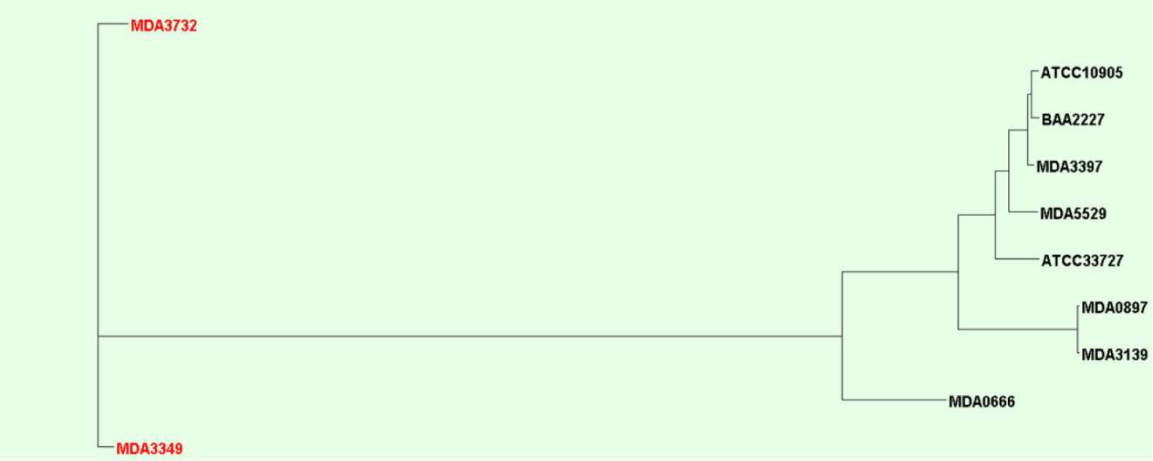

Figure S1. Phylogenetic relationship between the 10 *Nocardia* strains used in this study. A whole genome SNP-based maximum parsimony phylogenetic tree depicts the relationship of these strains based on the variations in the SNPs in their pan-genome. Horizontal branch lengths are expressed in terms of changes per number of SNPs. Strains in red belong to *N. cyriacigeorgica* and the rest belong to *N. nova*. It should be noted that this is a SNP based tree, not an alignment based tree and no evolutionary direction should be inferred from this tree. The tree is visualized using Dendroscope (1).

