltrehalose in the H37Ra strain of Myocobacterium tuberculosis. Biochemistry *15*, 441–447.

Tautenhahn, R., Böttcher, C., and Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. BMC Bioinformatics *9*, 504.

Taylor, P.J. (2005). Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. Clin. Biochem. *38*, 328–334.

van Meer, G., Leeflang, B.R., Liebisch, G., Schmitz, G., and Goñi, F.M. (2007). The European lipidomics initiative: enabling technologies. Methods Enzymol. *432*, 213–232.

van Summeren, R.P., Moody, D.B., Feringa, B.L., and Minnaard, A.J. (2006). Total synthesis of enantiopure beta-D-mannosyl phosphomycoketides from Mycobacterium tuberculosis. J. Am. Chem. Soc. *128*, 4546–4547.

Wayne, L.G. (1974). Simple pyrazinamidase and urease tests for routine identification of mycobacteria. Am. Rev. Respir. Dis. *109*, 147–151.

World Health Organization. (2009). Global Tuberculosis Control 2009: Epidemiology Strategy Financing (Geneva, Switzerland: World Health Organization).

Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., and Daffé, M. (2008). Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. J. Bacteriol. *190*, 5672–5680.