Network approach to mutagenesis sheds insight on phage resistance in mycobacteria

Saptarshi Sinha[1], Sourabh Samaddar[2], Sujoy K Das Gupta[2], and Soumen Roy[1*]

[1]Department of Physics, Bose Institute, 93/1 Acharya Prafulla Chandra Road, Kolkata 700009 WB India
[2]Department of Microbiology, Bose Institute, P-1/12 CIT Road, Scheme VIIM, Kolkata 700 054 WB India
[*]soumen@jcbose.ac.in

Abstract

Version of the manuscript accepted to Bioinformatics (Oxford). The final version is available on the publisher's website at https://doi.org/10.1093/bioinformatics/btaa1103

Motivation: A rigorous yet general mathematical approach to mutagenesis, especially one capable of delivering systems-level perspectives would be invaluable. Such systems-level understanding of phage resistance is also highly desirable for phage-bacteria interactions and phage therapy research. Independently, the ability to distinguish between two graphs with a set of common or identical nodes and identify the implications thereof, is important in network science.

Results: Herein we propose a measure called shortest path alteration fraction (SPAF) to compare any two networks by shortest paths, using sets. When SPAF is one, it can identify node pairs connected by at least one shortest path, which are present in either network but not both. Similarly, SPAF equaling zero identifies identical shortest paths,which are simultaneously present between a node pair in both networks. We study the utility of our measure theoretically in five diverse microbial species, to capture reported effects of well-studied mutations and predict new ones. We also scrutinise the effectiveness of our procedure through theoretical and experimental tests on *Mycobacterium smegmatis* mc^2155 and by generating a mutant of mc^2155, which is resistant to mycobacteriophage D29. This mutant of mc^2155, which is resistant to D29 exhibits significant phenotypic alterations. Whole-genome sequencing identifies mutations, which cannot readily explain the observed phenotypes. Exhaustive analyses of protein-protein interaction network of the mutant and wild-type, using the machinery of topological metrics and differential networks does not yield a clear picture. However, SPAF coherently identifies pairs of proteins at the end of a subset of shortest paths, from amongst hundreds of thousands of viable shortest paths in the networks.The altered functions associated with the protein pairs are strongly correlated with the observed phenotypes.

1 Introduction

Random mutations are one of the fundamental drivers of evolution. Mutations can sometimes lead to non-trivial genotype-phenotype relationships. A proper understanding of this relationship is of deep significance across all disciplines of life science. A rigorous yet broad mathematical formulation of mutagenesis is definitely desirable. Towards this end, herein, we introduce a general and potent measure called the Shortest Path Alteration Fraction (SPAF). It enables us to distinguish between any two networks with a set of common or identical nodes — irrespective of their origin, function, or kind. We demonstrate the importance of our method across five diverse microbial species. It successfully captures the reported effects of all given mutations. Indeed, it also predicts previously unknown phenotypic consequences of the mutations. Independently, we generate a D29 resistant mutant of *Mycobacterium smegmatis* mc^2155. Using results from extensive experiments reported herein, we also show how SPAF can shed new insight into phage resistance.

The study of mycobacteriophage resistance is significant in its own right. The emergence of multidrug-resistant (MDR) strains in microbes presents an unprecedented crisis in human civilisation. Alternatives to antibiotics are needed most urgently. MDR strains have also been reported in several mycobacteria, including in dreaded pathogens like *Mycobacterium leprae* and *Mycobacterium tuberculosis* (MTB) [1–3]. It is especially in this context that the bactericidal activity of phages, could play a vital role. Phages are associated with target specificity, minimal side effects and reproducibility. These properties poise phages to be a viable and indeed better potential alternative compared to most antibiotics. Phage therapy [4–6] is now a subject of intense scientific and clinical research [7–11].

A prime obstacle, however, is that the host bacteria adopt various robust resistance mechanisms against their phages to survive in the evolutionary run. Examples of such mechanisms are: (a) the inhibition of phage adsorption by alteration of the host cell sur-

face [12], (b) restriction-modification (R-M) systems [13], and, (c) anti-phage endonuclease systems like CRISPR (Clustered Regularly Interspersed Short Palindromic Repeats)-Cas to modify the genetic material of phages [14, 15]. A thorough systems-level understanding of phage resistance is not merely crucial for understanding phage-bacteria interaction, dynamics, and coevolution but also possesses immense therapeutic value.

Specifically for mycobacteria-mycobacteriophage system [16–18], the associated biological complexity hinders a proper understanding of these resistance mechanisms. Some reports indicate that alterations in certain genes and modification of cell wall components impart resistance to mycobacteriophages in some mycobacterial strains [3, 19, 20]. Examples of well-known mycobacteriophages are D29, L1, L5, Omega, and TM4. Among them, mycobacteriophage D29 is particularly well studied. It is a lytic phage and has high genomic similarity with mycobacteriophage L5 [21].

Here we use *Mycobacterium smegmatis* mc^2155 (henceforth mostly referred to as wild type or WT), as the host organism and D29 as the phage. Very few D29 phage resistant mutants of *Mycobacterium smegmatis* (MSG) are known in literature [20]. MSG is important because it is a close relative of the dangerous pathogen *Mycobacterium tuberculosis* (MTB). To understand the resistance mechanisms of *Mycobacterium smegmatis* towards D29, we have used random mutagenesis to generate a mutant of the wild type [22], which is resistant to D29. This D29 mycobacteriophage resistant mutant of mc^2155 will henceforth be frequently referred to as PRM.

Extensive characterisation of PRM shows non-trivial phenotypic differences with respect to the WT as regards: (a) resistance to D29 phage, (b) average cell length, (c) isoniazid sensitivity, and, (d) surface morphology. The vast majority of the identified mutations do not seem to show a direct relationship with PRM phenotypes, *prima facie*. Therefore, further insight is needed into how variations in the PRM genome as compared to the WT genome, lead to observed changes in phenotype. For this we analyse protein-protein interaction network (PPIN) of WT and PRM by mechanisms detailed in Section 2. Topological analyses of networks have been extensively applied to biology and medicine [23–25]. Further, the formalism of differential networks [26, 27] has been successfully used to provide insight into the structural determinants of optogenetics [28]. Use of differential networks coupled with extensive topological analyses results in the identification of a few proteins, which are affected due to these mutations. However, we do not obtain a clear picture. Therefore, we apply SPAF for detecting the effects of mutagenesis. SPAF can also be readily applied to many other biological problems.

SPAF identifies pairs of proteins located at the ends of select shortest paths, from hundreds of thousands of viable shortest paths. The identified node pairs are important because every pair is connected by at least one shortest path, which is present in either network but not both. These select protein pairs pinpoint specific biological functions that have been revised due to mutations. We find that the function of the altered pairs is strongly correlated with the observed phenotypes.

The method presented herein is most general in that it allows the comparison of any two networks possessing common or identical nodes by shortest paths irrespective of their origin, function, or kind. Therefore, in addition to the mutations in microbes demonstrated herein by us, it can be naturally applied to other classes of biological networks with common or identical nodes like ecological networks and protein contact networks [29].