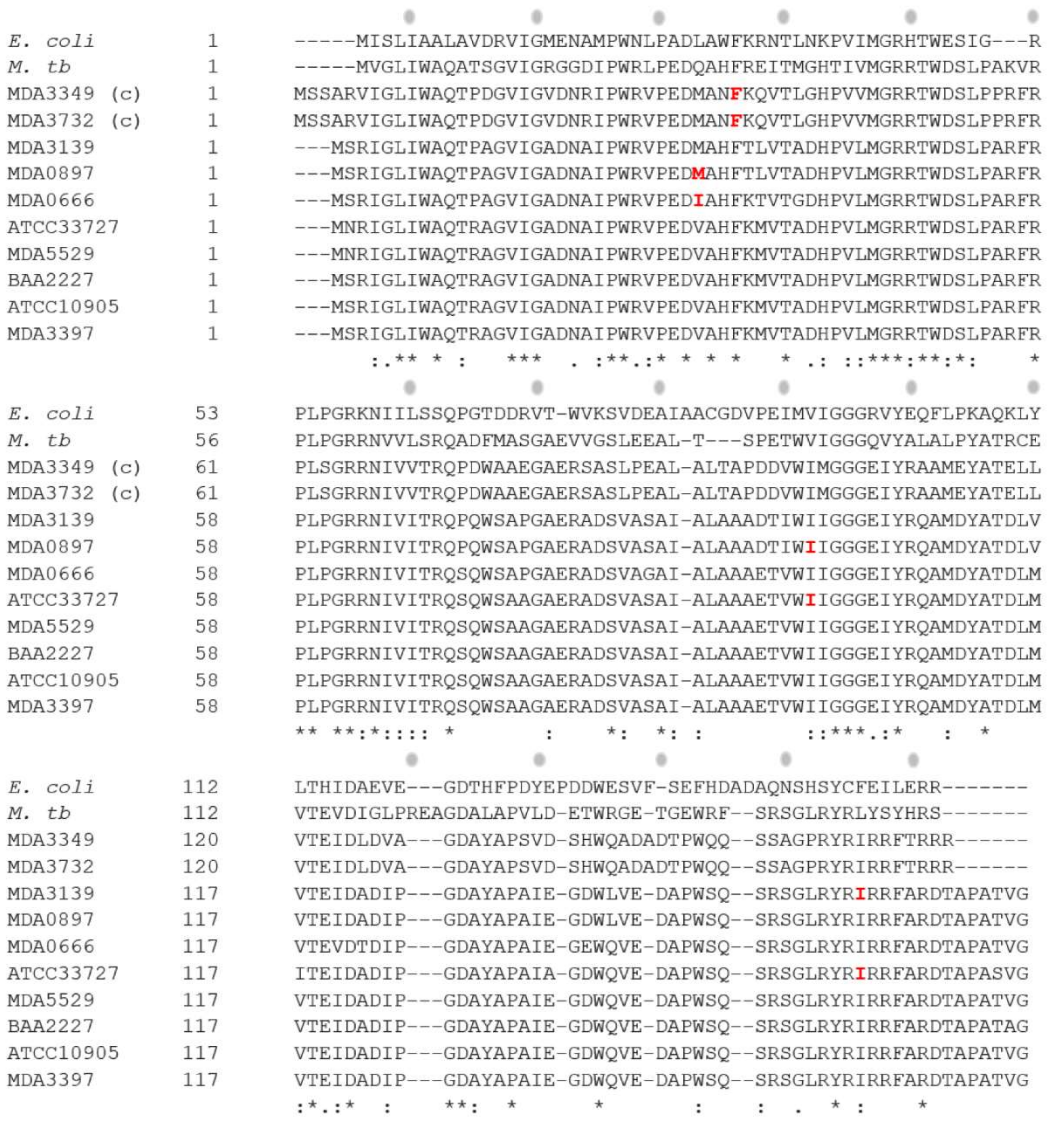

Figure S2. Multiple sequence alignment of DHFR (FolA). Protein sequences of DHFR from *E. coli, M. tuberculosis (M. tb)* and 9 pathogenic *Nocardia* isolates used in this study were aligned to identify conserved residues. An asterisk indicates a conserved residue, a colon indicates amino acids with strongly similar properties, a period indicates amino acids with weakly similar properties. Mutations identified in TMP-SMX evolved *Nocardia* isolates are highlighted in red.

