Supporting Information

A General Strategy for the Discovery of Metabolic Pathways: D-Threitol, L-Threitol, and Erythritol Utilization in *Mycobacterium smegmatis*

Hua Huang,[†] Michael S. Carter,[†] Matthew W. Vetting,[§] Nawar Al-Obaidi,[§] Yury Patskovsky,[§] Steven C. Almo[§], and John A. Gerlt*[,†,‡,Δ]

[†]Institute for Genomic Biology, [‡]Department of Biochemistry, [Δ]Department of Chemistry and [#]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

[§]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, United States

Contents:

1. Materials and Methods

2. General procedure for the enzyme preparation and assay

2.1 Solute binding protein preparation

2.2 *M.smegmatis* tetritol metabolism enzyme preparation

2.3 Solute binding protein crystallization and structure determination

2.4 Screening of solute binding proteins by DSF

2.5 Kinetic assay for the tetritol catabolic enzymes

2.5.1 Kinetic assay for the dehydrogenase

2.5.2 Kinetic assay for the kinase

2.5.3 Activity assay for the isomerase

3. *M. smegmatis* growth, gene transcriptional analysis and gene knockout studies

3.1 *M. smegmatis* growth condition

3.2 RNA extraction and qRT-PCR

3.3 *M. smegmatis* gene knockout studies

4. Tetritol metabolisms pathway distribution

5. Supporting figures and tables

1. Materials and Methods

Unless otherwise specified, all solvents and organic chemicals were purchased from Sigma-Aldrich (St. Louis, MO) and used without further purification.

All sugars were purchased from Carbosynth (Berkshire, UK). Isopropyl-β-D-thiogalactoside (IPTG), ampicillin and kanamycin were purchased from GoldBio (St. Louis, MO). Primers were purchased from Integrated DNA Technologies (Coralville, IA). Nucleotide sequences were determined by ACGT (Montgomery, MD). The pET23b and pET28a vectors were provided by Novagen PET vector from Life Technologies (Grand Island, NY). *Mycobacterium smegmatis* mc^{2}155, gDNA for *Mycobacterium smegmatis* mc^2 155 and the pk18mobsacB vector were purchased from ATCC (Manassas, VA). Restriction enzymes, T4 DNA ligase and Phusion DNA polymerase were purchased from New England Biolabs (Ipswich, MA).

Growth data were recorded with a Bioscreen C instrument from Growth Curves Limited (Piscataway, NJ).

Proteins were purified with Ni-NTA resin from Qiagen (Venlo, Limburg), dialyzed by membrane tubing from Spectrum Laboratories (Dominguez, CA) and analyzed by mini-protean TGX precast gels from Bio-Rad (Hercules, CA).

Protein concentrations were determined by absorbance at 280 nm as measured by a Nanodrop 2000 from Thermo Scientific (Grand Island, NY).

Enzyme kinetics were measured with a UV spectrophotometer (Varian CARY 300Bio). NMR spectra were collected with an Agilent 600MHz NMR Magnet.

Sequence similarity networks (SSNs) were constructed according to the procedure described elsewhere[1]. In short, a protein sequence, Pfam family number, or InterPro family/domain number were used as input for the EFI-EST webtool (http://efi.igb.illinois.edu/efi-est/). The initial SSN was generated with an alignment score threshold that corresponds to ~35% sequence identity, and nodes that represent proteins with Swiss-Prot curated functions were identified using the node attributes. To establish clusters that approximate isofunctionality, SSNs with increasingly stringent alignment score thresholds were generated until nodes with different SwissProt curated functions were segregated into unique clusters. Because different functions segregate at different alignment scores, this process may segregate divergent homologous proteins with the same catalytic function.

SSN clusters that contain the query protein were uploaded to the EFI-GNT webtool (http://efi.igb.illinois.edu/efi-gnt/) using the default "Neighborhood Size" (± 10 genes) and "Input % Co-Occurrence Lower Limit" (20%). The resulting genome neighborhood network (GNN) quantifies how frequently gene products (functionally identified by the Pfam families of the encoded proteins) are encoded proximal to the genes for members of the SSN query cluster. Given the tendency for genes of functionally linked proteins to be proximal, the GNN can be used to infer functional relationships of the query and proteins encoded by the proximal genes.

2. General procedure for enzyme preparation and assay

2.1 Solute binding protein preparation

Solute binding proteins were cloned by ligation independent cloning (LIC)[2] as previously described[3]. N-Terminal periplasmic signal sequences were determined using the SignalP server[4] and not included in the cloned product. PCR amplicons were produced using genomic DNA as template and the oligonucleotide primers listed in Table S1. PCR products were ligated into vector pNIC23-Bsa4[5], a pET23 based variant of the pNIC28-Bsa4 vector. Clones expressed from this vector yield an expressed product with an N-terminal TEV cleavable HIS-tag (Added N-terminal sequence MHHHHHHSSGVDLGTENLYFQSM). All subsequent growth media contained 100 μg mL^{-1} carbenicillin and 34 μg mL^{-1} chloramphenicol. Clones were transformed into BL21(DE3) containing the pRIL plasmid (Stratagene, Agilent Technologies) and grown overnight at 37°C in 20 ml of PA-0.5G outgrowth media[6]. Each overnight growth was used to inoculate 2 L of PASM-5052 autoinduction media with 125 μg L^{-1} selenomethionine[6a]. Cultures were incubated in a LEX48 airlift fermenter at 37°C for 4 hr and then at 22°C overnight (16–20 hr). Cells were collected by centrifugation and stored at -80°C. Cells were resuspended in buffer A (20 mM Hepes (pH 7.5), 500 mM NaCl, 20 mM imidazole, and 10% (vol/vol) glycerol) and lysed by sonication. Lysate was clarified by centrifugation, with the supernatant loaded onto a 1-mL HisTrap FF column (GE Healthcare). The column was washed with 10 column volumes of lysis buffer, and eluted with buffer B (20 mM Hepes (pH 7.5), 500 mM NaCl, 500 mM imidazole, and 5% (vol/vol) glycerol) directly onto a HiLoad S200 16/60 PR gel filtration column (GE Healthcare), which was equilibrated with buffer C (20 mM Hepes (pH 7.5), 150

mM NaCl, 5% glycerol, and 5 mM DTT). Protein was analyzed by SDS/PAGE, snap frozen in liquid nitrogen, and stored at -80 °C.

2.2 Tetritol metabolism enzyme preparation from *M. smegmatis*

Genes from *Mycobacterium smegmatis* mc^{2}155 were amplified by polymerase chain reaction (PCR) using genomic DNA (ATCC 700084D-5) as template, Phusion DNA polymerase, and the oligonucleotide primers listed in Table S1. PCR products were gel-purified and digested with the appropriate restriction enzymes. For each PCR product, pET-28a was similarly digested before the PCR product was ligated into the vector. The ligation product was used to transform *Escherichia coli* BL21 (DE3). After protein production was confirmed for cells from a single colony, plasmid was purified from the same cells and sequenced. For enzyme preparation, the transformed cells (2 L) were grown at 37 °C with agitation at 200 rpm in Luria-Bertani (LB) broth containing 50 μg/mL kanamycin until OD$_{600}$ of 0.5−0.6. Then, 0.4 mM IPTG was added to induce protein production at 20 °C for 12 h. The cells were harvested by centrifugation (4,648 x g for 12 min at 4 °C), and the cell pellet was suspended and homogenized (1 g of wet cells/10 mL) in ice-cold buffer A [100 mM P$_i$ (pH 8.0), 100 mM NaCl]. The cell suspension was passed through a French press at 1000 psi before centrifugation at 30,000g for 30 min. The supernatant was loaded onto a 10 mL Ni-NTA agarose column at 4 °C. After the column had been washed with 50 mL of buffer A, the enzyme was eluted with 50 mL of elution buffer B [100 mM P$_i$ (pH 8.0), 100 mM NaCl, and 250 mM imidazole]. The fractions were analyzed by SDS−PAGE, and the desired fractions were collected and dialyzed against 4 L buffer C (25 mM Tris, pH 8.0) three times. The purified protein was concentrated and frozen at -80 ^{o}C.

The genes encoding D-erythrulose kinase and L-erythrulose kinase (*msmeg_3271 &*

msmeg_6788) were cloned into pET23b to prepare C-terminally His-tagged proteins by the same

protocol described above. The erythritol/L-threitol dehydrogenase gene (*msmeg_3265*) was

expressed from the pET23b vector to produce a tagless protein. After the protein was produced

in *E. coli*, similar to the method described above, the cell supernatant was loaded onto 30 cm x 5

cm DEAE-Sepharose column, which was eluted with a 2 L linear gradient of NaCl (from 0 to 0.5

M) in buffer C. The column fractions were analyzed by SDS−PAGE. The desired protein

fractions were combined, adjusted to 10% $(NH_4)_2SO_4$ (w/v), and loaded onto 15 cm x 3 cm

Butyl-Sepharose column pre-equilibrated with buffer C containing 10% $(NH_4)_2SO_4$. The column

was eluted with a 0.5 L linear gradient of $(NH_4)_2SO_4$ (from 5 to 0%) in buffer C. The desired

protein fractions were combined, dialyzed against buffer C, concentrated, and stored at -80 ^{o}C.

2.3 Solute binding protein crystallization and structure determination

Protein was rapidly thawed, mixed with TEV protease[7] in a 1/80 (mg TEV / mg target)

ratio and incubated on ice for 2 hr. Protein was buffer-exchanged (20 mM HEPES pH 7.5, 5 mM

DTT) by dilution and ultracentrifugal concentration. Crystallization was by sitting-drop vapor

diffusion with 0.5 µL protein (40-60 mg/mL) and 0.5 µL crystallization agent equilibrated

against 70 µL of crystallization agent in 96 well Intelliplates (Art Robbins Instruments) stored at

18 °C. Crystals of MSMEG_3598 grew from solutions containing 0.1 M imidazole pH 8.0, 0.2

M zinc acetate, 2.5 M sodium chloride, and 10 mM xylitol. Crystals of MSMEG_3599 were

grown from solutions containing 0.1 M citric acid pH 4.0, 25% PEG3350, and 10 mM DL-

threitol. Crystals of MSMEG_3598 and MSMEG_3599 were vitrified and stored by mounting on

nylon loops and immersion in liquid nitrogen without added cryoprotectant. Data from selenomethionine-containing crystals were collected at 100 K on beamline 31-ID (Lilly-CAT; Advanced Photon Source) using a wavelength of 0.97929 Å. Data were integrated and scaled using HKL3000[8] (MSMEG_3598) or XDS (MSMED_3599). The heavy atom substructure and initial phases were calculated utilizing SHELX[9] with the initial models calculated by the automated fitting modules of ArpWarp[10] and SHELX. Iterative cycles of manual rebuilding within the molecular graphics program COOT[11], and refinement against the data within REFMAC[12] were performed until convergence was achieved. Ligands and waters were added during the final cycles of fitting and refinement. Data collection and refinement statistics are given in Table S2.

2.4 Screening of solute binding proteins by differential scanning fluorimetry (DSF)

Solute binding proteins were screened by DSF as previously described[3]. In this instance, a total of 405 ligands consisting of hexoses, acid sugars, amino acids, polyols, and other environmentally available small molecules were arrayed in a 384 well plate as single ligands or combinations of ligands (up to six), with each condition in duplicate. Final DSF reaction mixtures (20 μL final volume) were composed of 10 μM protein, 1 mM ligand, 5x Sypro Orange (Thermo Fisher Scientific) in DSF buffer (100 mM Hepes pH 7.5, 150 mM NaCl, 5 mM DTT). Fluorescence intensities were measured using an Applied Biosystems 7900HT Fast Real-Time PCR System with excitation at 490 nm and emission at 530 nm. The value of ΔT_m for each specific ligand was calculated as the difference of the T_m values measured with a ligand (average of 2 measurements) and without ligand (average of 8 control wells). The following polyols were

in the ligand library: meso-erythritol, DL-threitol, D-mannitol, L-mannitol, D-iditol, galacitol, L-fucitol, D-sorbitol, L-sorbitol, ribitol, D-arabitol, L-arabitol, lactitol, D-talitol, galactinol, maltotetraitol, maltotriitol, isomaltitol, maltitol, allitol, L-altritol, xylitol, and volemitol.