2 Material and methods

2.1 A discrete mathematical approach to mutagenesis: Shortest path alteration fraction

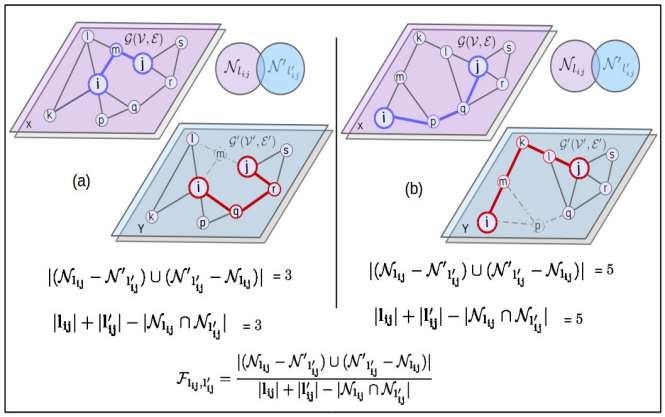

Figure 1: How does the mutation or deletion of even a single node in a network affect the shortest paths between all other nodes? Mutation of a node can lead to alterations in the *set of all possible shortest paths*, $\mathcal{L}_{ij}$ and $\mathcal{L}'_{ij}$, between nodes i and j in networks $\mathcal{G}(\mathcal{V}, \mathcal{E})$ and $\mathcal{G}'(\mathcal{V}', \mathcal{E}')$ respectively. Here, the shortest paths before and after mutation of $'m'$ in (a) and $'p'$ in (b) are denoted in blue and red respectively. The shortest path alteration fraction, $\mathcal{F}_{l_{ij},l'_{ij}}$, compares shortest paths $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$. $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ identifies shortest paths, which are present in either $\mathcal{G}(\mathcal{V}, \mathcal{E})$ or $\mathcal{G}'(\mathcal{V}', \mathcal{E}')$, but not in both.

We present rigorous results, which will allow comparison between any two networks with common or identical nodes, using shortest paths. This section is solely concerned with rigorous mathematical details regarding the shortest path alteration fraction (SPAF). Therefore, all subsequent sections(Section 2.2 onwards), which are mostly concerned with biological details, can be read independent of this section. An illustration of the notion

of SPAF and the essence of the most important results of this section is provided in Fig. 1. The applications of SPAF and its biological importance across microbial species is presented from Section 3 onwards.

Definition: Throughout this paper, the term graph is meant to be synonymous with a network. Let $\mathcal{G}(\mathcal{V},\mathcal{E})$ and $\mathcal{G}'(\mathcal{V}',\mathcal{E}')$ denote two undirected graphs; where, $\mathcal{V} \stackrel{?}{=} \mathcal{V}'$ and $\mathcal{E} \stackrel{?}{=} \mathcal{E}'$ represent the set of nodes and edges respectively.

Hereinafter, we restrict ourselves to $\mathcal{G}_{\mathcal{L}}(\mathcal{V}_{\mathcal{L}},\mathcal{E}_{\mathcal{L}})$ and $\mathcal{G}'_{\mathcal{L}'}(\mathcal{V}'_{\mathcal{L}'},\mathcal{E}'_{\mathcal{L}'})$, which represent the largest connected subgraph of $\mathcal{G}(\mathcal{V},\mathcal{E})$ and $\mathcal{G}'(\mathcal{V}',\mathcal{E}')$ respectively.

Given two nodes $i,j \in \mathcal{V}_{\mathcal{L}} \cap \mathcal{V}'_{\mathcal{L}'}$ *which are not connected by an edge* — one can further identify the *set of shortest paths*: (a) $\mathcal{L}_{ij}$ between $i,j \in \mathcal{V}_{\mathcal{L}}$, and, (b) $\mathcal{L}'_{ij}$ between $i,j \in \mathcal{V}'_{\mathcal{L}'}$. To reiterate; $i,j \in \mathcal{V}_{\mathcal{L}} \cap \mathcal{V}'_{\mathcal{L}'}$ for both (a) and (b). Furthermore, let $\mathcal{N}_{l_{ij}}$ and $\mathcal{N}'_{l'_{ij}}$ denote the *set of nodes* lying on the shortest paths $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$ respectively. $|l_{ij}| = (|\mathcal{N}_{l_{ij}}| - 1)$ and $|l'_{ij}| = (|\mathcal{N}'_{l'_{ij}}| - 1)$ denote the length of shortest paths l_{ij} and l'_{ij} respectively. Herein, we introduce the following measure, the shortest path alteration fraction between $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$,

$$\mathcal{F}_{l_{ij},l'_{ij}} = \frac{|(\mathcal{N}_{l_{ij}} - \mathcal{N}'_{l'_{ij}}) \cup (\mathcal{N}'_{l'_{ij}} - \mathcal{N}_{l_{ij}})|}{|l_{ij}| + |l'_{ij}| - |\mathcal{N}_{l_{ij}} \cap \mathcal{N}_{l'_{ij}}|} \tag{1}$$

Comments: $0 \leq \mathcal{F}_{l_{ij},l'_{ij}} \leq 1$. Furthermore:

(**1**) The trivial case of $\mathcal{F}_{l_{ij},l'_{ij}} = 0, \forall\{i,j\}$ is realised when $\mathcal{V}_{\mathcal{L}} = \mathcal{V}'_{\mathcal{L}'}$ and $\mathcal{E}_{\mathcal{L}} = \mathcal{E}'_{\mathcal{L}'}$. If we compare $\mathcal{G}(\mathcal{V},\mathcal{E})$ with $\mathcal{G}'(\mathcal{V}',\mathcal{E}')$ instead of merely comparing $\mathcal{G}_{\mathcal{L}}(\mathcal{V}_{\mathcal{L}},\mathcal{E}_{\mathcal{L}})$ with $\mathcal{G}'_{\mathcal{L}'}(\mathcal{V}'_{\mathcal{L}'},\mathcal{E}'_{\mathcal{L}'})$; $\mathcal{F}_{l_{ij},l'_{ij}} = 0, \forall\{i,j\}$ would obviously be realised when $\mathcal{V} = \mathcal{V}'$ and $\mathcal{E} = \mathcal{E}'$.

(**2**) If $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ and $\mathcal{L}_{i,j} - \mathcal{L}'_{i,j} = \varnothing$, then $\mathcal{L}'_{i,j} - \mathcal{L}_{i,j}$ contains only those shortest paths which are present in $\mathcal{L}'_{ij}$ and $\mathcal{G}'(\mathcal{V}',\mathcal{E}')$ but not in $\mathcal{L}_{ij}$ or $\mathcal{G}(\mathcal{V},\mathcal{E})$. Similarly, if $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ and $\mathcal{L}'_{i,j} - \mathcal{L}_{i,j} = \varnothing$, then $\mathcal{L}_{i,j} - \mathcal{L}'_{i,j}$ contains only those shortest paths which are present in $\mathcal{L}_{ij}$ and $\mathcal{G}(\mathcal{V},\mathcal{E})$ but not in $\mathcal{L}'_{ij}$ or $\mathcal{G}'(\mathcal{V}',\mathcal{E}')$.

(**3**) Comparison of two highly different networks is mathematically possible but is usually uninformative besides being computationally expensive. Substantial useful information from interesting real-world examples like mutagenesis could be gained when $|\mathcal{V} - \mathcal{V}'| \approx |\mathcal{V}' - \mathcal{V}|$.