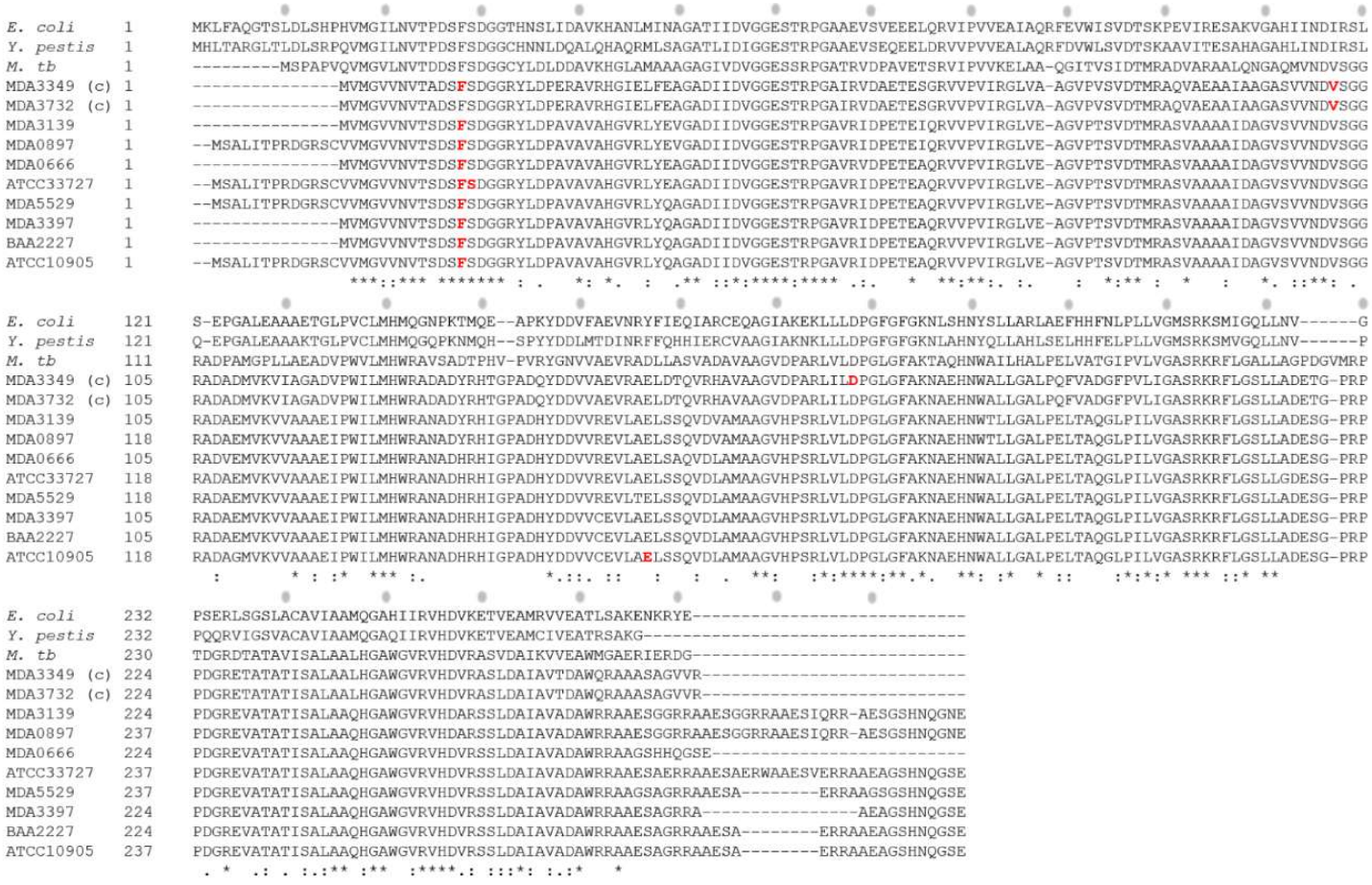

Figure S3. Multiple sequence alignment of DHPS (FolP). Protein sequences of DHPS from *E. coli, Y. pestis, M. tuberculosis (M. tb)* and 9 pathogenic *Nocardia* isolates used in this study were aligned to identify conserved residues. An asterisk indicates a conserved residue, a colon indicates amino acids with strongly similar properties, a period indicates amino acids with weakly similar properties. Mutations identified in TMP-SMX evolved *Nocardia* isolates are highlighted in red.

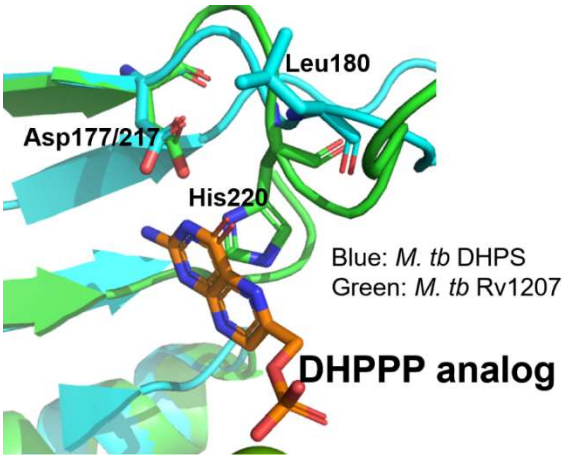

Figure S4. Superimposition of DHPS (FolP1) from *M. tb* (PDB:1eye) and Rv1207 (PDB: 2vp8). The Leu180 residue in the substrate (DHPPP analog) binding pocket of the DHPS structure (blue) is replaced by His220 in Rv1207 (green). It was concluded that this arrangement prevents binding of the pterin substrate to Rv1207 which could be a reason for the lack of DHPS enzymatic activity observed in Rv1207. The image was generated using PyMol (The PyMOL Molecular Graphics System, Version 2.0.1 Schrödinger, LLC).