2.5 Kinetic assay for the tetritol catabolic enzymes

2.5.1 Kinetic assay for dehydrogenases

The enzymatic activities of the dehydrogenases were measured at room temperature by monitoring the formation or depletion of NADH/NADPH at 340 nm (ε = 6.2 mM^{-1} cm^{-1}). For the oxidation reaction, the solutions initially contained 1 μM dehydrogenase, 1 mM substrate, 1 mM NAD$^+$/NADP$^+$, and 1 mM MgCl$_2$ in 100 mM glycine/NaOH buffer (pH 9.5). For the reduction reaction, the solution contained 1 μM dehydrogenase, 1 mM substrate, 0.17 mM NADH/NADPH, and 1 mM MgCl$_2$ in 100 mM citrate/Na$_2$HPO$_4$ buffer (pH 6.5). For determination of K_m and k_{cat}, the initial velocities under different concentration of substrate (0.5 K_m to 5 K_m) were fit to eq 1 using KinetAsyst I.

$$V_0 = (V_{max}[S]) / ([S] + K_m) \tag{1}$$

where V_0 is the initial velocity, V_{max} the maximum velocity, [S] the substrate concentration, and K_m is the Michaelis constant for the substrate. The k_{cat} value was calculated from V_{max} and [E] according to the equation $k_{cat} = V_{max}/[E]$, where [E] is the enzyme concentration.

2.5.2 Kinetic assay for kinases

The activities of the kinases were measured by coupling the reaction with pyruvate kinase and lactate dehydrogenase in the presence of NADH and phosphoenolpyruvate (PEP). The depletion of NADH at 340 nm ($\varepsilon = 6.2$ mM^{-1} cm^{-1}) was recorded. In a standard kinase reaction, the solution contained 1 μM kinase, 1 mM substrate, 0.17 mM NADH, 1.0 mM MgCl$_2$, 1.0 mM KCl, 2.5 mM ATP, 2.5 mM PEP, 5 U/mL pyruvate kinase and 5 U/mL lactate dehydrogenase in 100 mM Tris/HCl buffer (pH 8.0). The steady-state kinetic constants (K$_m$ and k_{cat}) were determined using the same conditions to measure the initial velocity under varied concentrations of substrate (0.5 K$_m$ to 5 K$_m$). Then, the data were fit to eq 1 using KinetAsyst I.

2.5.3 Activity assay for isomerases

Isomerization activities were measured by ^{1}H-NMR at 2 and 12 hrs after the isomerases were added to solutions of D-erythrulose-4P or L-erythrulose-4P prepared in D$_2$O buffer. In a typical reaction, D-erythrulose-4P or L-erythrulose-4P was prepared by *in situ* phosphorylation of D-erythrulose or L-erythrulose with the following reagents: 10 mM D- or L-erythrulose, 2.5 mM MgCl$_2$, 50 mM KCl, 0.5 mM ATP, 9.5 mM PEP, 5 U/mL pyruvate kinase and 1 μM of the corresponding D or L-erythrulose kinase were mixed in 25 mM P$_i$ buffer (pD = 8.0). After the kinase reaction was complete as confirmed by ^{1}H-NMR (~30 min at 30°C), the corresponding isomerases were added to a final concentration of 1 μM, and the ^{1}H-NMR spectra were recorded after 2 hours and 12 hours incubation at 30°C.

3. Bacterial growth, transcript analysis and gene truncations.

3.1 Strains and growth condition

$M.\ smegmatis$ mc^{2}155 (ATCC 700084) was purchased from ATCC. The wild type and strains derived from the wild type were regularly grown on Middlebrook 7H10 agar plates without ADC enrichment purchased from Moltox (Boone, NC). For growth studies, strains were grown at pH 7.0 and 37 ^{o}C in minimal medium **A**[13] containing (per liter) 5.0 mM NH$_4$Cl, 0.4 mM Tween 80, 20 mM potassium phosphate buffer (pH 7.0), 2.4 mM MgSO$_4$·7H$_2$O, 1 ml of vitamin solution (per liter: 10 mg pyridoxal hydrochloride), 1 mL of trace element solution (per liter: 2 mg CaCl$_2$, 5 mg CoCl$_2$·6H$_2$O, 2 mg FeCl$_3$·7H$_2$O, 10 mg MnCl$_2$·4H$_2$O, and 3 mg ZnSO$_4$·7H$_2$O) and 21 mM of the carbon sources. Growth data were collected aerobically in a Bioscreen C instrument at 37 ^{o}C with continuous shaking. Strains were precultured in LB broth. Cells were washed with minimal medium A before being used as inoculum.

Wild type $M.\ smegmatis$ is slow to grow with L-threitol, requiring ~60 hr of incubation before permanently adapting to L-threitol growth. The adapted $M.\ smegmatis$ strain (HH001) was prepared with following protocol. Wild type $M.\ smegmatis$ was inoculated into minimal medium A containing 21 mM L-threitol. After incubating at 37 ^{o}C and shaking at 225 rpm for 120 hrs, an aliquot of the culture was streaked onto an LB plate. Strain HH001 was then isolated from a single colony. Even after passage through rich media or minimal glycerol media, strain HH001 grows with L-threitol without a lag, and growth with glycerol, D-threitol, or erythritol is comparable to wild type (Figure S3B). Transcript analysis revealed that during glycerol growth, the genes in Locus III in HH001 were elevated 5-20 fold compared to the wild type strain,

suggesting a deregulation of the gene cluster (Figure S7). Each of the relevant coding regions in Locus III (*msmeg_6789*, *lerK*, *derI2*, and *lerI*) and 1000 bp upstream of Locus III were sequenced; no mutations were identified. Because HH001 was capable of growth with L-threitol, it was used during studies regarding metabolism of L-threitol.

3.2 RNA extraction and qRT-PCR

The extraction of RNA for qRT-PCR from wild-type *M. smegmatis* mc^{2}155 was performed as described previously[14]. Briefly, 5 mL of aerated culture in exponential phase (OD$_{600}$ 0.4-0.5) was added into 10 mL of RNAprotect Bacteria Reagent (Qiagen), After centrifuging (10 min, 4,000g, 4 ^{o}C), the pellet was suspended in 200 μL water, and the culture was transferred to screwcap tubes containing 0.5 mL of 0.1 mm zirconia/silica beads (Biospec) before 700 μL of Buffer RLT (Qiagen) was added. The bacteria were lysed using a vortex mixer (Fisher Scientific) at maximum speed for 5 min, and the cell lysates were recovered by centrifugation (5 min, 13,000 g, 4 ^{o}C). RNA was purified from the lysate using the RNeasy kit (Qiagen) and treated with DNase (Qiagen) both on and off column according to the manufacturer's instruction. The purified RNA was used as template to synthesize cDNA with a Protoscript first strand cDNA synthesis kit (NEB) following the manufacturer's instruction. Reactions for qPCR were set up using the iTaq Universal SYBR green supermix (Bio-Rad) and performed on a Lightcycler 480II (Roche). The reactions were 10 μL and contained 1μL of cDNA, 0.5 μM of primers and 2x SYBR supermix. Sequences of each primer are given in Table S3. Reactions were heated to 95 ^{o}C for 30s before cycling for 40 cycles of 95 ^{o}C for 10 s, 50 ^{o}C for 20 s, and 72 ^{o}C for 3 0s. Fluorescence was measured at the end of each cycle after heating to

80 $^{\circ}$C. At the end of the PCR, melt curve analysis was performed, and PCR products were analyzed on agarose gel to ensure product specificity. Concentrations of cDNA for each transcript were measured twice, and technical triplicates were used for each duplicate measurement. Transcript levels in cells from each condition were normalized with levels of *msmeg_2758* (sigma 70) transcript. Fold changes of transcripts were determined for cells grown with each carbon source relative to levels in cells of the same strain grown with glycerol.

3.3 Deletion of *M. smegmatis* genes

All deletion mutations of the *M. smegmatis* genome were performed according to the following protocol for double homologous recombination. Regions (~1000 bp) that flanked the upstream and downstream of the coding region of interest were amplified using the primers indicated in Table S4. Each product included ≥27 bp of the coding region. Using the Gibson Assembly Cloning Kit (NEB), the upstream and downstream fragments were ligated into pK18mobsacB[15] that had been linearized with *Xba*I/*Bam*HI or *Eco*RI/*Hind*III. *M. smegmatis* was then electroporated with the assembled plasmid as previously described[16]. Single crossover strains were isolated as kanamycin resistant and were incubated overnight with shaking at 37 °C in nonselective LB broth. A 100 µL aliquot of a 100-fold diluted overnight culture was plated on LB supplemented with 10% sucrose. The resulting colonies were sequentially patched onto LB and LB with kanamycin. Wild type revertants were separated from mutants by colony PCR using primers that directed amplification across the truncated region. Mutant genotypes were confirmed by sequencing the entire region of the genome that would be available for

recombination with the vector. All mutations were in-frame and introduced no additional nucleotides.

4. Phylogenetic distribution of tetritol metabolic pathways

The Uniprot IDs for the homologues of D-threitol dehydrogenase, erythritol/L-threitol dehydrogenase, D- or L-erythrulose kinases, and L-erythrulose-4P epimerase were extracted from their clusters in their respective SSNs. The Uniprot IDs then were converted into NCBI Taxonomy IDs by Uniprot Retrieve/ID mapping tool (http://www.uniprot.org/uploadlists/). Finally, the NCBI Taxonomy IDs were used to generate the phylogenetic trees using PhyloT (http://phylot.biobyte.de/); the trees were visualized and colored by iTOL (http://itol.embl.de/)[17]. The presence of L-threitol catabolic pathways was determined by the co-existence of erythritol/L-threitol dehydrogenase and L-erythrulose-4P epimerase.