(**4**) We can hardly overemphasise the fact that $\mathcal{F}_{l_{ij},l'_{ij}}$ could be 1 for many pairs of $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$. However, only those $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$, which are *not simultaneously present in both* $\mathcal{G}(\mathcal{V},\mathcal{E})$ and $\mathcal{G}'(\mathcal{V}',\mathcal{E}')$, would yield the most insightful results. Throughout, this paper we use only those $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$ with $\mathcal{F}_{l_{ij},l'_{ij}} = 1$, which are present in *either* $\mathcal{G}(\mathcal{V},\mathcal{E})$ *or* $\mathcal{G}'(\mathcal{V}',\mathcal{E}')$*, but not simultaneously in both.*

Theorem 1 *For a pair of nodes, i and j, which is simultaneously present in two nonidentical graphs but not directly connected by an edge— if shortest paths, $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$ possess: (i) all nodes in common, then $\mathcal{F}_{l_{ij},l'_{ij}} = 0$; (ii) more than two but not all nodes in common, then $0 < \mathcal{F}_{l_{ij},l'_{ij}} < 1$; and (iii) exactly two nodes, i and j, in common, then $\mathcal{F}_{l_{ij},l'_{ij}} = 1$.*

Corollary 1 *If $\mathcal{N}'_{l'_{ij}} \neq \mathcal{N}_{l_{ij}}$ yet $|\mathcal{N}'_{l'_{ij}}| = |\mathcal{N}_{l_{ij}}|$, then generally $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$ can not be described easily in terms of each other when $\mathcal{F}_{l_{ij},l'_{ij}} = 1$.*

Proof: Let us recall that i and j are not connected by an edge as specified in the definition above but rather by shortest paths. If all nodes between $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L}'_{ij}$ are common, then $\mathcal{N}_{l_{ij}} = \mathcal{N}'_{l'_{ij}}$, which trivially leads to $\mathcal{F}_{l_{ij},l'_{ij}} = 0$.

We discuss below, the conditions under which $0 < \mathcal{F}_{l_{ij},l'_{ij}} \leq 1$, when $\mathcal{N}_{l_{ij}} \neq \mathcal{N}'_{l'_{ij}}$. Let us write

$$|\mathcal{N}_{l_{ij}} \cap \mathcal{N}'_{l'_{ij}}| = m \tag{2}$$

We will first restrict ourselves to the scenario when neither $\mathcal{N}_{l_{ij}} \not\subset \mathcal{N}'_{l'_{ij}}$ nor $\mathcal{N}'_{l'_{ij}} \not\subset \mathcal{N}_{l_{ij}}$. Every pair of shortest paths between i and j will always have their end nodes, i and j, in common — even if they share no other node. Thus, $m \geq 2$ between i and j. Therefore, a pair of:

(A) shortest paths possessing no other common node apart from their end nodes leads to $m = 2$.

(B) partially overlapping shortest paths between i and j obviously always has other common nodes apart from the end nodes, leading to $m > 2$.

Using Eqn. 2 in conjunction with a well-known basic result from Set theory on the numerator of Eqn. 1 yields, the following, which is valid for both (A) and (B),

$$|(\mathcal{N}_{l_{ij}} - \mathcal{N}'_{l'_{ij}}) \cup (\mathcal{N}'_{l'_{ij}} - \mathcal{N}_{l_{ij}})| = (|\mathcal{N}_{l_{ij}}| - m) + (|\mathcal{N}'_{l'_{ij}}| - m) \tag{3}$$

As defined above,

$$|l_{ij}| = (|\mathcal{N}_{l_{ij}}| - 1) \tag{4}$$

$$|l'_{ij}| = (|\mathcal{N}'_{l'_{ij}}| - 1) \tag{5}$$

Eqns. 2, 4 and 5 in the denominator of Eqn. 1 yield,

$$|l_{ij}| + |l'_{ij}| - |\mathcal{N}_{l_{ij}} \cap \mathcal{N}_{l'_{ij}}| = (|\mathcal{N}_{l_{ij}}| - 1) + |(|\mathcal{N}'_{l'_{ij}}| - 1) - m \tag{6}$$

Plugging R.H.S. of Eqns. 3 and 6 in the numerator and denominator of Eqn. 1 respectively, leads to

$$\mathcal{F}_{l_{ij},l'_{ij}} = \frac{|\mathcal{N}_{l_{ij}}| + |\mathcal{N}'_{l'_{ij}}| - 2m}{|\mathcal{N}_{l_{ij}}| + |\mathcal{N}'_{l'_{ij}}| - m - 2} \tag{7}$$

We are currently restricted to $\mathcal{N}_{l_{ij}} \neq \mathcal{N}'_{l'_{ij}}$, when neither $\mathcal{N}_{l_{ij}} \not\subset \mathcal{N}'_{l'_{ij}}$ nor $\mathcal{N}'_{l'_{ij}} \not\subset \mathcal{N}_{l_{ij}}$. Therefore, it follows from Eq. 7 that: (I) $m > 2 \implies 0 < \mathcal{F}_{l_{ij},l'_{ij}} < 1$, and, (II) $m = 2 \implies \mathcal{F}_{l_{ij},l'_{ij}} = 1$.

To complete the remaining possibilities for $\mathcal{N}_{l_{ij}} \neq \mathcal{N}'_{l'_{ij}}$, we now consider that either of the following could hold true:

(C) $\mathcal{N}_{l_{ij}} \subset \mathcal{N}'_{l'_{ij}}$, whence $\mathcal{N}_{l_{ij}} - \mathcal{N}'_{l'_{ij}} = \varnothing$ and $\mathcal{N}'_{l'_{ij}} - \mathcal{N}_{l_{ij}} = \mathcal{N}'_{l'_{ij}} - m \implies$ numerator of Eqn. 1 is $(|\mathcal{N}'_{l'_{ij}}| - m)$

(D) $\mathcal{N'}_{l'_{ij}} \subset \mathcal{N}_{l_{ij}}$, whence $\mathcal{N'}_{l'_{ij}} - \mathcal{N}_{l_{ij}} = \varnothing$ and $\mathcal{N}_{l_{ij}} - \mathcal{N'}_{l'_{ij}} = \mathcal{N}_{l_{ij}} - m \implies$ numerator of Eqn. 1 is $(|\mathcal{N}_{l_{ij}}| - m)$

For both (C) and (D) the R.H.S. of Eqn. 6 is still valid as the denominator of Eqn. 1. Additionally, for both (C) and (D), nodes i and j are defined to be connected not by an edge but by shortest path/(s). Since $\mathcal{N'}_{l'_{ij}} \subset \mathcal{N}_{l_{ij}}$ or $\mathcal{N'}_{l'_{ij}} \subset \mathcal{N}_{l_{ij}}$, therefore $m > 2$ always for both (C) and (D), thus leading to $0 < \mathcal{F}_{l_{ij},l'_{ij}} < 1$.
Therefore, Theorem 1 stands proved. We know that $\mathcal{F}_{l_{ij},l'_{ij}} = 0$, when $\mathcal{N'}_{l'_{ij}} = \mathcal{N}_{l_{ij}}$. We have also proved above that $0 < \mathcal{F}_{l_{ij},l'_{ij}} \leq 1$, when $\mathcal{N'}_{l'_{ij}} \neq \mathcal{N}_{l_{ij}}$. What happens when $\mathcal{N'}_{l'_{ij}} \neq \mathcal{N}_{l_{ij}}$ yet $|\mathcal{N'}_{l'_{ij}}| = |\mathcal{N}_{l_{ij}}|$? The number of overlapping nodes between $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L'}_{ij}$ can then either be: (i) two, or, (ii) more than two but not all. For the former $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ and for the latter $\mathcal{F}_{l_{ij},l'_{ij}} < 1$. For both cases, especially the former, $l_{ij} \in \mathcal{L}_{ij}$ and $l'_{ij} \in \mathcal{L'}_{ij}$ can not be generally described easily in terms of one another. Corollary 1 thus follows from Theorem 1.

2.2 Generation and characterization of mycobacteriophage resistant mutant

2.2.1 Bacteria and bacteriophage

We use mycobacteriophage D29 and *Mycobacterium smegmatis* mc^2155 as the host bacterium (WT). The host bacterium is a laboratory strain and the phage was obtained from Dr. Ruth McNerney (London).