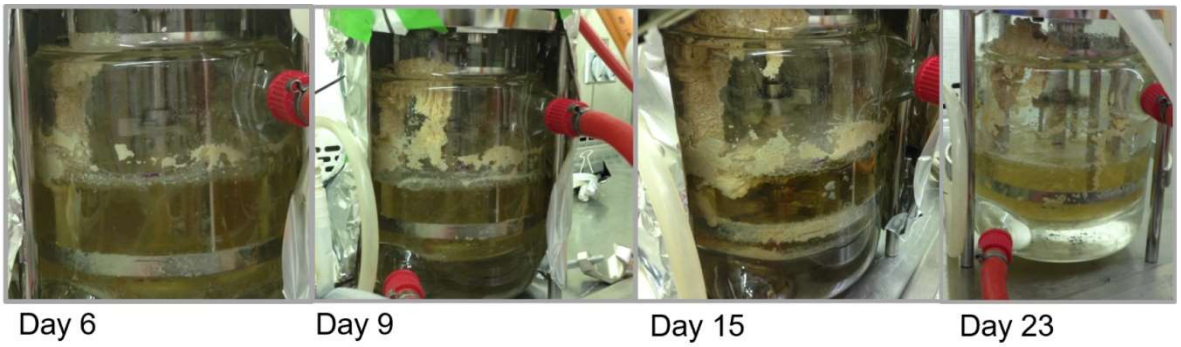

Figure S5. Biofilm formation during growth and adaptation of *N. cyriacigeorgica* MDA3349 to TMP-SMX in a bioreactor. The figure shows the bioreactor vessel that was used to maintain the continuous culture for evolution of MDA3349 to TMP-SMX. It can be observed that thick biofilm was seen coating the metal framework of the vessel as early as day 6 post inoculation which spread on to the vessel walls as the experiment progressed. Shown here are images from

days 6, 9, 15 and 23 post inoculation. The culture was first exposed to TMP-SMX on day 3 post inoculation.

Supplementary file S1

Phylogenetic analysis shows substantial genome diversity for the 10 *Nocardia* strains under study.

Two species of *Nocardia, N. nova* and *N. cyriacigeorgica* were used in this work. The rationale for the use of these species was based on previous works that studied the emergence of TMP-SMX resistance in *Nocardia.* Resistant strains of *N. cyriacigeorgica* were identified in a study involving detection of TMP-SMX resistant nocardial infections in patients receiving TMP-SMX prophylaxis (2). In a different study conducted by the Centers for Disease Control and Prevention (CDC) from 1995 to 2004, 42% of 765 *Nocardia* isolates studied were found to be TMP-SMX resistant out of which 112 belonged to *N. nova* (3). For this study, TMP-SMX susceptible *N. nova* and *N. cyriacigeorgica* strains were obtained from American Type Culture Collection (ATCC) and from cancer patients at the M. D. Anderson Cancer Center, Houston, Texas to study their ability to acquire TMP-SMX resistance and the genetic basis of this resistance.

The 10 different pathogenic strains of *Nocardia* (Table S1) used here had not been studied extensively prior to this work. This, together with the lack of genome sequence available on any of these isolates made it important to sequence the genomes of the strains and to establish their phylogenetic relationship.

As shown in Figure S1, the presence of a very long internal branch in the phylogenetic tree suggests that the two species, *N. nova* and *N. cyriacigeorgica* are sufficiently different from one another. While the two *N. cyriacigeorgica* strains shared similar genome sizes and a tight

phylogeny, the 8 *N. nova* strains were more diverse in genome sizes as well as phylogeny. The *N. nova* strains exhibited clustering in the clade containing ATCC10905, BAA2227 and MDA3397, as well as in the clade containing MDA0897 and MDA3139, with varying amounts of divergence between the remaining strains. The genome sizes and phylogenetic distance between the two species are consistent with species-level differences. However, the distinct genome sizes of of the 5 *N. nova* strains (ranging from 7 to 8 Mb) isolated at a single location (M.D. Anderson Cancer Center, Houston TX) is interesting. Since differences in genomes could affect adaptation to TMP-SMX, we surveyed all 10 strains to provide a broader representation of accessible strategies.