5. Supporting Figures and Tables

Table S1. Primers used for enzyme preparation

Gene Name	Primers seq 5' to 3'
msmeg_3603	ggatacac**catatg**gtaggaaccgagatcag ttgtggatgtcggtgca**ctcgag**cagagcgttg
msmeg_3604	gagcacc**catatg**gatgtgcgttcggcattc gtcgtcgagcgcgct**ctcgag**gacccttctg
msmeg_3605	ggaaaggaaac**catatg**aacgccgtcgcagc gtgcgcccgctcag**ctcgag**tgccgaacgcac
msmeg_3607	gcaaagcg**catatg**actcaggcacaagag gcacgccgacggca**aagctt**ccctgttgtgac
msmeg_3265	cgaaaggtattt**catatg**tccaatcaagtcc ctcggctcatctcag**aagctt**tcgctatgc
msmeg_3266	catagcgaaagcg**catatg**atgagccgagaatc caatgtcatgtcgc**aagctt**ctacaggttgac
msmeg_3271	gagcggtt**gctagc**accaagt tgttcaacga tc cgacccgtagtg**ctcgag**gttctctcct tg
msmeg_3275	ccaaggagagaa**catatg**gcactacgggtc caaacgtcgggcgc**ctcgag**actagctcg
msmeg_3598 *(LIC SITES)*	**(tacttccaatccatggcg)**ggcgacaccgcg ccccgtctgggtcgg**(gattgtcatttccacctat)**
msmeg_3599 *(LIC SITES)*	**(tacttccaatccatggcg)**ggtgaccccgcc ggccggtcttcgtcgg**(gattgtcatttccacctat)**
msmeg_6785	catctgcgcgtacg**catatg**cctgacgctcg cacctaccggcc**aagctt**cactcaccggtg
msmeg_6787	ggagaaggagagc**catatg**gcgttgaagatcg ggtgccgatcca**aagctt**tgcggcgccgagc
msmeg_6788	gaaggacgga**catatg**acgtacctcctgaac gtcggtcgggc**ctcgag**ccgggtgaggtgc

Table S2. Crystallographic Parameters

TARGET IDENTITY		
PROTEIN LOCUS / PDB CODE	MSMEG_3598 / 4RS3	MSMEG_3599 / 4RSM
Co-crystallization Ligand	10 mM xylitol	10 mM DL-threitol
DATA COLLECTION*		
Space Group	P212121	P21
Unit Cell (Å , Degree)	a=62.8, b-63.1, c=73.1	a=62.7, b=128.3, c=66.6, beta=93.1
Resolution (Å)	50.0 - 1.40 (1.42-1.40)	50-1.60 (1.63-1.60)
Completeness (%)	99.7 (98.5)	96.2 (90.7)
Redundancy	14.6 (13.9)	6.7 (6.8)
Mean(I)/sd(I)	4.1	2.5
Rsym	0.111 (0.72)	0.093 (0.68)
Heavy Atom Phasing	SEMET SAD	SEMET SAD
REFINEMENT STATISTICS*		
Resolution (Å)	50-1.4 (1.44-1.40)	50-1.60 (1.64-1.60)
Rcryst (%)	12.5 (15.2)	17.7 (22.8)
Rfree (%, 5% of data)	16.3 (21.7)	21.7 (25.7)
Molecules per ASU	1	4
Residues / Waters / Atoms total	315 / 228 / 2647	1257 / 890 / 10269
Bfactor Protein/Waters/Ligand	18.1 / 29.3 / 18.7	28.0 / 31.8 / 16.1
Ligand [number]	xylitol [x1]	D-threitol [x4]
RMSD Bond Lengths (Å) / Angles (°)	0.019 / 1.959	0.012 / 1.55

* Data in parenthesis are the statistics in the highest resolution bin

Table S3. Primers used for gene transcript analysis

Gene Name	Primers seq 5' to 3'
msmeg_3598	acgttgtccttggtgatcg, tttcatcggcacctcactg
msmeg_3599	acgttgtccttggtgatcg, gtcaagagcggtgatttcatc
msmeg_3603	aggactttcacgttctcacc, ccgccatcgacatgctc
msmeg_3604	ccatgtcggtcttcacgtatc, acctcacgaaatcccttgc
msmeg_3605	agaactcacgagatggattgc, tcggctacatccagaacatg
msmeg_3606	catctcccggttcaaggtc, gggttgtggatcttgatctcg
msmeg_3607	tcaacatggcgtcacagg, tgggcgagatggtgttg
msmeg_3262	caggttgaccgcgatgg, tggacgtgctggtgaac
msmeg_3264	gttgatcagcgacgggtac, gtaccagaccaacgagatcac
msmeg_3265	tccaatcaagtccccgaaaag, tgagatcgcttgcacagatg
msmeg_3266	acggcaactacaagaccaac, gccttgttcatcgattccttg
msmeg_3271	caccaagtacgaagagctgt, gagccacatcacggtcag
msmeg_3272	tacaaggaagcactcaaggg, cttcggcgatcagacgtg
msmeg_6785	tcaaggatccgcacgtaac, aatgggattgctctacggtg
msmeg_6787	agatgtcattgaccttcgcg, tccgtcctgtcgaacaatg
msmeg_6788	acaacggtgaatccttgctg, tgattctcaacggcctgg
msmeg_6789	aacagctccatcagacgc, atccgacatgagaaccgattc
msmeg_6790	catcaaggacgtcgaccac, tgatgtgccagaagttcgag
msmeg_6791	attcaaggtccagatcgcc, ctcgaattcccgtgcctc
msmeg_6792	gaacccaaggattatccggag, agatcagcaccaacggaag

Table S4. Primers used for the construction of gene knockout

| Primer Name	Primers seq 5' to 3'
msmeg_3599	cacacaggaaacagctatgacatgattacggctaccgcagcgcccggttcgccaccccgggttgatcagttcg gtgaggccgtcgaggaacgcggcgggtcacccgcaccgcaagccgtacggcttgcggtgcgggtgaccc cgccgcgttcctcgacggcctcaccgaactgatcaacacgttgtaaaacgacggccagtgccaagctcttgaca tcctcgcccttgaccagcgcgtc;
msmeg_3607	cacacaggaaacagctatgacatgattacggcgtccagatcgccgaggagttcggcatctgtcggacgcgagg aacactgcggccgcggcgccgatcccggacgcggcgccggtcaccagctggtgaccggcgccgcgtccgg gatcggcgccgcggccgcagtgttcctcgcgtccgacacgttgtaaaacgacggccagtgccaagctttcagc gccgtggccgccaggaacccgcac;
msmeg_3271	ctgcaggtcgactctagatgtcggtgtcgggttcgagagctacggtggcaccaacgcccacgacgtaacggga attggtgggcgttggtgccacctcgctggcattgtgcgagctcggtacccggggatccgcgggctctcagctgt ggtg
msmeg_3572	ctgcaggtcgactctagagagaagcggatcgcggtgattctcaaccttgctggtcgcgccgacgacgacccgt agtggtcggcgcgaccagcaagtcggcggagaaggtctcgagctcggtacccggggatcccgttgctcttcg catgtgacgcatc
msmeg_3265	cacacaggaaacagctatgacatgattacgcacgtcgcccaccgccgaccgcagcgaactcttctggaattcg gtgagcgggaactgatggcgctgcggtaccgcgacctcctcgagccgcgggctcgaggaggtcgcggtaccg cagcgccatcagttcccgctcaccgaattccagaagacgttgtaaaacgacggccagtgccaagctgtggtcca cgacgtcgggccggccttgtgc
msmeg_3266	cacacaggaaacagctatgacatgattacgggcccgcgtgggccgagaccatggtgatcctcgtcaccgccgt agatgtcctgcagcgccacggccgccgcgccggccaggcccatggccggccatgggcctggccggcgcgg cggccgtggcgctgcaggacatctacggcggtgacgaacgttgtaaaacgacggccagtgccaagcttagcg cgcggtgatgaacacgatgacccac
msmeg_3271	cacacaggaaacagctatgacatgattacgccagcggctcgatcaagatcggcgaccgcgcagcgaggtggc accggcgtccggcgtgccaacgcccacgacgtaacgggaattggcgtcgacgccaattcccgttacgtcgtgg gcgttggcacgccggacgccggtgccacctcgctgacgttgtaaaacgacggccagtgccaagctcgcgggc tctcagctgtggtgtgctcggat
msmeg_3272	cacacaggaaacagctatgacatgattacgtgttgcgcgaccgtgggtacacgatcgtcgggtcgaactcgtag ccgagccactcgcgcggacgcggtcgtccgcggcgaggtcgcccttaagggcgacctcgccgcggacgac cgcgtccgcgcgagtggctcggctacgagttcgaccacgttgtaaaacgacggccagtgccaagctgcctgtc ggcgacgtccttgtcatagacgc
msmeg_6785	cacacaggaaacagctatgacatgattacggtgggccgacgcggccggcggaacctcgggctgccaggccg cccgtccgatgaacaacccgtaaccacgcgactcggcgagcccccttgttaacaaggggctcgccgagtcgcgt ggttacgggttgttcatcggacgggcggcctggcagacgttgtaaaacgacggccagtgccaagctgcgcac gtgcgtgttcggtgtgaactccgc
msmeg_6787	cacacaggaaacagctatgacatgattacgcgcagcaaggattcaccgttgtgcgcccgggaactcgaggccc agccactccttcaccaggacgcggtcgtcggcctcgaggtccttgcgcgcaaggacctcgaggccgacgacc gcgtcctggtgaaggagtggctgggcctcgagttcacgttgtaaaacgacggccagtgccaagctacagaccg gcgacgatccgatcctgcgtgg
msmeg_6788	cacacaggaaacagctatgacatgattacgggcagccaatgggaccgcgcggatgcgaaaatcgggcgtgc cgacgctcttctccccgagcacctcggtgagcagatcggggttggcggcgccgccaacccgatctgctcacc gaggtgctcggggagaagagcgtcggcacgcccgatacgttgtaaaacgacggccagtgccaagctcgatg cacagcagcggcaccagaccgtgat

Table S5. Names and physiological catalytic functions of identified enzymes

Suggested Name	Locus Tag	Uniprot ID	Physiological Function
EltD	MSMEG_3265	A0QXD8	Erythritol/L-Threitol Dehydrogenase
EltP	MSMEG_3266	A0QXD9	Erythritol/L-Threitol SBP
DerK	MSMEG_3271	A0QXE4	D-Erythrulose Kinase
DerI1	MSMEG_3272	A0QXE5	D-Erythrulose-4P Isomerase
XypA	MSMEG_3598	A0QYB3	Xylitol SBP
ThpA	MSMEG_3599	A0QYB5	D-Threitol SBP
DthD	MSMEG_3607	A0QYC2	D-Threitol Dehydrogenase
LerI	MSMEG_6785	A0R756	L-Erythrulose-4P Epimerase
DerI2	MSMEG_6787	A0R757	D-Erythrulose-4P Isomerase
LerK	MSMEG_6788	A0R758	L-Erythrulose Kinase