2.2.2 Media and other chemicals

For the growth of *Mycobacterium smegmatis* mc^2155, we use Middlebrook 7H9 minimal media (MB7H9) powder (Difco). For preparation of broth, we add $2.5g/l$ (0.25%) BSA (HiMedia Laboratories) as a nitrogen source and $2ml/l$ (0.2%v/v) Glycerol (SRL) as a carbon source with $4.66g/l$ MB7H9. For preparing solid media, we use $5.6g/l$ MB7H9 along with $15g/l$ (1.5%w/v) agar powder (HiMedia). For growing cells in the liquid medium, we add 0.01% Tween 80. However for phage infection, Tween 80 is replaced by $2mM\mathrm{CaCl_2}$.

2.2.3 Random mutagenesis

Dimethyl sulfate (DMS) is used as a mutagen to generate random mutagenesis in the host bacterium. Cells were allowed to grow in the presence of $1mM$ DMS in MB7H9 broth for 2 days at $37°C$. DMS treated cells are incubated with mycobacteriophage D29. Only two tubes among six indicate growth in the presence of phage. After re-incubation in a fresh medium, cells were plated in the presence of phages. A single colony was isolated as phage resistant *Mycobacterium smegmatis* mc^2155 and analysed for phenotypic characterisation.

2.2.4 Comparison of growth and interaction dynamics and adsorption affinity

Both types of cells were grown in MB7H9 broth in the presence of 0.01% Tween 80. For the comparison of growth rate, samples were withdrawn at regular intervals for $O.D.$ measurements at $600nm(OD_{600})$. During interaction dynamics experiments of D29 with WT and PRM cells, we measure the CFU count at regular time intervals. Initial cells densities were chosen so as to maintain MOI 0.1. For the adsorption affinity and comparison of growth in the presence of phage, Tween 80 was replaced by $2mM\mathrm{CaCl_2}$. Purified phage solution with Multiplicity of Infection, MOI 1.0, was added in the bacterial solution when OD_{600} reached 0.1. For adsorption assay, after every 15 min time interval, samples were collected from the same phage-bacteria mixture of MOI 0.1 till 75 min. For counting the adsorbed phages on the cell surface, the samples were centrifuged at 6000 RPM to separate cells and free phages. After a 10^5 fold dilution, the samples were pour plated with host bacteria (as indicator bacteria) in soft agar. The adsorbed phages were the same as PFU count [30].

2.2.5 INH sensitivity assay

Over a wide range of concentrations of INH (10μg/ml to 50μg/ml), both types of cells were incubated overnight in MB7H9 at $37°C$. Following the McFarland standards, the inoculum density of both cell types was kept at 10^5 cell/ml. After overnight incubation, the OD of all sets was measured at $600nm$ for growth measurements.

2.2.6 Scanning Electron Microscopy (SEM)

For Scanning Electron Microscopy (SEM), samples were collected and fixed with glutaraldehyde (1.5%w/v). After centrifugation at 6000 RPM, the pellet was dissolved in 20% ethanol solution and spread on the glass slide for drying. The gold coating was done in vacuum for 3 min on the aluminium stab. Subsequently, the samples were mounted for SEM [17].

2.2.7 Whole genome sequencing

The whole genome of PRM and the WT were sequenced by Illumina Nextseq 500 platform. Quality of the reads was checked by FastQC. Sequence assembly and mapping of mutations were done by CLC bio Genomics Workbench version 9 (CGWB) software. Variant identification was performed using Qiagen's CLC Fixed Ploidy Variant Detection module. This process was executed by Bioserve Biotechnologies. Proteins are identified from 'Mycobrowser' database [31].

3 Results

3.1 Characterisation of mycobacteriophage resistant mutant

3.1.1 Altered interaction dynamics of PRM cells with D29

The results of the growth dynamics of WT and PRM cells in absence of D29 is presented in Fig. $S1$ of Supplementary Information 1 (SI 1). As observable therein, the behaviour is very similar. However, the scenario differs in the presence of D29. The dynamics of the interaction of WT cells with D29, which has a latent period of 60 min has been reported [8]. The D29 particles infect WT cells and the lytic cycle continues for about 4 $hour$. This interaction results in a decrease in Optical Density (OD) of the phage-bacteria mixture by two log folds. However, PRM cells by their very nature, possess phage resistance. Therefore, even after 1 $hour$ PRM cells continue to grow. This ultimately results in an overall increase in OD of the phage-bacteria mixture by one log fold, after 4 $hour$ of incubation as seen in Fig. 2(c). We measure optical density of the growth media at 600 nm (OD_{600}) for the growth dynamics experiments and CFU count for interaction dynamics experiments. Further details are presented in Section 2.2. To probe into the mechanisms of resistance, adsorption assay of both types of cells is investigated. Here we use the standard method outlined earlier to compare D29 adsorption affinity between both types of cells. It is observed that, in comparison with WT cells, PRM cells have a distinct pattern of phage adsorption with a decreased affinity, which is represented in the inset of Fig. 2(c). Unlike WT cells, PRM cells do not have an adsorption maxima after 60 min. This decrease in the phage adsorption affinity of PRM cells could be a manifestation of cell surface modification or other physiological changes [12]. Therefore, to examine the cell surface morphology and physiological properties, we performed microscopy analyses.

3.1.2 Scanning electron microscopy reveals morphological differences

Initial observations with an optical microscope reveal some phenotypic alterations in PRM cells in comparison to WT cells. Significant changes include alteration of cellular aggregation property and average cell length as depicted in Fig. $S2$ and Fig. $S3$ respectively of SI 1. A slightly reduced acceptance of acid fast stain is also observed. In order to investigate those alterations in further detail, we conduct scanning electron microscopy (SEM). The SEM images of WT and PRM cells are shown in Fig. 2(a) and Fig. 2(b) respectively. These images show that the average length of PRM cells is significantly shorter than WT cells ($t\ value = 0.0078$ and $df = 20$). According to measurements, the average length of PRM cells is about three-quarters of the average length of WT cells as seen in Fig. 2(c). The forward scatter measurement in FACS analysis of a cell population lends clues towards the overall average cell size of that population [44]. The size difference revealed by SEM is further confirmed by FACS analyses. The ratio of forward scatteringing of PRM to WT cells after 5 $hour$ of growth in log phase is approximately three-quarters. SEM also uncovers morphological differences between these two types of cells, as regards the cellular surface. PRM cells seem less undulated than WT cells. FACS analyses also attest to the same. The average side scatter ratio of the PRM cell population is much lower than the WT cell population. This indicates that the PRM cells are likely to possess lesser surface irregularities as compared to the WT cells [45]. Further details of FACS analyses are provided in SI 1. Results from SEM and FACS provide leads regarding alterations in surface properties of PRM cells. Therefore, we study the effect of isonicotinylhydrazide (more commonly referred to as isoniazid), which targets a unique component of the cell envelope in mycobacteria, namely mycolic acid.

3.1.3 Antibiotic sensitivity of PRM cells

Isoniazid (INH) is a widely used antibiotic to treat mycobacterial infections. It is well known that INH inhibits the synthesis of mycolic acid, which is one of the key components of the mycobacterial cell wall [46]. We observe that the INH sensitivity of PRM cells differs from WT cells. In our experimental setup, both WT and PRM cells were incubated overnight over a wide range of INH concentrations to check their susceptibility, as detailed in Section 2.2. OD_{600} of each sample obtained after overnight incubation, is plotted against the corresponding INH concentration for both cultures in Fig. 2(d). We observe that PRM cells are less sensitive towards INH than WT cells. All these physiochemical properties of PRM cells appear to be associated with multiple alterations at the genome level. In order to identify these genetic alterations of PRM cells, whole-genome Illumina sequencing was conducted.