During submission of genomes to NCBI for curation, the Average Nucleotide Identity (ANI) test conducted by NCBI showed that the genome sequence of the strain MDA0666 shared 97.25% identity with *Nocardia aobensis* while ATCC 33727 shared 99.99% identity with *Nocardia elegans* NBRC 108235. MDA0666 was characterized as *N. nova* based on the sequence of the 16s rRNA gene (4) while ATCC 33727 was classified as *N. nova* based on biochemical characteristics of the strain (5). With improvement in technology for taxonomic classification of organisms, the classification of these strains may be modified. However, without conducting more rigorous experiments to confirm this, in this study we will retain the original taxonomic classification of MDA0666 and ATCC 33727 which is *Nocardia nova.*

Supplementary file S2.

Additional information regarding Material and Methods

Experimental evolution by serial flask transfer

In duplicate experiments, 9 strains were grown by picking a single colony from a non-selective

blood agar plate and resuspended in 10 ml LB broth. After initial growth under non-selective conditions, a 1:100 dilution was made into 2 tubes, one containing the current drug concentration (no drug in this case; serves as a control) and one at a concentration equal to 0.25X the starting MIC of the strain. When the culture at the higher concentration reached a density equal to the control tube, 1% of it was used to inoculate 4 new tubes, one containing the same concentration (control) and 3 with higher drug concentrations (typically 1.5, 2 and 2.5X the current drug concentration). This process was repeated until the population reached at least clinical levels of resistance (MIC>4-76 mg/ml TMP-SMX). Before each dilution and increment in drug concentration, 1ml of the population was frozen as a 20% (v/v) glycerol stock at -80ºC. Out of the 18 evolving populations, 1 population of MDA3397 was lost to contamination. Each final resistant population was serially diluted and spread on a non-selective blood agar plate. 2 to 4 end point isolates were picked from each population for phenotypic characterization and whole genome sequencing.

Bioreactor set up

A 600 ml culture volume was used to grow this relatively slow growing organism, *N. cyriacigeorgica* MDA3349 at a steady rate. LB broth was the growth medium. The bioreactor vessel was maintained at 37ºC with 0.2 lpm air flow and a rotor speed of 50 rpm. A 10 ml culture in LB broth was started from a single colony picked from a blood agar plate. After growth for 48 hours, 6 ml of this culture was used to inoculate the vessel. After 24 hours of growth in the bioreactor, the culture reached mid-exponential phase at which point, a pump steadily feeding media into the vessel was turned on at 3% speed. Unlike organisms previously grown in this bioreactor (6–8) that had high metabolic activity and produced CO2 which could be used as a proxy for turbidity, MDA3349 grew much slower and despite the large culture volume, the level

of CO2 in the vessel did not rise significantly or in proportion to the growth of the cells. Thus, the very slow growth rate of nocardia favored a manual adjustment of the media feed to maintain the fastest growth rate. Also, pilot experiments done to observe growth of this organism in the bioreactor suggested that a steady drip of media at 3% pump speed was enough to maintain cells in the desired exponential growth phase.

A 50 ml sample from the bioreactor was collected every 24 hours. 10 ml of this was used to make a 20% glycerol (v/v) stock and frozen at -80ºC. The remaining culture was pelleted and split into 3 tubes which were frozen for future genomic DNA preparation. Each sample was streaked on a blood agar plate and incubated at 37ºC for 48 hours to check for contaminants.

The MIC of the ancestor strain was determined to be 1-19 mg/l TMP-SMX. The adaptation of MDA3349 to TMP-SMX was initiated by exposing the cells growing in the bioreactor to 0.25X MIC drug concentration on day 3 of growth. The cells were allowed to grow for 24 hours in the vessel. The sample collected after 24 hours was subjected to an MIC test in 5 ml cultures set up in glass tubes. The highest drug concentration at which cells grew as well as they did at the previous concentration was determined to be the next drug concentration to be added to the vessel. Duplicate adaptation runs were conducted in the bioreactor, one lasting 22 days and the other one lasting 23 days till the cells were able to grow at 8-152 mg/l TMP-SMX which is 2 fold higher than the clinical breakpoint for this organism (9).