Table S6. Kinetic data for characterized enzymes

MSMEG_3604 (NADP$^+$/NADPH)	k_{cat} (s^{-1})	K$_m$(mM)	k_{cat}/K$_m$(M^{-1}s^{-1})
D-sorbitol	28.4 $\pm$ 0.8	19.6 $\pm$ 0.06	1.4 x 10^3
D-iditol	3.62 $\pm$ 0.03	8.06 $\pm$ 0.2	4.5 x 10^2
D-galactitol	4.53 $\pm$ 0.1	16.3 $\pm$ 0.9	2.8 x 10^2
D-allitol	1.40 $\pm$ 0.08	6.67 $\pm$ 1.0	2.1 x 10^2
D-sorbose	11.9 $\pm$ 0.5	39.7 $\pm$ 3.0	3.0 x 10^2
MSMEG_3605 (NAD$^+$/NADH)	k_{cat} (s^{-1})	K$_m$(mM)	k_{cat}/K$_m$(M^{-1}s^{-1})
Xylitol	0.68 $\pm$ 0.04	1.35 $\pm$ 0.3	5.0 x 10^2
D-sorbitol	1.07 $\pm$ 0.04	3.80 $\pm$ 0.5	2.8 x 10^2
MSMEG_3607 (NAD$^+$/NADH)	k_{cat} (s^{-1})	K$_m$(mM)	k_{cat}/K$_m$(M^{-1}s^{-1})
D-threitol	6.97 $\pm$ 0.1	0.52 $\pm$ 0.04	1.3 x 10^4
D-threose	0.71 $\pm$ 0.01	1.20 $\pm$ 0.08	6.0 x 10^2
D-erythrulose	7.23 $\pm$ 0.2	0.048 $\pm$ 0.004	1.5 x 10^5
MSMEG_3265 (NAD$^+$/NADH)	k_{cat} (s^{-1})	K$_m$(mM)	k_{cat}/K$_m$(M^{-1}s^{-1})
L-threitol	10.2 $\pm$ 0.03	1.64 $\pm$ 0.2	6.2 x 10^3
Erythritol	8.46 $\pm$ 0.2	1.50 $\pm$ 0.02	5.6 x 10^3
L-arabitol	8.73 $\pm$ 0.4	1.47 $\pm$ 0.3	5.9 x 10^3
Xylitol	8.20 $\pm$ 0.5	2.76 $\pm$ 0.5	3.0 x 10^3
Ribitol	6.56 $\pm$ 0.08	2.21 $\pm$ 0.1	3.0 x 10^3
D-erythrulose	9.88 $\pm$ 0.7	0.026 $\pm$ 0.01	3.8 x 10^5
L-erythrulose	13.3 $\pm$ 0.7	0.066 $\pm$ 0.01	2.0 x 10^5
D-xylulose	19.3 $\pm$ 1.9	0.73 $\pm$ 0.2	2.6 x 10^4
L-xylulose	7.75 $\pm$ 0.4	0.25 $\pm$ 0.04	3.1 x 10^4
Ribulose	21.0 $\pm$ 0.7	1.23 $\pm$ 0.08	1.7 x 10^4
MSMEG_3271 (kinase)	k_{cat} (s^{-1})	K$_m$(μM)	k_{cat}/K$_m$(M^{-1}s^{-1})
D-erythrulose	2.88 $\pm$ 0.08	4.6 $\pm$ 0.6	6.3 x 10^5
L-erythrulose	0.091 $\pm$ 0.003	396.2 $\pm$ 62	2.3 x 10^2
Dihydroxyacetone	0.063 $\pm$ 0.002	338.9 $\pm$ 41	1.8 x 10^2
MSMEG_6788 (kinase)	k_{cat} (s^{-1})	K$_m$(μM)	k_{cat}/K$_m$(M^{-1}s^{-1})
L-erythrulose	1.79 $\pm$ 0.03	3.0 $\pm$ 0.3	6.0 x 10^5
D-erythrulose	1.60 $\pm$ 0.05	10.6 $\pm$ 1.4	1.5 x 10^5
Dihydroxyacetone	2.12 $\pm$ 0.05	37.8 $\pm$ 3.8	5.6 x 10^4

A:
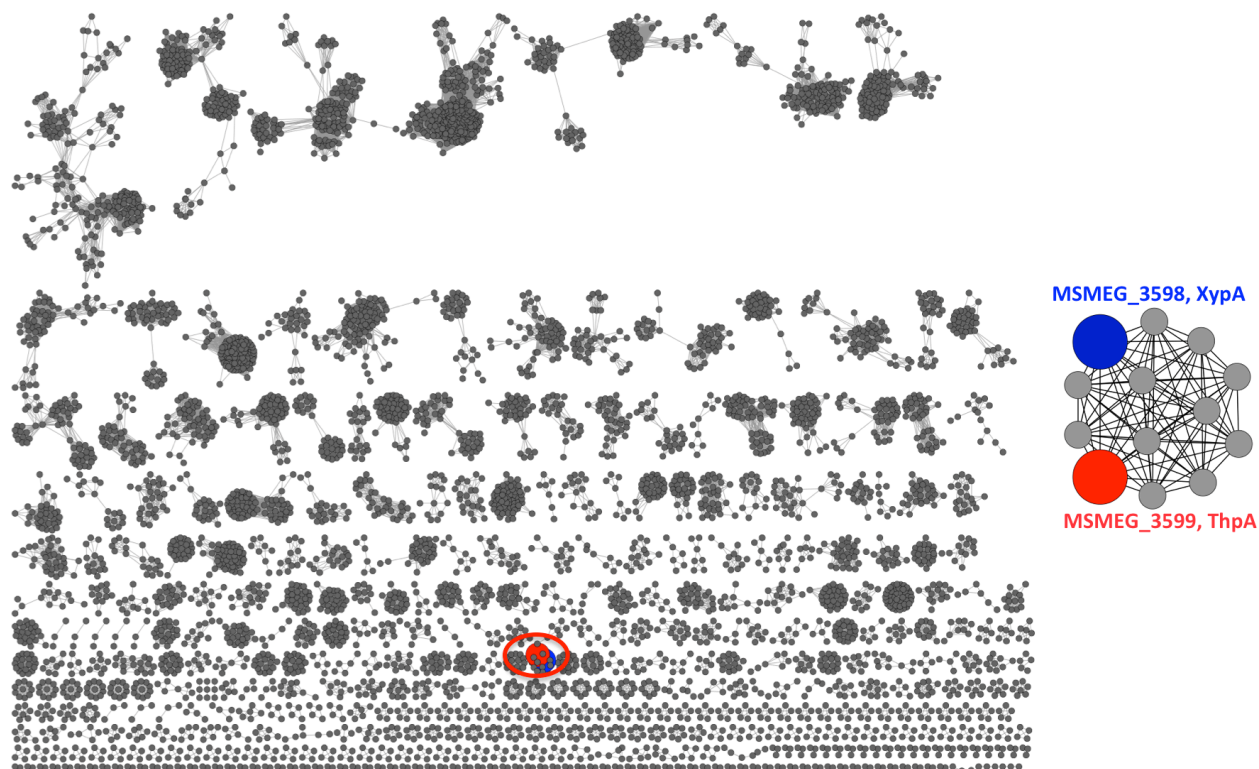
B:

Differential Scanning Fluorimetry of MSMEG_3598(XypA) and MSMEG_3599(ThpA)*

LIGAND	MSMEG_3598 (ΔT, C°)	MSMEG_3599 (ΔT, C°)
Xylitol	14.7	12.6
D-sorbitol (L-sorbitol)	-1.4 (13.0)	-0.5 (2.8)
D-sorbose (L-sorbose)	3.7 (11.5)	1.7 (7.5)
D-idose (L-idose)	-0.2 (7.8)	0.4 (4.0)
D-threitol (L-threitol)	6.2 (-0.3)	13.6 (3.1)
D-xylose (L-xylose)	-0.2 (3.0)	0.3 (6.7)
D-iditol (L-iditol)	1.9 (-0.1)	7.6 (0.2)

* L-isomers and their thermal stabilization effect shown in parenthesis

Figure S1: **Panel A**: SSN for ABC SBPs (PF13407, alignment score 100, 55% identity, 90% repnode network) that illustrates the likely breadth of specificity diversity; the alignment score was selected to separate different experimentally determined specificities (data to be published). The D-threitol transporters are highlighted. **Panel B**: Detailed DSF results for two ABC SBPs from *M. smegmatis* mc^{2}155 (MSMEG_3598,XypA & MSMEG_3599,ThpA: 85% identity).
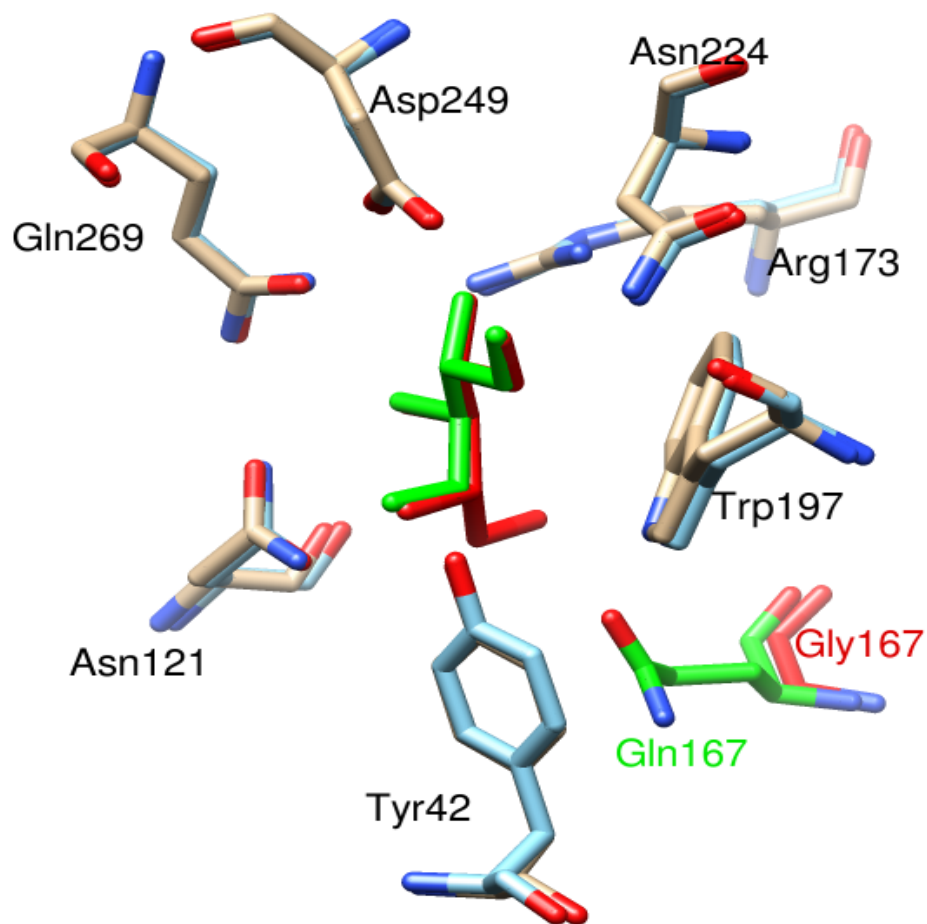
Figure S2: Comparison of the ligand binding sites for MSMEG_3598 (XypA) and MSMEG_3599 (ThpA). Substrate binding site overlay of ThpA (PDB: 4RSM, cyan) and XypA (PDB: 4RS3, blue) with D-threitol (green) and xylitol (red). The unique residue (Gln167 in ThpA and Gly167 in XypA) is highlighted in green and red, respectively.