3.1.4 Comparative analysis of the whole-genome sequence of PRM using Illumina next-generation sequencing

Illumina next-generation sequencing of PRM genome initially provides us 10646 contig sequences. The longest contig sequence indicates the highest mapping reads with *Mycobacterium smegmatis* mc^2155 complete genome, which has $CP009494.1$ as its gene bank id. Hence, MSG mc^2155 is used as the reference for assembly and variant identification. Application of Qiagen's CLC Fixed Ploidy Variant Detection module yields 116 variants located in various genes. Variants that lead to a change in the resultant amino acid are classified as non-synonymous. Variants that do not lead to a change in the resultant amino acid are classified as synonymous. Thus 57 mutations are

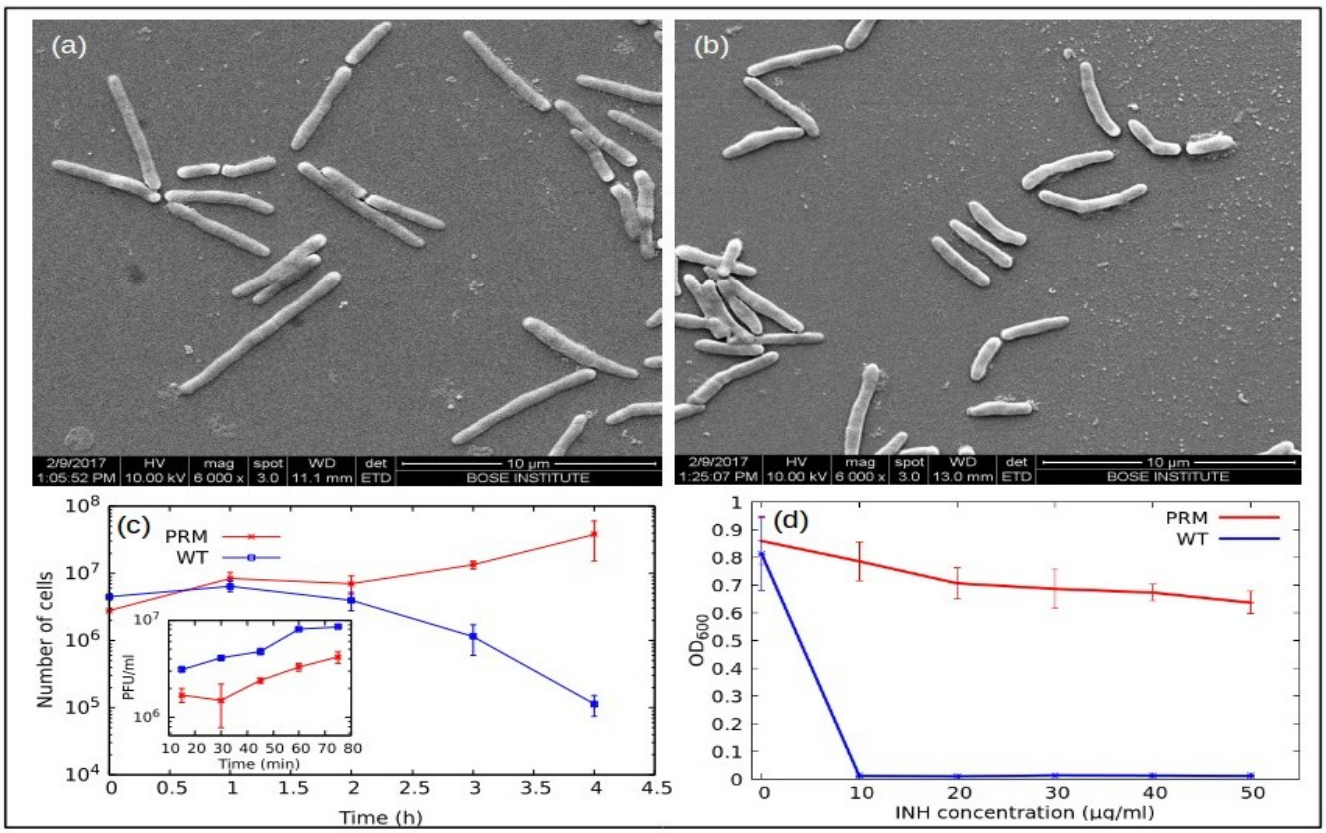

Figure 2: Comparison of physiochemical properties of WT and PRM cells. SEM image of cells of: (a) WT, (b) PRM. Comparison of WT and PRM demonstrates thereduction in the average length of PRM cells. The mean length of randomly picked statistically significant number of cells is provided with standard deviation. (c) Comparison of interaction dynamics of D29 with wild type (WT) and phage resistant mutant (PRM) cells of *Mycobacterium smegmatis* mc^2155. In the presence of phages, PRM cells grow well though the number of WT cells decreases after 1 *hour*. (Inset) Compared to PRM, WT cells adsorb more phages — which saturates by the end of its latent period of 60 *min*. Upon overnight incubation at various concentrations of INH: (d) shows O.D. count of WT and PRM cells. Each data point represents the mean ± standard deviation derived from three biological replicates.

non-synonymous and 59 are synonymous. Functional enrichment indicates that 20 functional proteins and 8 hypothetical proteins have been altered in the PRM genome as an effect of the random mutagenesis. Details of all mutations are listed in *SI* 3. However, most of the identified mutations do not exhibit a direct relationship with the observed phenotype.

3.2 Shortest path alteration fraction leads to a systems view of the observed phenotypes in the mutant

Hereafter, we will mostly refer to the PPIN for WT and PRM as $\mathcal{WT}$ and $\mathcal{PRM}$ respectively. We present the results of extensive network analyses using well-known topological metrics and the machinery of differential networks [26] in SI 2. Further biological details are presented in *SI* 4. However, we do not obtain a clear picture towards understanding the relationship between the mutations and phenotypes. We now demonstrate how SPAF can shed valuable insights, when conventional topological analyses of networks might not present a clear picture.

Of particular interest is the presence of non-overlapping shortest paths between a pair of proteins, where these paths are present in either but not both PPIN. Would the biological function associated with such protein pairs likely have been modified? If so, the signature of these modified paths should be manifested in the observed phenotypes. $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ identifies all such pairs of nodes at the ends of shortest paths in $\mathcal{WT}$ and $\mathcal{PRM}$. From hundreds of thousands of viable shortest paths, $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ selects 3428 such protein pairs. From the function related to the terminal proteins, all identified pairs of proteins at the end of these 3428 shortest paths can be broadly classified into 5 groups, which are respectively associated with: (1) cellular metabolism– especially fatty acid metabolism, (2) cell wall formation and cell division, (3) cellular transport and signalling, (4) information processing — especially translation, and, (5) cell redox homeostasis. The five groups and their possi-