At the end of adaptation, the planktonic culture from the vessel was stored at -80ºC as a glycerol stock. Biofilm accumulated on the glass wall and metal surfaces on the vessel was scraped using a sterile spatula and frozen in 20% glycerol (v/v). The planktonic and biofilm cultures were serially diluted and spread on blood agar plates to isolate individual colonies which were called end point isolates.

MIC test to characterize end point isolates

1 ml culture of each isolate was grown in LB broth in a 96 well deep block plate. After 48 hours of growth, 200 µl of each culture was transferred to a sterile 96 well plate and with the help of a Boekel 96 pin microplate replicator, was spotted on Mueller Hinton agar plates ranging from 0.25-4.75 to 32-608 mg/l TMP-SMX. The estimated MIC of the isolate was the lowest drug concentration at which no visible growth was observed after 48 hours of incubation at 37ºC. Based on the phenotypes and the estimated MICs, 15 isolates from run 1 and 17 from run 2 were selected for whole genome sequencing.

End point isolates from flask transfer and bioreactor adapted populations selected for whole genome sequencing were subjected to agar dilution MIC tests in triplicate. For each isolate, 3 single colonies from a blood agar plate were selected. Individual 1 ml LB liquid cultures in deep block plates were set up for each replicate. After growth for 48 hours, 500 µl of the culture was placed in a sterile microfuge tube and cell clumps were allowed to settle for 15 to 20 minutes. 200 µl of the clear supernatant was transferred to a sterile 96 well plate and its OD at 600 nm was measured on the Synergy 2 (Biotek) Gen 5 plate reader. All ODs were normalized to 0.05 by diluting the culture in Mueller Hinton broth. After dilution, 5 µl of each was spotted onto Mueller Hinton Agar plates ranging in concentration from 0.25-4.75 to 32-608 mg/l TMP-SMX. Plates were incubated for 48 hours at 37ºC. MIC of the isolate was the lowest drug concentration at which no visible growth was observed in each of the three replicates.

Genomic DNA isolation, whole genome sequencing and analysis

MO BIO Ultraclean Microbial DNA Isolation Kit (now called Qiagen DNeasy Ultraclean Microbial kit) was used with an additional lysis step for Gram positive *Nocardia.* 5 µl of 5 kU/ml

mutanolysis and 50 µl of 100 mg/ml lysozyme were added to 300 µl of pelleted cells and incubated at 37ºC for 1 hr before the cells were mechanically lysed. The rest of the protocol was followed as per manufacturer's instructions. Isolated genomic DNA was sent to Genewiz (South Plainfield, NJ) for library preparation and pair ended whole genome sequencing on the Hi-Seq platform. Each metagenomic sample from the daily bioreactor population was sequenced at minimum 300X coverage while the end point isolates were sequenced with at least 100X coverage.

All the strains used in this work did not have complete reference genomes available. Hence, all the ancestor strains were sequenced and genomes were *de novo* assembled. For the bioreactor strain, *N. cyriacigeorgica* MDA3349, the genomic DNA was sent to the Institute of Genome Sciences at the University of Maryland School of Medicine for PacBio sequencing. The long reads obtained from this were *do novo* assembled by the same facility to give a complete closed reference genome of MDA3349. The rest of the *Nocardia* ancestors were sequenced on the Hi-Seq platform and short reads obtained from that were *de novo* assembled into contigs using SPAdes (10). The raw sequencing reads were put through the software with k-mer values of up to 99. All resulting contigs equal to or less than 300 base pairs were deleted.

All *de novo* assembled genomes were submitted to RAST for annotation with default parameters (11). A phylogenetic tree was constructed using kSNP3.0 (12). First, the two utilities MakeFasta and Kchooser, included in the kSNP3.0 package, were run to find an optimal value of k=21 for the k-mer analysis. The default kSNP3 command with k=21 was then run to find the maximum parsimony phylogenetic tree. The computational pipeline Breseq (13) was used for identification of mutations in the evolving populations as well as the end point isolates. RAST annotated genomes were used as reference genomes. Raw sequencing reads from the end point isolates were trimmed using Sickle (14) and aligned to the corresponding reference genome using the consensus

mode on Breseq0.30.0. The trimmed raw reads from metagenomic populations from the bioreactor were run using the polymorphism mode on Breseq0.30.0 with the following parameters: --polymorphism-reject-surrounding-homopolymer-length 5, --polymorphism-reject-indel-homopolymer-length 0 and --polymorphism-minimum-coverage-each-strand 6. A default frequency cut-off of 5% was used for this analysis.