A:
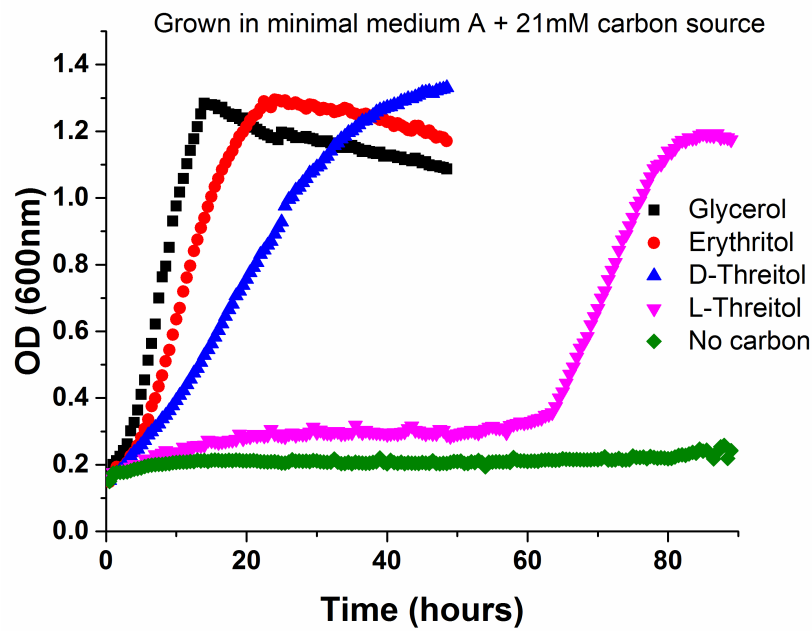
B:
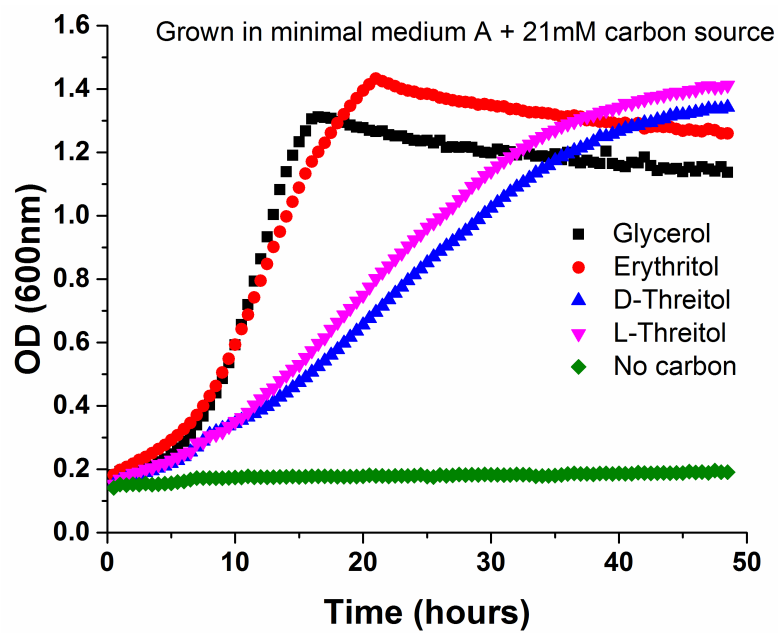
Figure S3: Growth curves for *M. smegmatis* with tetritols. **Panel A**: Wild type *M. smegmatis* with tetritols. **Panel B**: Threitol-adapted *M. smegmatis* (strain HH001) with tetritols.
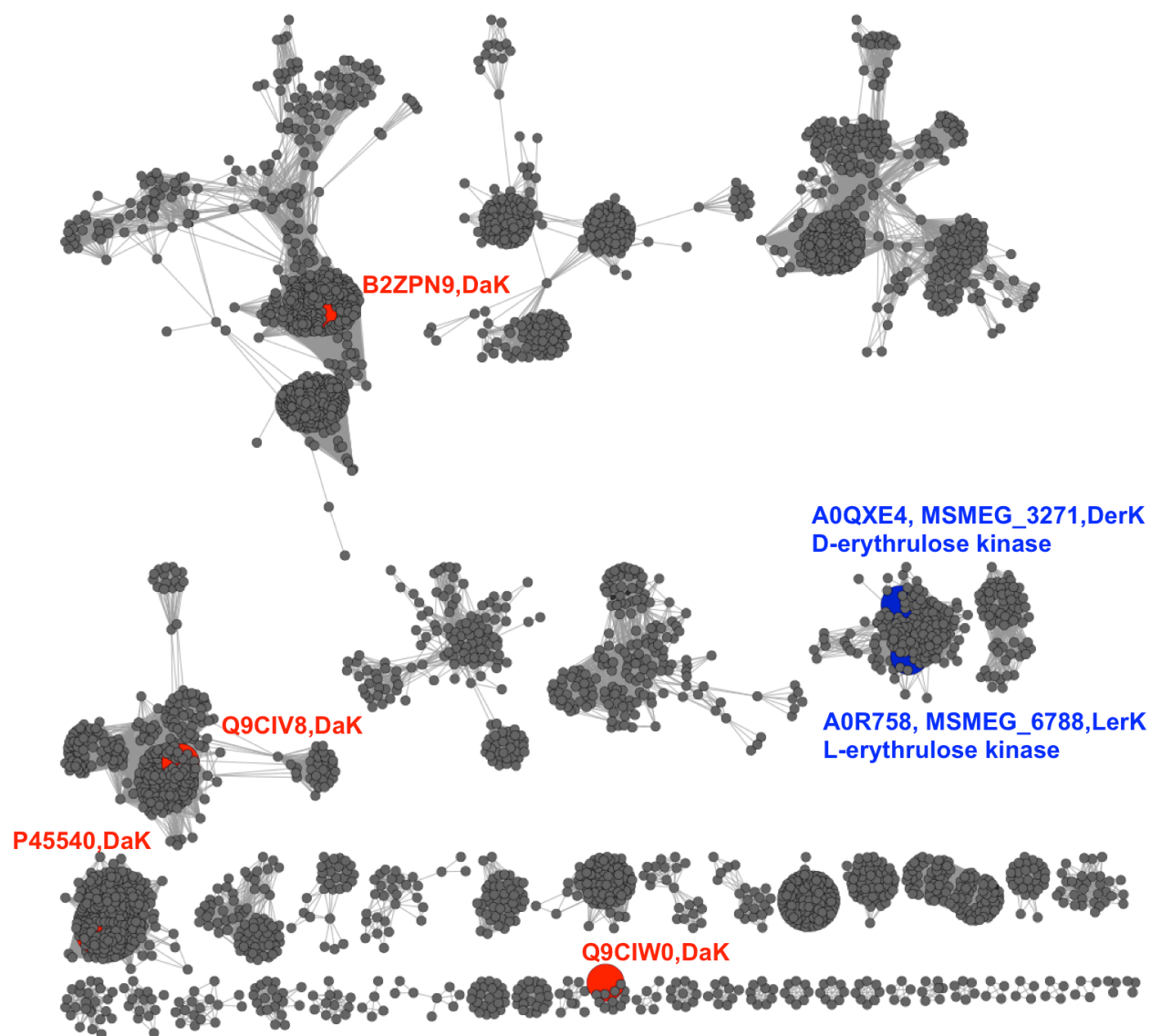
Figure S4: SSN for PF02733 (Dak1; alignment score 120, 40% sequence identity, 100% rep node network,singletons omitted). Blue nodes represent two annotated dihydroxyacetone kinases (MSMEG_3271, DerK and MSMEG_6788, LerK, which are confirmed here as D or L-erythrulose kinase) in *M. smegmatis*. Red nodes indicate authentic dihydroxyacetone kinases (DaK)[18].
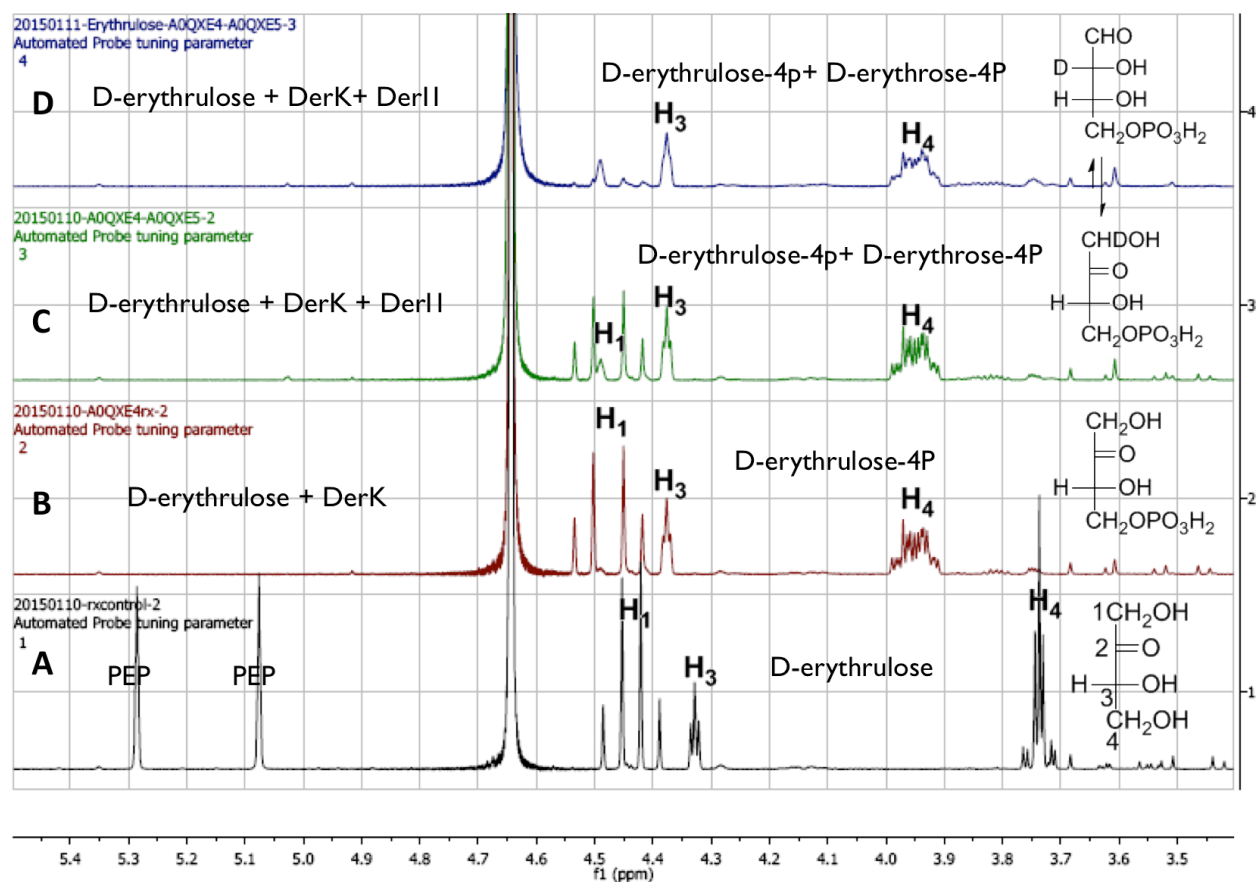
Figure S5: Confirmation of the D-erythrulose kinase (DerK, MSMEG_3271) and D-erythrulose-4P isomerase (DerI1, MSMEG_3272) activities. ^{1}H-NMR spectra were recorded after D-erythrulose was incubated with DerK and then after addition of DerI1. The resonances were assigned for each compound. The reaction was monitored before (**Panel A**) and 30 min after (**Panel B**) addition of DerKto the standard kinase assay described in Materials and Methods. Then, 1 µM DerI1was added; spectra were recorded after 2 hr (**Panel C**) and 12 hr (**Panel D**) incubation at 30 °C.
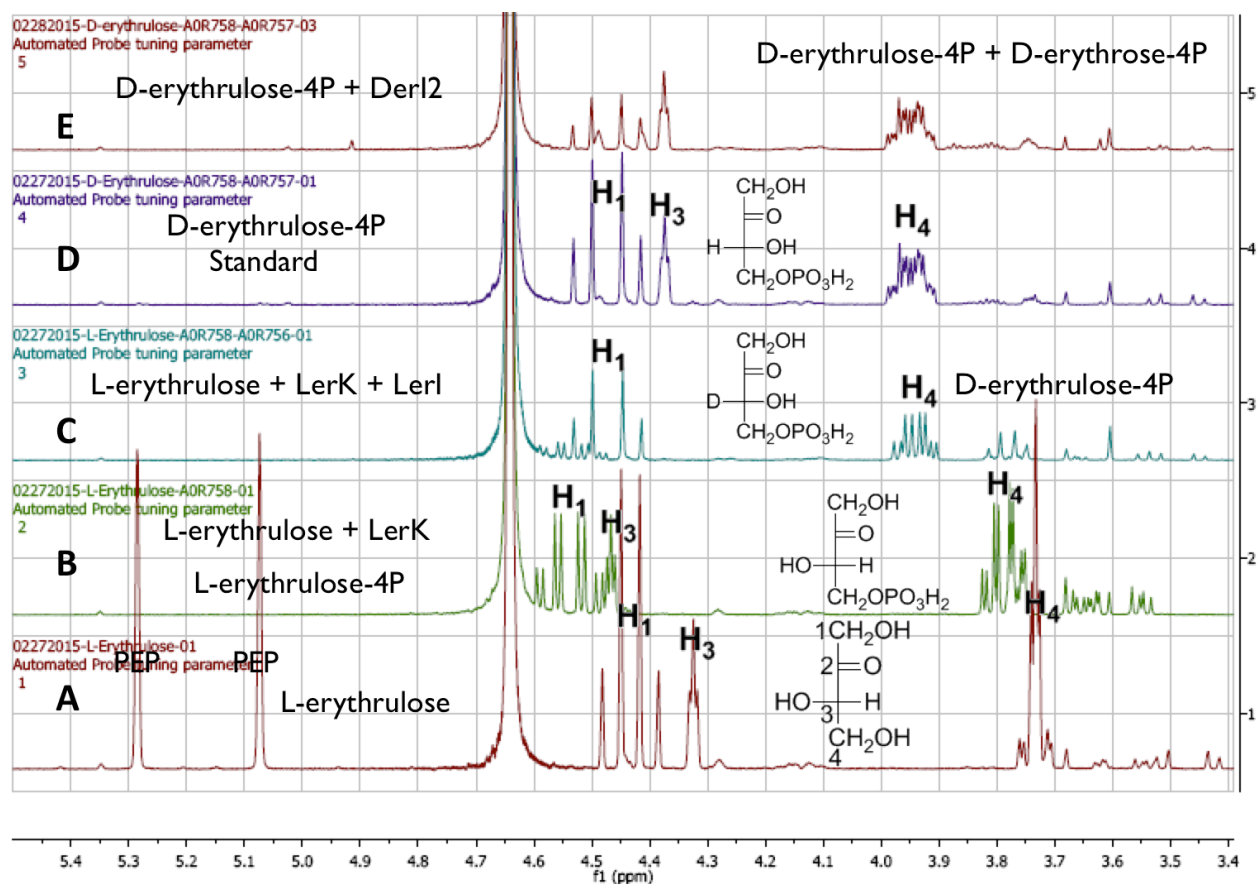
Figure S6: Confirmation of L-erythrulose kinase (LerK, MSMEG_6788) and L-erythrulose-4P isomerase (LerI, MSMEG_6785; DerI2, MSMEG_6787) activities. ^{1}H-NMR spectra were recorded for L-erythrulose incubated with LerK and LerI as well as for D-erythrulose-4P incubated with DerI2. The resonances were assigned for each compound. The reaction was monitored before (**Panel A**) and after (**Panel B**) the addition of LerK to the standard kinase assay described in Materials and Methods. Then 1 μM LerI was added; a spectrum was recorded after 1 hr incubation at 30 ^{o}C (**Panel C**). A solution containing a D-erythrulose-4P was prepared as described above (**Panel D**) and incubated with 1 μM DerI2; a spectrum was recorded after 2 hr at 30 ^{o}C (**Panel E**).
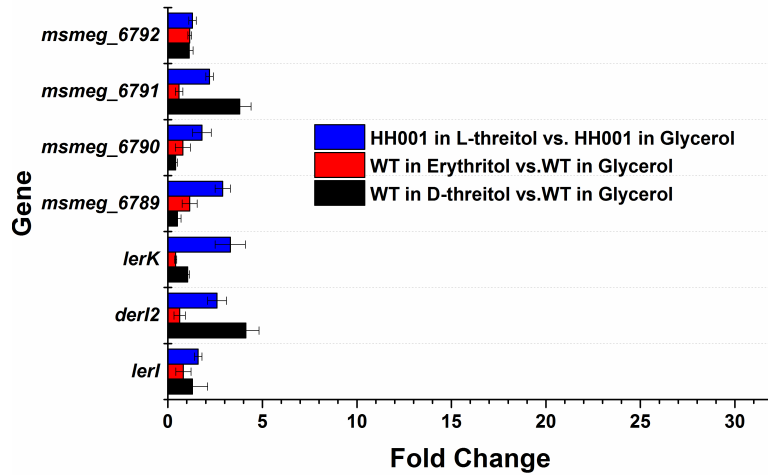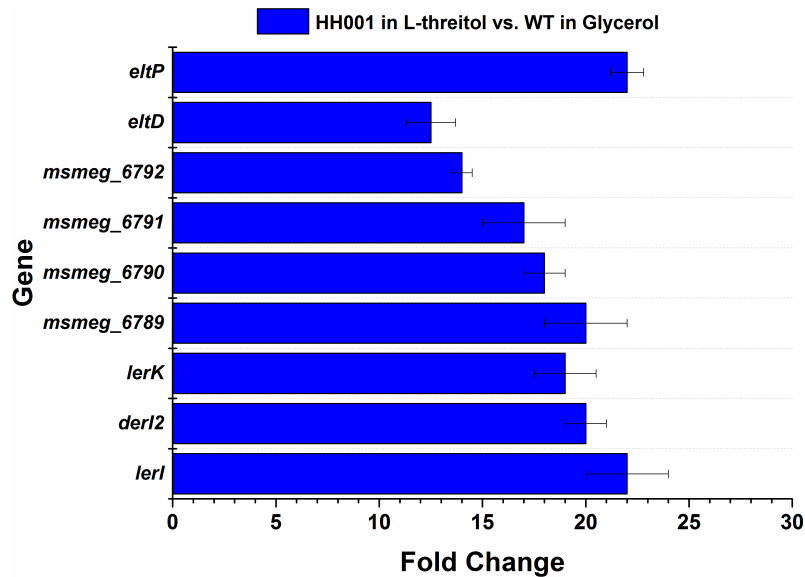
A:

Legend:
- HH001 in L-threitol vs. HH001 in Glycerol (blue)
- WT in Erythritol vs. WT in Glyerol (red)
- WT in D-threitol vs.WT in Glycerol (black)

Genes (top to bottom): dthD, msmeg_3606, msmeg_3605, msmeg_3604, msmeg_3603, thpA, xypA

X-axis: Fold Change (0–16)
Y-axis: Gene

B:

Legend:
- HH001 in L-threitol vs. HH001 in Glycerol (blue)
- WT in Erythritol vs. WT in Glycerol (red)
- WT in D-threitol vs. WT in Glycerol (black)

Genes (top to bottom): derl1, derK, eltP, eltD, msmeg_3264, msmeg_3262

X-axis: Fold Change (0–40)
Y-axis: Gene

Figure S7: Regulation of transcripts in cells grown with D-threitol, erythritol, or L-threitol. . **Panel A**: Relative transcript levels of Locus I genes (Figure 2A); **Panel B**: Relative transcript levels of Locus II genes(Figure 2B); **Panel C**:Relative transcript levels of Locus III genes(Figure 2C); **Panel D**: Comparison of transcript levels of genes encoding proteins for import and catabolism of L-threitol in L-threitol-grown HH001 and glycerol-grown wild type *M. smegmatis*. HH001: L-threitol adapted strain; WT: Wild type *M. smegmatis* (Refer to strains and growth conditions in Materials and Methods)
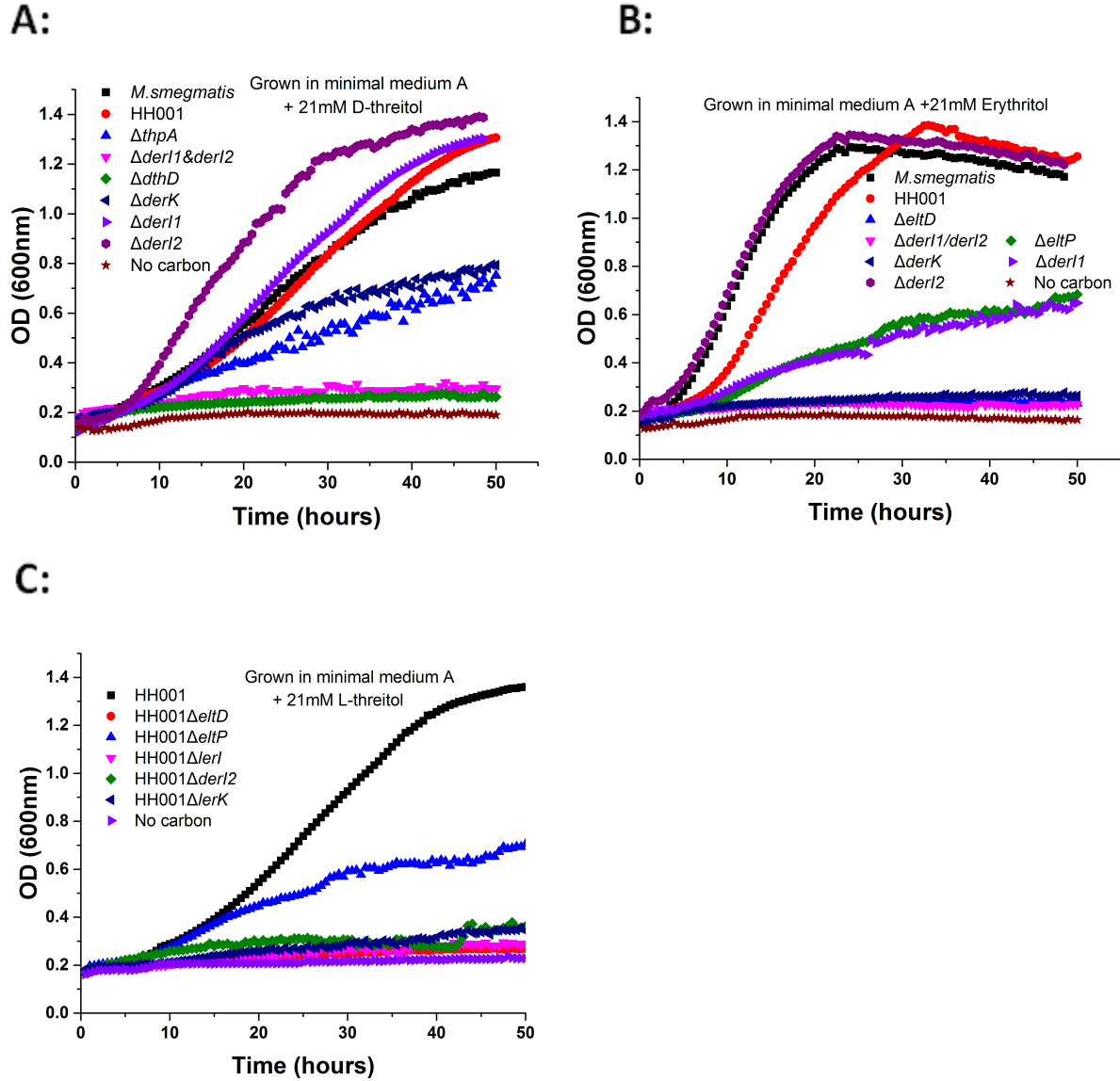
Figure S8: Growth of wild type *M. smegmatis* mc^{2}155, deletion mutants of the wild type strain, the L-threitol adapted strain (HH001), and deletion mutants of HH001 with D-threitol (**Panel A**), erythritol (**Panel B**), or L-threitol (**Panel C**). See Figure 2 for gene assignments and relative genomic locations.
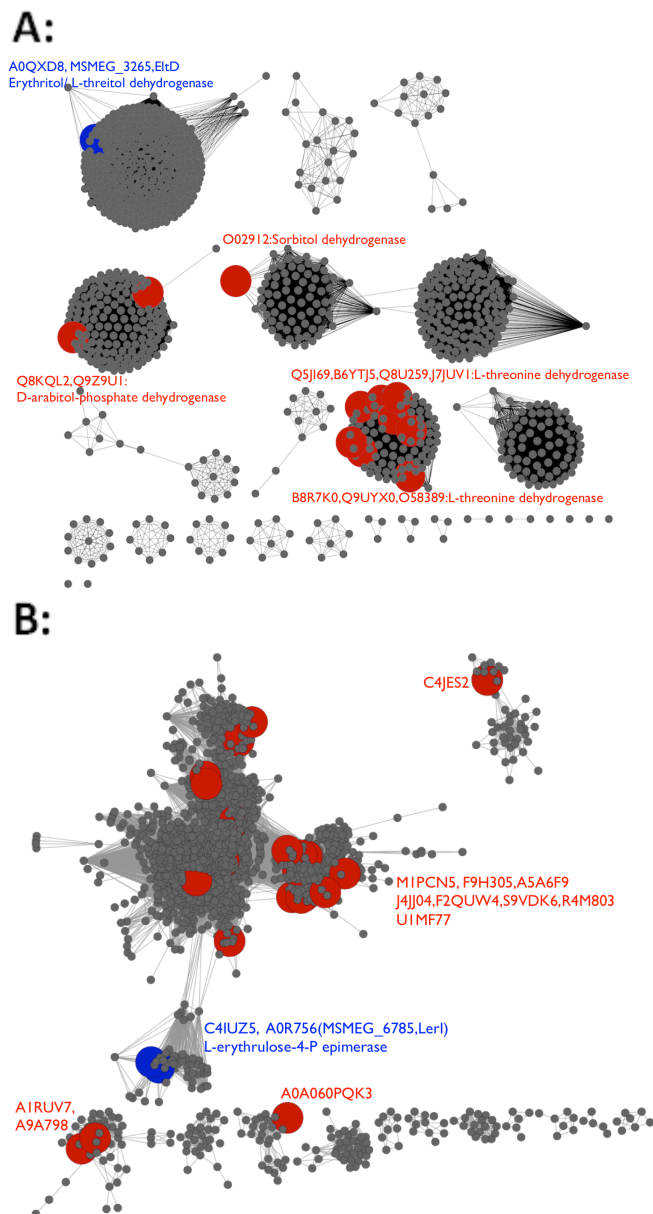