Microbial species $(\mathcal{N}_\mathcal{W}, \mathcal{E}_\mathcal{W})$	**Mutations** $(\mathcal{N}_\mathcal{M}, \mathcal{E}_\mathcal{M})$ $\mathcal{N}_\mathcal{L}$	**Reported phenotypic changes**	**Functions associated with end proteins identified using** $\mathcal{F}_{l_{ij},l'_{ij}} = 1$	**P, Z p-value s, t-value** $(\mathcal{F}_{l_{ij},l'_{ij}} = 1)$	**P′ , Z′ p′-value s′, t′-value** (random)
Bacillus subtilis $(1346, 1542)$	ccpA gene $(1345, 1538)$ *3586*	Defects in biosynthesis of acetoin and other metabolites [32], Repression of sporulation [33], Transcriptional down regulation [33]	*Cellular metabolism, Sporulation cascade, Transcriptional regulation,* Stress response	0.799, 14.931 $< 10^{-12}$ 0.161, 37.312	0.229, -3.042 $< 1.17 \times 10^{-3}$ 0.176, -7.243
Escherichia coli $(1577, 6456)$	RecA gene $(1574, 6440)$ *5738*	Deficiency in DNA repair [34], Noisy gene expression [35], SOS response [36]	*DNA repair, Gene regulation,* Metabolic regulation	0.888, 45.204 $< 10^{-12}$ 0.099, 143.403	0.192, -11.143 $< 10^{-12}$ 0.155, -28.288
Helicobacter pylori $(374, 2218)$	FlaA gene $(373, 2213)$ *14*	Cellular motility [37]	*Flagellar assembly*	——	——
Saccharo -myces cerevisiae $(3821, 40936)$	SSD1 gene $(3820, 40929)$ *9411*	Cell wall integrity [38], Altered virulence [39], Cell cycle regulation [40]	*Cell wall and membrane integrity, Cellular growth and cell division, Virulence activity,* Gene regulation and signaling	0.715, 3.365 $< 3.9 \times 10^{-4}$ 0.203, 7.462	0.114, -41.555 $< 10^{-12}$ 0.1, -130.8
Salmonella typhimurium $(1136, 4281)$	Aro genes (total 8) $(1124, 4249)$ *7497*	Cell wall & membrane integrity [41], Alteration in lipid and amino acid metabolism and transport [42, 43]	*Cell wall and membrane formation, Lipid and amino acid metabolism and transport,* Metabolic regulation	0.726, 5.158 $< 1.3 \times 10^{-7}$ 0.198, 11.574	0.156, -22.451 $< 10^{-12}$ 0.132, -61.886

Table 1: Well-studied phenotypic alterations in mutants across five microbes. $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ successfully leads to known phenotypes from literature (*in italics*), while also predicting hitherto unknown phenotypes towards experimental verification. $\mathcal{N}_\mathcal{W}$ and $\mathcal{E}_\mathcal{W}$ represent the number of nodes and edges in the wild type PPIN respectively, while $\mathcal{N}_\mathcal{M}$ and $\mathcal{E}_\mathcal{M}$ represent the same for the mutant PPIN. $\mathcal{N}_\mathcal{L}$ represents the number of shortest paths with $\mathcal{F}_{l_{ij},l'_{ij}} = 1$. Values of P, Z, s, t and P', Z', s', t' for one proportion test of significance of the identified and random pairs respectively are displayed in the last two columns. The results demonstrate that 70% or even more of protein pairs identified via $\mathcal{F}_{l_{ij},l'_{ij}} = 1$, are reported to be functionally associated with the phenotypes at a confidence level of 99%. In contrast, 25% or even lesser of an identical number of randomly selected protein pairs are reported to be functionally associated with the phenotypes at a confidence level of 99%. Tests have not been conducted on *H. pylori* as $\mathcal{N}_\mathcal{L} < 30$.

ble impact on the observed phenotypic characteristics are depicted in Fig. 3. The complete constitution of these groups is provided in *SI* 5, where the columns in a group represent proteins at the end of shortest paths obtained using $\mathcal{F}_{l_{ij},l'_{ij}} = 1$.

3.3 The importance of shortest path alteration fraction: application to diverse microbes

We evaluate the importance of SPAF in a well-studied mutant in each of five different microbial species. These mutations are *ccpA* for *Bacillus subtilis*, *recA* for *Escherichia coli*, *flaA* for *Helicobacter pylori*, *SSD1* for *Saccharomyces cerevisiae* and *aro* for *Salmonella typhimurium*. The first four mutants are generated as a result of a single mutation in the respective species. The last is due to eight mutations. The wild type PPIN for SPAF analysis were obtained from the STRING database. For all the cases studied above, a cutoff score of 950 was selected to minimise hypothetical interactions, as described in Section $S1$ of *SI* 2. PPIN of the mutant has been constructed by deleting the specific mutated protein/(s) from the respective PPIN of the wild type. A summary of these mutants and their phenotypic characteristics, both reported as well as predicted, is provided in Table 1. The protein pairs associated with these phenotypic groups are located at the ends of select shortest paths. These paths are successfully identified using $\mathcal{F}_{l_{ij},l'_{ij}} = 1$, as prescribed in Theorem 1. A detailed account of the protein pairs for the microbes thus identified, is provided in *SI* 6. Their classification has been undertaken by biological function.

To test the statistical significance of these results for every organism, we randomly select node pairs equal in number to that identified via $\mathcal{F}_{l_{ij},l'_{ij}} = 1$. These randomly chosen pairs are drawn from a distribution of shortest path lengths, akin to the one governing the protein pairs identified via $\mathcal{F}_{l_{ij},l'_{ij}} = 1$. We then perform two statistical tests to verify whether the reported functional association of the pairs identified by $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ is significantly higher than the pairs selected at random.

We first perform a one proportion Z-test for protein pairs identified via $\mathcal{F}_{l_{ij},l'_{ij}} = 1$; against the null hypothesis, $\mathcal{H}_0$, that *less than* P_0% of the identified protein pairs is reported in literature to be functionally associated with the phenotypes of the given mutant. We choose $P_0 = 70$ and a confidence level of 99%. We next

perform this test for the randomly selected protein pairs. Here, our null hypothesis, $\mathcal{H}'_0$, is that *greater than* $P'_0\%$ of the randomly selected pairs is reported to be functionally associated with these phenotypes. As before, we choose a confidence level of 99% and $P'_0 = 25$. The protein pairs identified using $\mathcal{F}_{l_{ij},l'_{ij}} = 1$, are listed in SI 6. Similarly, randomly identified protein pairs are listed in SI 7. $\mathcal{N}_\mathcal{L}$ represents the number of identified shortest paths possessing $\mathcal{F}_{l_{ij},l'_{ij}} = 1$. Out of $\mathcal{N}_\mathcal{L}$ pairs, $\mathcal{N}_\mathcal{R}$ pairs have been reported in literature. Thus, $P = \mathcal{N}_\mathcal{R}/\mathcal{N}_\mathcal{L}$. We test our null hypothesis $\mathcal{H}_0 : P < P_0$, against the alternative hypothesis, $\mathcal{H}_a : P \geq P_0$ for proteins identified using $\mathcal{F}_{l_{ij},l'_{ij}} = 1$. Alternately, for $\mathcal{N}'_\mathcal{L}$ randomly selected protein pairs (where obviously $\mathcal{N}'_\mathcal{L} = \mathcal{N}_\mathcal{L}$); $\mathcal{N}'_\mathcal{R}$ pairs have been reported in literature. Thus, $P' = \mathcal{N}'_\mathcal{R}/\mathcal{N}'_\mathcal{L}$. We test our null hypothesis, $\mathcal{H}'_0 : P' > P'_0$, against the alternative hypothesis, $\mathcal{H}'_a : P' \leq P'_0$. We shall reject $\mathcal{H}_0$ and accept $\mathcal{H}_a$, if Z exceeds the tabulated values [47] at a confidence level of 99%. To test against the null hypothesis, we calculate $Z = (P - P_0)/\sqrt{P(1-P)/\mathcal{N}_\mathcal{L}}$ for proteins identified via $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ and $Z' = (P' - P'_0)/\sqrt{P'(1-P')/\mathcal{N}'_\mathcal{L}}$ for randomly selected proteins. Table 1 lists the relevant values P, P', Z and Z'. We find that we can reject $\mathcal{H}_0$ and accept $\mathcal{H}_a$, implying that 70% or even more of the protein pairs identified by our method possess clear evidence in literature regarding functional association with the phenotypes. We also find that we can reject $\mathcal{H}'_0$ and accept $\mathcal{H}'_a$ at a confidence level of 99%, thereby implying that 25% or even lesser of the randomly identified protein pairs are reported to be functionally associated with the phenotypes.