Table S1. Strains used in this work.

Strain	Species	Ancestor MIC (mg/l TMP-SMX)	Source[a]	Sequencing platform	Genome length	Evolution strategy
MDA3349	*N. cyriacigeorgica*	1-19	MDACC	PacBio and Illumina	6,320,090 bp chromosome, 142,547 bp plasmid	Bioreactor
MDA3732	*N. cyriacigeorgica*	1-19	MDACC	Illumina	6,613,593 bp (127 contigs)	Flask transfer
MDA0666	*N. nova*	1-19	MDACC	Illumina	7,040,727 bp (87 contigs)	Flask transfer
MDA0897	*N. nova*	1-19	MDACC	Illumina	7,098,143 bp (167 contigs)	Flask transfer
MDA3139	*N. nova*	1-19	MDACC	Illumina	7,522,631 bp (177 contigs)	Flask transfer
MDA3397	*N. nova*	2-38	MDACC	Illumina	8,035,742 bp (122 contigs)	Flask transfer
MDA5529	*N. nova*	2-38	MDACC	Illumina	7,733,147 bp (118 contigs)	Flask transfer
BAA2227	*N. nova*	2-38	ATCC	Illumina	7,988,778 bp (128 contigs)	Flask transfer
ATCC10905	*N. nova*	2-38	ATCC	Illumina	7,875,623 bp (108 contigs)	Flask transfer
ATCC33727	*N. nova*	2-38	ATCC	Illumina	7,578,591 bp (105 contigs)	Flask transfer

[a] MDACC: M.D. Anderson Cancer Center; ATCC: American Type Culture Collection

References

1. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. 2007. Dendroscope : An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460.
2. Minero MV, Marın M, Cercenado E, Rabadan PM, Bouza E, Munoz P. 2009. Nocardiosis at the turn of the century. Medicine (Baltimore) 88:250–261.
3. Uhde KB, Pathak S, Mccullum Jr I, Jannat-Khah DP, Shadomy S V, Dykewicz CA, Clark TA, Smith TL, Brown JM. 2010. Antimicrobial-resistant *Nocardia* isolates, United States, 1995–2004. Clin Infect Dis 51:1445–1448.
4. Wang HL, Seo YH, LaSala PR, Tarrand JJ, Han XY. 2014. Nocardiosis in 132 patients with cancer: Microbiological and clinical analyses. Am J Clin Pathol 142:513–523.
5. Tsukamura M. 1982. Numerical analysis of the taxonomy of nocardiae and rhodococci. Microbiol Immunol 26:1101–1119.
6. Mehta HH, Prater AG, Shamoo Y. 2018. Using experimental evolution to identify druggable targets that could inhibit the evolution of antimicrobial resistance. J Antibiot (Tokyo) 71:279–286.
7. Hammerstrom TG, Beabout K, Clements TP, Saxer G, Shamoo Y. 2015. *Acinetobacter baumannii* repeatedly evolves a hypermutator phenotype in response to tigecycline that effectively surveys evolutionary trajectories to resistance. PLoS One 10:e0140489.
8. Beabout K, Hammerstrom TG, Wang TT, Bhatty M, Christie PJ, Saxer G, Shamoo Y. 2015. Rampant parasexuality evolves in a hospital pathogen during antibiotic selection. Mol Biol Evol 32:2585–97.

9. National Committee for Clinical Laboratory Standards. 2011. Susceptibility Testing of mycobacteria, nocardiae, and other aerobic actinomycetes. Approved Standard M24-A2, 2nded. Clinical Laboratory Standards Institute [CLSI], Wayne, PA.
10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.
11. Aziz RK, Bartels D, Best AA, Dejongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, Mcneil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server : Rapid Annotations using Subsystems Technology. BMC Genomics 9:75.
12. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0 : SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31:2877–2878.
13. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165–188.
14. Joshi NA, Fass JN. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.