Figure S9: **Panel A**: SSN for MSMEG_3265 (alignment score 70, 40% sequence identity, 100% rep node network, singletons omitted), characterized here as erythritol/L-threitol dehydrogenase (EltD). Red nodes represent reviewed enzymes[19]. Note that EltD is a member of a large cluster with no other characterized members. **Panel B**: SSN for PF00121 (triosephosphate isomerase, alignment score 60, 45% identity, 100% repnodes, Singletons omitted), which includes MSMEG_6785 (LerI): Red nodes represent authentic triosephosphate isomerases.[20] Blue nodes represent the authentic L-erythrulose-4P epimerase (D0AT64) from *Brucella abortus*[21]. Note that LerI represents a novel function within the network of characterized triosephosphate isomerases.
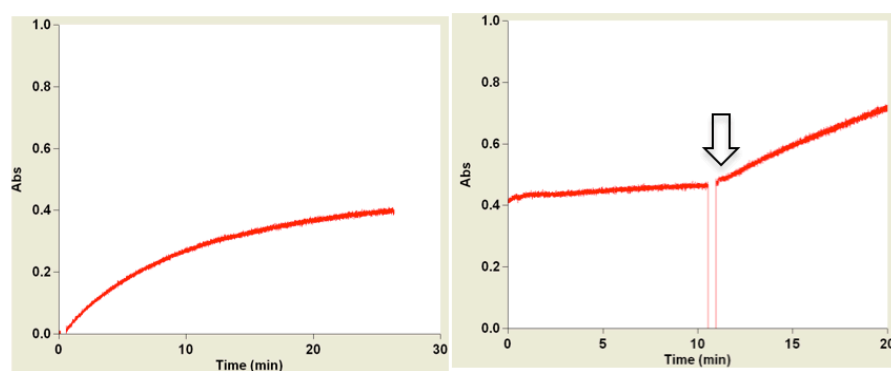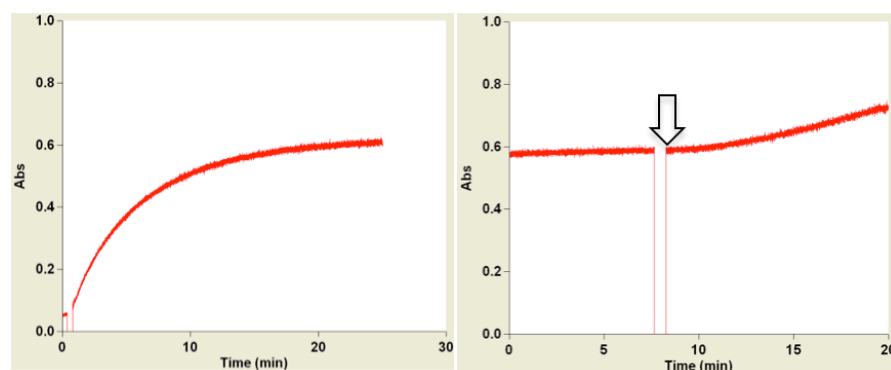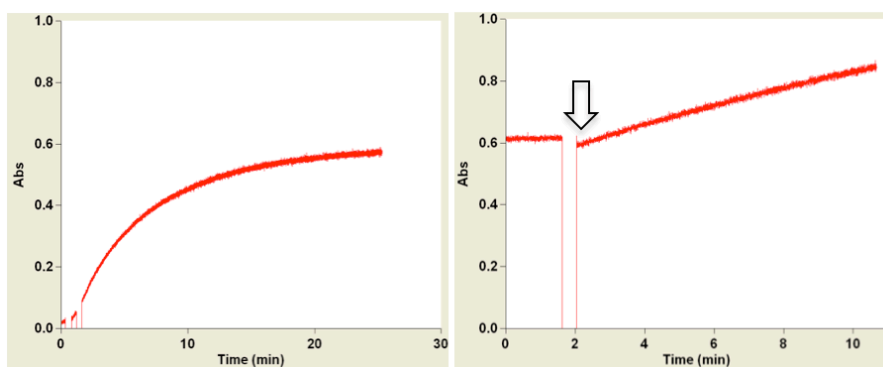
Figure S10: Erythritol/L-threitol oxidation product confirmation by *in situ* UV (340 nm) monitoring. A 200 μL reaction solution contained 1 mM polyol, 1 mM Mg^{2+}, 1 mM NAD^+, 100 mM glycine/NaOH pH 9.5 and 1 μM of dehydrogenase (MSMEG_3265, EltD). After the reaction reached equilibrium, 5 μM of the appropriate substrate-specific erythrulose kinase was added (indicated by arrows). A: Erythritol oxidation catalyzed by EltD. The product was confirmed by observing activity following the addition of D-erythrulose kinase (MSMEG_3271, DerK). B: L-threitol oxidation catalyzed by EltD. The product was confirmed by observing

activity following the addition of D-erythrulose kinase (MSMEG_3271, DerK). C: L-threitol oxidation by EltD. The product was confirmed by observing activity following the addition of L-erythrulose kinase (MSMEG_6788, LerK).
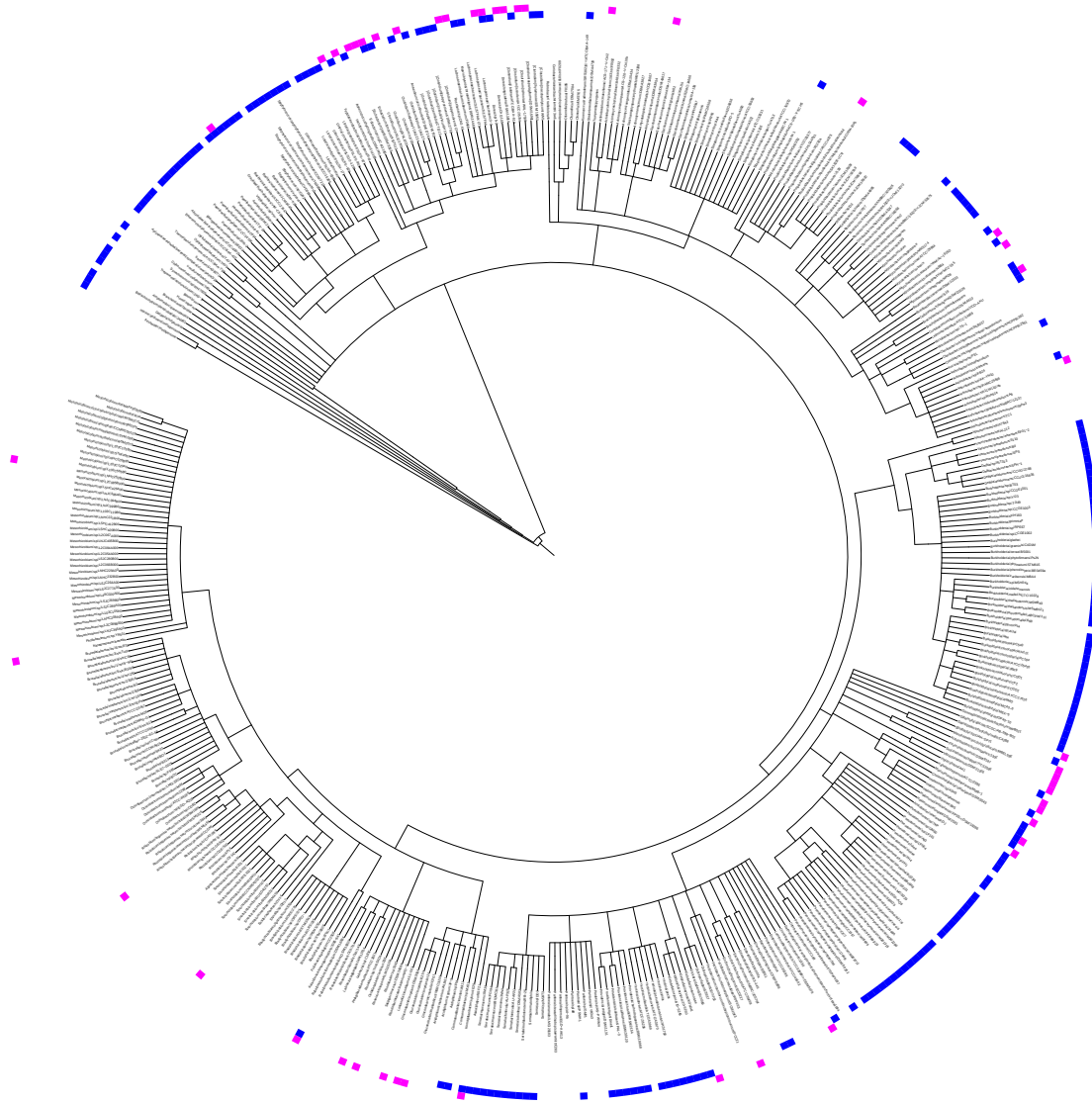
A:
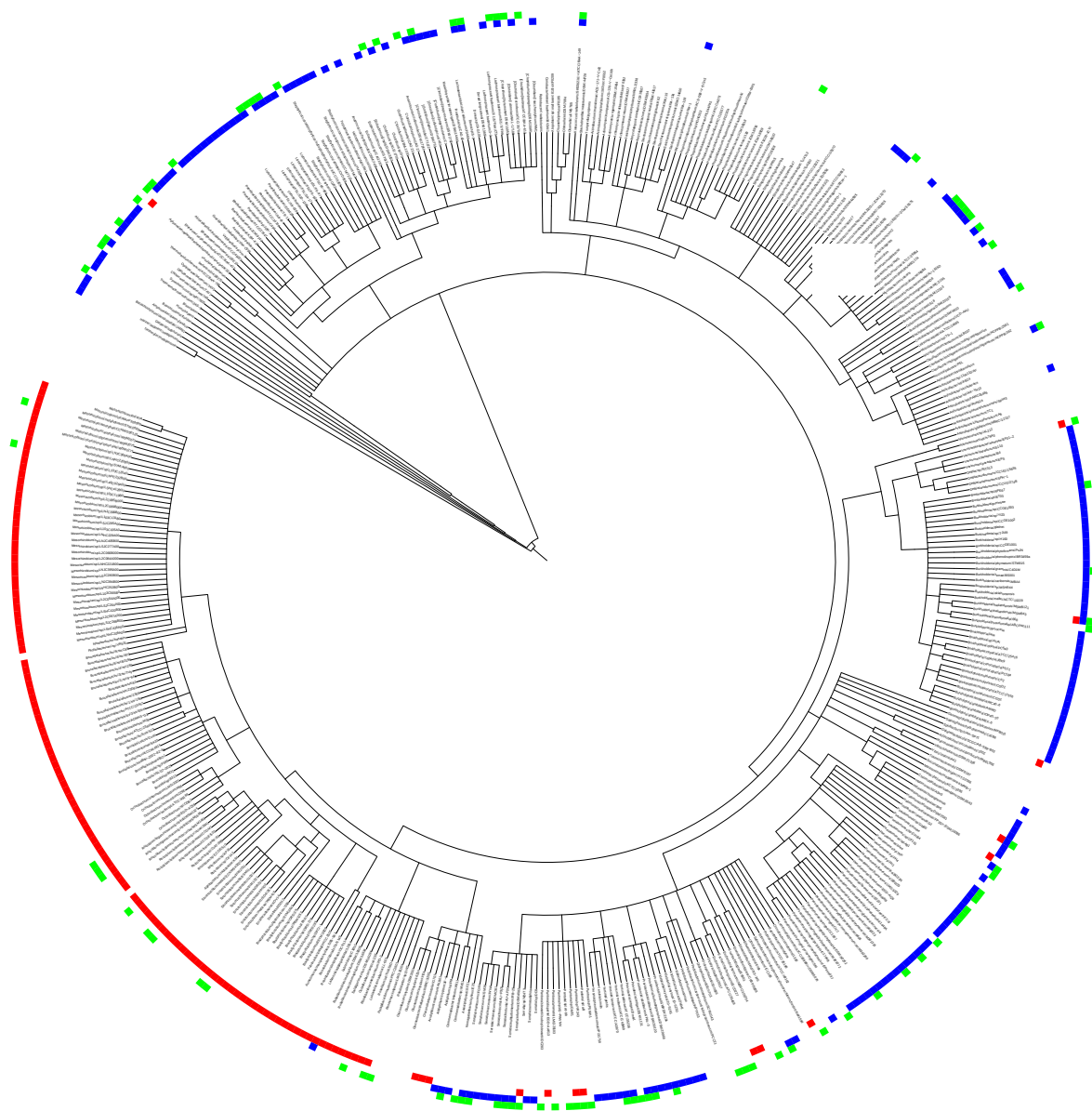
B:

Figure S11: Distribution of tetritol metabolism pathways in different organisms. **Panel A**: Distribution of the *Brucella*-like erythritol metabolic pathway as determined by the presence of erythritol kinase (red boxes) and the *Mycobacterium*-like pathway as determined by the presence of erythritol dehydrogenase (EltD, blue boxes). Distribution of the D-threitol metabolic pathway as determined by D-threitol dehydrogenase (DthD, green boxes). **Panel B**: Distribution of the L-threitol metabolic pathway as determined by the co-occurrence of L-threitol dehydrogenase (EltD, blue boxes) and L-erythrulose-4P epimerase (LerI, pink boxes).

References

1. (a) Zhao, S.; Sakai, A.; Zhang, X.; Vetting, M. W.; Kumar, R.; Hillerich, B.; San Francisco, B.; Solbiati, J.; Steves, A.; Brown, S.; Akiva, E.; Barber, A.; Seidel, R. D.; Babbitt, P. C.; Almo, S. C.; Gerlt, J. A.; Jacobson, M. P., *eLife* **2014**, DOI: 10.7554/eLife.03275; (b) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L., Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2015,** *1854* (8), 1019-1037.

2. (a) Aslanidis, C.; de Jong, P. J.; Schmitz, G., Minimal length requirement of the single-stranded tails for ligation-independent cloning (LIC) of PCR products. *Genome Research* **1994,** *4* (3), 172-177; (b) Aslanidis, C.; de Jong, P. J.; Schmitz, G., Minimal length requirement of the single-stranded tails for ligation-independent cloning (LIC) of PCR products. *PCR Methods Appl* **1994,** *4* (3), 172-7.

3. (a) Vetting, M. W.; Al-Obaidi, N.; Zhao, S.; San Francisco, B.; Kim, J.; Wichelecki, D. J.; Bouvier, J. T.; Solbiati, J. O.; Vu, H.; Zhang, X.; Rodionov, D. A.; Love, J. D.; Hillerich, B. S.; Seidel, R. D.; Quinn, R. J.; Osterman, A. L.; Cronan, J. E.; Jacobson, M. P.; Gerlt, J. A.; Almo, S. C., Experimental Strategies for Functional Annotation and Metabolism Discovery: Targeted Screening of Solute Binding Proteins and Unbiased Panning of Metabolomes. *Biochemistry-US* **2015,** *54* (3), 909-931; (b) Vetting, M. W.; Al-Obaidi, N.; Zhao, S.; San Francisco, B.; Kim, J.; Wichelecki, D. J.; Bouvier, J. T.; Solbiati, J. O.; Vu, H.; Zhang, X.; Rodionov, D. A.; Love, J. D.; Hillerich, B. S.; Seidel, R. D.; Quinn, R. J.; Osterman, A. L.; Cronan, J. E.; Jacobson, M. P.; Gerlt, J. A.; Almo, S. C., Experimental strategies for functional annotation and metabolism

discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* **2015,** *54* (3), 909-31.

4.   (a) Dyrløv Bendtsen, J.; Nielsen, H.; von Heijne, G.; Brunak, S., Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology* **2004,** *340* (4), 783-795; (b) Bendtsen, J. D.; Nielsen, H.; von Heijne, G.; Brunak, S., Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **2004,** *340* (4), 783-95.

5.   (a) Savitsky, P.; Bray, J.; Cooper, C. D.; Marsden, B. D.; Mahajan, P.; Burgess-Brown, N. A.; Gileadi, O., High-throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol* **2010,** *172* (1), 3-13; (b) Savitsky, P.; Bray, J.; Cooper, C. D. O.; Marsden, B. D.; Mahajan, P.; Burgess-Brown, N. A.; Gileadi, O., High-throughput production of human proteins for crystallization: The SGC experience. *Journal of Structural Biology* **2010,** *172* (1), 3-13.

6.   (a) Studier, F. W., Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **2005,** *41* (1), 207-234; (b) Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif* **2005,** *41* (1), 207-34.

7.   Blommel, P. G.; Fox, B. G., A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr Purif* **2007,** *55* (1), 53-68.

8.   (a) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M., HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. *Acta Crystallographica Section D* **2006,** *62* (8), 859-866; (b) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M., HKL-3000: the integration of data reduction

and structure solution--from diffraction images to an initial model in minutes. *Acta Cryst.* **2006,** *62* (Pt 8), 859-66.

9.	(a) Sheldrick, G. M., A short history of SHELX. *Acta Crystallogr A* **2008,** *64* (Pt 1), 112-22; (b) Sheldrick, G., A short history of SHELX. *Acta Crystallographica Section A* **2008,** *64* (1), 112-122.

10.	(a) Langer, G.; Cohen, S. X.; Lamzin, V. S.; Perrakis, A., Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **2008,** *3* (7), 1171-9; (b) Langer, G. G.; Cohen, S. X.; Lamzin, V. S.; Perrakis, A., Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature protocols* **2008,** *3* (7), 1171-1179.

11.	(a) Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **2004,** *60* (12 Part 1), 2126-2132; (b) Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Cryst.* **2004,** *D60* (Pt 12 Pt 1), 2126-32.

12.	(a) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* **1997,** *53* (3), 240-255; (b) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **1997,** *53* (Pt 3), 240-55.

13.	Chacon, O.; Feng, Z.; Harris, N. B.; Cáceres, N. E.; Adams, L. G.; Barletta, R. G., Mycobacterium smegmatis d-Alanine Racemase Mutants Are Not Dependent on d-Alanine for Growth. *Antimicrob. Agents Ch.* **2002,** *46* (1), 47-54.

14.	Kendall, S. L.; Withers, M.; Soffair, C. N.; Moreland, N. J.; Gurcha, S.; Sidders, B.; Frita, R.; Ten Bokum, A.; Besra, G. S.; Lott, J. S.; Stoker, N. G., A highly conserved transcriptional

repressor controls a large regulon involved in lipid degradation in Mycobacterium smegmatis and Mycobacterium tuberculosis. *Mol. Microbiol.* **2007,** *65* (3), 684-699.

15.   Simon, R.; Priefer, U.; Puhler, A., A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Nat. Biotechnol.* **1983,** *1* (9), 784-791.

16.   Goude, R.; Parish, T., Electroporation of Mycobacteria. *J. Visual. Exp.* **2008,** (15), 761.

17.   (a) Letunic, I.; Bork, P., Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **2007,** *23* (1), 127-128; (b) Letunic, I.; Bork, P., Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **2011,** *39* (suppl 2), W475-W478.

18.   (a) Siebold, C.; Arnold, I.; Garcia-Alles, L. F.; Baumann, U.; Erni, B., Crystal Structure of the Citrobacter freundii Dihydroxyacetone Kinase Reveals an Eight-stranded α-Helical Barrel ATP-binding Domain. *J. Biol. Chem.* **2003,** *278* (48), 48236-48244; (b) Zurbriggen, A.; Jeckelmann, J.-M.; Christen, S.; Bieniossek, C.; Baumann, U.; Erni, B., X-ray Structures of the Three Lactococcus lactis Dihydroxyacetone Kinase Subunits and of a Transient Intersubunit Complex. *J. Biol. Chem* **2008,** *283* (51), 35789-35796.

19.   (a) Povelainen, M.; Eneyskaya, E. V.; Kulminskaya, A. A.; Ivanen, D. R.; Kalkkinen, N.; Neustroev, K. N.; Miasnikov, A. N., Biochemical and genetic characterization of a novel enzyme of pentitol metabolism: D-arabitol-phosphate dehydrogenase. *Biochem. J.* **2003,** *371* (Pt 1), 191-197; (b) Fukui, T.; Atomi, H.; Kanai, T.; Matsumi, R.; Fujiwara, S.; Imanaka, T., Complete genome sequence of the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1 and comparison with Pyrococcus genomes. *Genome Res.* **2005,** *15* (3), 352-363.

20. (a) Araujo-Montoya, B. O.; Rofatto, H. K.; Tararam, C. A.; Farias, L. P.; Oliveira, K. C.; Verjovski-Almeida, S.; Wilson, R. A.; Leite, L. C. C., Schistosoma mansoni: Molecular characterization of Alkaline Phosphatase and expression patterns across life cycle stages. *Exp. Parasitol.* **2011,** *129* (3), 284-291; (b) Lambeir, A.-M.; Opperdoes, F. R.; Wierenga, R. K., Kinetic properties of triose-phosphate isomerase from Trypanosoma brucei brucei. *Eur. J. Biochem.* **1987,** *168* (1), 69-74; (c) Lolis, E.; Petsko, G. A., Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5-.ANG. resolution: implications for catalysis. *Biochemistry* **1990,** *29* (28), 6619-6625.

21. Barbier, T.; Collard, F.; Zúñiga-Ripa, A.; Moriyón, I.; Godard, T.; Becker, J.; Wittmann, C.; Van Schaftingen, E.; Letesson, J.-J., Erythritol feeds the pentose phosphate pathway via three new isomerases leading to D-erythrose-4-phosphate in Brucella. *Proc. Nat. Acad. Sci.* **2014,** *111* (50), 17815-17820.