We also perform the t-test on the same null hypothesis. The sample mean reflects the number of successes, i.e. the number of protein pairs demonstrating clear evidence of functional association in literature. The sample mean is thus easily be found to be P, and the variance, $s = P(1-P)$. We calculate $t = (P - P_0)/(P(1-P)/\sqrt{\mathcal{N}_\mathcal{L}})$. Similarly, we calculate $s' = P'(1-P')$ and $t' = (P' - P'_0)/(P'(1-P')/\sqrt{\mathcal{N'}_\mathcal{L}})$ for the randomly selected pairs. The values of t and t' are tabulated in Table 1. These indicate that we can accept $\mathcal{H}_a$ for protein pairs identified using $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ at a confidence level of 99% [47]. Similarly, the t'-values lead us to accept $\mathcal{H}'_a$ for the randomly selected pairs, at a confidence level of 99%. Thus, the conclusions of the t-test are in complete agreement with those of the Z test.

4 Discussion

We rigorously introduce a mathematical approach and the measure of shortest path alteration fraction to compare any two networks with common or identical nodes. Herein, it enables us to propose a broad mathematical approach to mutagenesis and empowers us with a systems perspective. We apply our measure for five well-studied cases of mutants across the microbial domain. SPAF based PPIN analysis successfully identifies all phenotypic changes previously reported in the literature. In addition, it makes predictions which have not yet been reported but can be tested experimentally. Simultaneously, we generate a phage resistant mutant (PRM) of MSG mc^2155 (WT). WGS results indicate various non-synonymous mutations at 20 protein-coding regions in the PRM genome. The distinguishing phenotypic characteristics of PRM cells as compared to WT cells are: (a) resistance to D29 phage, (b) average cell length, (c) isoniazid sensitivity, and, (d) surface morphology. As aforementioned, we construct the PPIN of WT and PRM, which are referred to as $\mathcal{WT}$ and $\mathcal{PRM}$ respectively. Comparison and analyses of $\mathcal{WT}$ and $\mathcal{PRM}$ using differential networks and well-known topological network measures fail to provide a clear picture. A detailed study of proteins at the ends of all these paths, reveals strong correlations with the observed phenotypes. Further, they can be sorted into five groups as outlined in Fig. 3. Modifications associated with these five groups of protein interactions are strongly correlated with the observed phenotypes. Alterations in the metabolism of fatty acids and their interactions with cell surface components seem to affect the cell wall character. These also affect the surface-associated receptors. Such changes would cause altered surface morphology and abnormal phage adsorption, leading to phage resistance [48, 49]. Moreover, phage resistance would also be caused by the altered translational paths and redox homeostasis as these affect phage propagation and phage-mediated secondary killing of the host cell respectively [8, 50]. Furthermore, alterations in various metabolic pathways and paths associated with cellular transport would cause alterations in the INH sensitivity of cells [51, 52]. Likewise, changes related to cell division proteins and their interactions with cell wall components would alter the average cell length [53, 54]. As well known, a phenotype may emerge from complex interactions of proteins arising as a result of mutations in several genes [55]. Indeed, the phenomenon of genetic pleiotropy [56] portrays a completely different aspect of such complexity; wherein a single gene can even be responsible for two or more completely unrelated effects. However, it is not limited merely to protein-protein interaction networks and can be applied to diverse problems, across the sciences. Mutations in one or more residues at the level of a single protein could be tackled with our approach using the framework of protein contact networks or residue interaction graphs [29]. It can also be naturally applied to a diverse variety of biological networks — ranging from structural and molecular biology to ecology. To illustrate, seasonal or environment effects would likely leave their imprint over time on ecological networks, which could be studied using SPAF. Indeed, we expect the wide use of SPAF not merely in biological networks but in all kinds of networks.

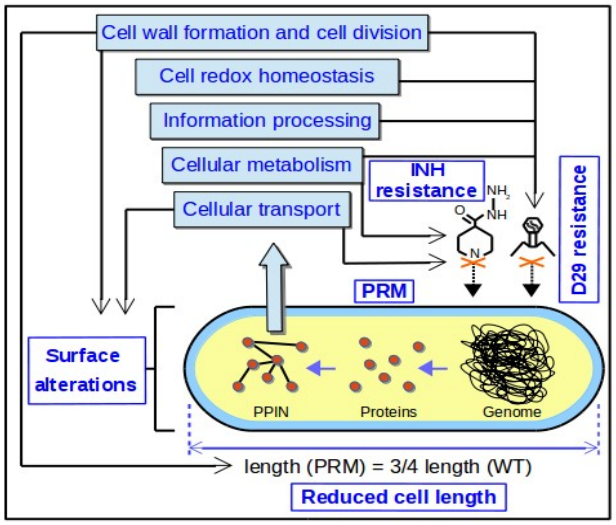

Figure 3: $\mathcal{F}_{l_{ij},l'_{ij}} = 1$ leads to the identification of five groups of protein pairs at both the ends of select shortest paths. These groups of altered interactions are strongly correlated with the observed phenotypes.

Acknowledgements

We thank Ritabrata Munshi (ISI, Kolkata), Mahan Mj. (TIFR, Mumbai), Deep Nath (Bose Institute, Kolkata) and the anonymous Reviewers of our manuscript for helpful discussions and comments.

Author contributions: Sinha (SS1) and Samaddar (SS2) conducted the experiments. All authors contributed to experimental design. SR obtained the mathematical results. SS1 conducted the network analyses. SR and SS1 designed theoretical research, conducted statistical analyses and wrote the manuscript.

Conflict of interest: none declared

References

[1] Rattan,A. *et al.* (1998). Multidrug-resistant *Mycobacterium tuberculosis*: molecular perspectives. Emerg. Infect. Dis. 4, 195.

[2] Maeda,S. *et al.* (2001). Multidrug resistant *Mycobacterium leprae* from patients with leprosy. Antimicrob. Agents Chemother. 45, 3635-3639.

[3] Hatfull,G.F. (2014). Mycobacteriophages: windows into tuberculosis. PLOS Pathog. 10, e1003953.

[4] Hankin,E.H. (1896). L'action bactéricide des eaux de la Jumna et du Gange sur le vibrion du cholera. Annales de l'Institut Pasteur, 10, 511-523.

[5] d'Herelle,F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. Comptes Rendus Academie des Sciences, 165, 373-375.

[6] Kutateladze,M. & Adamia,R. (2008). L'expérience de l'Institut Eliava en phagothérapie. Médecine et maladies infectieuses, 38, 426-430.

[7] McNerney,R. (1999). TB: the return of the phage. A review of fifty years of mycobacteriophage research. Int. J. Tuberc. Lung D. 3, 179-184.

[8] Samaddar,S. *et al.* (2016). Dynamics of mycobacteriophage-mycobacterial host interaction: evidence for secondary mechanisms for host lethality. Appl. Environ. Microbiol. 82, 124-133.

[9] Sinha,S. *et al.* (2018). Modeling bacteria-phage interactions and its implications for phage therapy. Adv. Appl. Microbiol. 103, 103-141.

[10] Sinha,S. *et al.* (2020). Modeling bacteria-phage dynamics, Methods in Mol. Biol. 2131, 309-327

[11] Matsuzaki,S. *et al.* (2005). Bacteriophage therapy: a revitalized therapy against bacterial infectious diseases. J. Infect. Chemother. 11, 211-219.

[12] Labrie,S.J. *et al.* (2010). Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317.

[13] Pingoud,A. *et al.* (2005). Type II restriction endonucleases: structure and mechanism. Cell. Mol. Life Sci. 62, 685.

[14] Barrangou,R. *et al.* (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315, 1709-1712.

[15] Bernheim,A., & Sorek,R. (2020). The pan-immune system of bacteria: antiviral defense as a community resource. Nature Reviews Microbiology, 18, 113-119.

[16] Marrakchi,H. *et al.* (2014). Mycolic acids: structures, biosynthesis, and beyond. Chem. Biol. 21, 67-85.

[17] Rastogi,N. *et al.* (1981). Multiple drug resistance in *Mycobacterium avium*: is the wall architecture responsible for exclusion of antimicrobial agents? Antimicrob. Agents Chemother. 20, 666-677.

[18] Baldwin,S.L. *et al.* (2019). The complexities and challenges of preventing and treating nontuberculous mycobacterial diseases. PLOS Negl. Trop. Dis. 13, e0007083.

[19] Hatfull,G.F. (2014). Molecular genetics of mycobacteriophages. Molecular Genetics of Mycobacteria, 81-119.

[20] McNerney,R. & Traore,H. (2005). Mycobacteriophage and their application to disease control. J. Appl. Microbiol. 99, 223-233.

[21] Hatfull,G.F. *et al.* (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS Genet. 2, e92.

[22] Garza-Ramos,G. *et al.* (2001). Binding site of macrolide antibiotics on the ribosome: new resistance mutation identifies a specific interaction of ketolides with rRNA. J. Bacteriol. 183, 6898-6907.

[23] Roy,S. (2012). Systems biology beyond degree, hubs and scale-free networks: the case for multiple metrics in complex networks. Syst. Synth. Biol., 6, 31-34.

[24] Banerjee,S.J. *et al.* (2015). Slow poisoning and destruction of networks: Edge proximity and its implications for biological and infrastructure networks. Phys. Rev. E, 91, 022807.

[25] Banerjee,S.J. *et al.* (2015). Using complex networks towards information retrieval and diagnostics in multidimensional imaging. Sci. Rep. 5, 17271

[26] Grewal,R.K. *et al.* (2015). Mapping networks of light-dark transition in LOV photoreceptors. Bioinformatics, 31, 3608-3616.

[27] Deb,A. *et al.* (2020). Residue Interaction Dynamics in Vaucheria Aureochrome1 light-oxygen-voltage: Bridging Theory and Experiments. Proteins, 88, 1660-1674

[28] Banerjee,S. & Mitra,D., (2020). Structural basis of design and engineering for advanced plant optogenetics, Trends Plant Sci., 25, 35-65.

[29] Grewal,R.K. & Roy,S (2015). Modeling proteins as residue interaction networks. Prot. Pept. Lett. 22, 923-933.

[30] Delbrück,M. (1940). The growth of bacteriophage and lysis of the host. J. Gen. Physiol. 23, 643-660.

[31] Kapopoulou,A. *et al.* (2011). The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis, 91, 8-13.

[32] Turinsky,A.J. *et al.* (2000). Bacillus subtilis ccpA gene mutants specifically defective in activation of acetoin biosynthesis. J. Bacteriol. 182, 5611-5614.

[33] Moreno,M.S. *et al.* (2001). Catabolite repression mediated by the CcpA protein in Bacillus subtilis: novel modes of regulation revealed by whole-genome analyses. Mol. Microbiol. 39, 1366-1381.

[34] Bryant,F.R. (1988). Construction of a recombinase-deficient mutant recA protein that retains single-stranded DNA-dependent ATPase activity. J. Biol. Chem. 263, 8716-8723.

[35] Elowitz,M.B. *et al.* (2002). Stochastic gene expression in a single cell. Science, 297, 1183-1186.

[36] Cox,M.M. (2007). Regulation of bacterial RecA protein function. Crit. Rev. Biochem. Mol. Biol. 42, 41-63.

[37] Josenhans,C. *et al.* (1995). Comparative ultrastructural and functional studies of Helicobacter pylori and Helicobacter mustelae flagellin mutants: both flagellin subunits, FlaA and FlaB, are necessary for full motility in Helicobacter species. J. Bacteriol. 177, 3010-3020.

[38] Kaeberlein,M. & Guarente,L. (2002). Saccharomyces cerevisiae MPT5 and SSD1 function in parallel pathways to promote cell wall integrity. Genetics, 160, 83-95.

[39] Wheeler,R.T. *et al.* (2003). A Saccharomyces cerevisiae mutant with increased virulence. Proc. Natl. Acad. Sci. U.S.A. 100, 2766-2770.

[40] Li,L. *et al.* (2009). Budding yeast SSD1-V regulates transcript levels of many longevity genes and extends chronological life span in purified quiescent cells. Mol. Biol. Cell, 20, 3851-3864.

[41] Sebkova,A. *et al.* (2008). aro mutations in Salmonella enterica cause defects in cell wall and outer membrane integrity. J. Bacteriol. 190, 3155-3160.

[42] Felgner,S. *et al.* (2016). aroA-deficient Salmonella enterica serovar Typhimurium is more than a metabolically attenuated mutant. mBio, 7, e01220-16.

[43] Thorne,G.M., & Corwin,L.M. (1975). Mutations affecting aromatic amino acid transport in *Escherichia coli* and *Salmonella typhimurium*. Microbiol., 90, 203-216.

[44] Basiji,D.A. *et al.* (2007). Cellular image analysis and imaging by flow cytometry. Clin. Lab. Med. 27, 653-670.

[45] Ramirez,J.M. *et al.* (2013). Side scatter intensity is highly heterogeneous in undifferentiated pluripotent stem cells and predicts clonogenic self-renewal. Stem Cells Dev. 22, 1851-1860.

[46] Somoskovi,A. *et al.* (2001). The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in *Mycobacterium tuberculosis*. Respir. Res. 2, 164.

[47] Rosner,B (2015). Fundamentals of Biostatistics. Eighth Edition, Cengage, Boston, MA, USA

[48] Chen,J. *et al.* (2009). Defects in glycopeptidolipid biosynthesis confer phage I3 resistance in *Mycobacterium smegmatis*. Microbiology, 155, 4050-4057.

[49] Zhao,X. *et al.* (2017). Transcriptomic and Metabolomic profiling of Phage-Host Interactions between Phage PaP1 and *Pseudomonas aeruginosa*. Front. Microbiol. 8, 548.

[50] Swift,B.M. *et al.* (2014). Factors affecting phage D29 infection: a tool to investigate different growth states of mycobacteria. PLOS ONE, 9, e106690.

[51] Quemard,A. *et al.* (1991). Isoniazid inhibition of mycolic acid synthesis by cell extracts of sensitive and resistant strains of *Mycobacterium aurum*. Antimicrob. Agents Chemother. 35, 1035-1039.

[52] Bardou,F. *et al.* (1998). Mechanism of isoniazid uptake in *M. tuberculosis*. Microbiology, 144, 2539-2544.

[53] Jain,P. *et al.* (2018). Delineating FtsQ-mediated regulation of cell division in *Mycobacterium tuberculosis*. J. Biol. Chem. 293, 12331-12349.

[54] Dasgupta,A. *et al.* (2006). The serine/threonine kinase PknB of *M. tuberculosis* phosphorylates PBPA, a penicillin-binding protein required for cell division. Microbiology, 152, 493-504.

[55] McGary,K.L., Lee,I. & Marcotte,E.M. (2007). Broad network-based predictability of *Saccharomyces cerevisiae* gene loss-of-function phenotypes. Genome Biol. 8, R258.

[56] Stearns,F.W. (2010). One hundred years of pleiotropy: a retrospective. Genetics, 186, 767-773.