IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF HYPOTHETICAL PROTEINS IN YERSINIA PESTIS ANGOLA

NAVYA DALVI, SNEHAL HATWAR
NAGPUR
G H RAISONI UNIVERSITY

ABSTRACT:

The notorious gram-negative Coccobacillus enterobacterium Yersinia pestis is the cause of three devastating global plague pandemics. In order to effectively develop drugs and vaccines, in silico analysis of the putative proteins is required, given the current outbreak of this zoonotic illness. Since a large fraction of the proteome consists of hypothetical proteins, it is crucial to identify them both structurally and functionally. Using a variety of bioinformatic techniques and software, the current study assessed the physicochemical parameters, predicted homology-based 3D structure, and annotated functions of the hypothetical Y. pestis protein. The tertiary model, which was minimized energetically, was predicted using the Swiss Model. The model was deemed to be good by the quality assessment servers. Analysis of protein-protein interactions was done using the STRING service. Tools such as Pfam and InterPro were utilized for functional prediction.

INTRODUCTION:

A class of widely spread rod-shaped, gram-negative bacteria found in the environment is called Yersinia. There are 17 species in this genus, and 3 of them are harmful to people. Yersinia pestis, Yersinia enterocolitica, and Yersinia pseudotuberculosis are a few of these. Enterocolitica and pseudotuberculosis are both known to induce gastrointestinal illnesses; however, enterocolitica is more frequently encountered in people than pseudotuberculosis

Since its discovery in Hong Kong in 1894, Yersinia pestis has undergone taxonomy and typing study. The earliest use of phenotyping techniques based on phenotypic traits was in the subspecies level categorization of Y. pestis. These techniques included bio typing, serotyping, antibiogram analysis, bacteriocin typing, phage typing, and plasmid typing. Then, when molecular biological technology developed, techniques based on fatty acid content, outer membrane protein profiles, and bacterial mass fingerprinting were also applied to identify the populations within Y. pestis. Nevertheless, because Y. pestis is a very homogeneous species, the typing techniques described above could only offer poor resolution; for example, just one serotype and one phage type were found for the entire species.

A LOOK BACK AT YERSINIA PESTIS' HISTORY:

A zoonotic illness that has established persistent foci in the Americas, Africa, and Eurasia, the lethal plague is caused by the Gram-negative bacterium Yersinia pestis. The delicate equilibrium between Y. pestis-contaminated soils, burrowing and nonburrowing mammals with varying degrees of plague sensitivity, and the fleas that accompany them is what allows it to remain in the ecosystem. The pathogen is mostly spread by bites from infected fleas, which cause the painful, swollen lymph nodes known as "buboes" and subsequent septicemic propagation of the infection. On the other hand, primary pneumonic plague is brought on by droplet inhalation following close contact with infected mammals.
Despite remaining mortality, epidemic control is aided by point-of-care diagnosis, prompt antibiotic treatment, and confinement measures. Ectoparasite management and diligent surveillance of known plague foci are essential components of mandatory primary prevention. In Eurasia, plague is known to have afflicted human populations for at least 5,000 years. Yersinia pseudotuberculosis, a closely related enteric pathogen, is thought to have shared a common ancestor with Y. pestis genomes recovered from affected archaeological sites. Additionally, Y. pestis's ectoparasite transmissibility was maintained while it acquired the ymt gene during the Bronze Age. There have been three recorded historical pandemics, the first occurring in 541 AD and continuing to this day

Y. pestis is a closed pan-genome bacterial species that is extremely clonal. The genome of Y. pestis underwent extremely frequent rearrangement and genome degradation events during its history, according to comparative genomic research. There are five main branches in the genealogy of Y. pestis, and four of them appear to have originated from a "big bang" node connected to the Black Death. The branch length in the genealogical tree revealed over dispersion, which was allegedly caused by varied historical molecular clock that is associated with demographical effect by alternate cycles of enzootic disease and epizootic disease in sylvatic plague foci, even though the whole genome-wide variation of Y. pestis reflected a neutral evolutionary process.

REVIEW OF LITERATURE:

FEATURES OF THE YERSINIA PESTIS:

They examined the epidemiologic features and dispersion of 5,958 Yersinia pestis isolates obtained from humans, host animals, and insect vectors between 1950 and 2019 in 4 Marmota plague foci in China, as well as 1,067 human cases of the disease. The human case-fatality rate from the plague was 68.88%; overall, the trend declined gradually over time but varied significantly. The Marmota Himalaya plague focus accounted for the majority of detected human cases (98.31%) and isolates (82.06%) from all sources. Three stages can be distinguished from the tendency among human cases: 1950–1969, 1970–2003, and 2004–2019. Since 1926, there have been no confirmed human cases or isolates from the Marmota Siberian plague center. Nevertheless, Y. pestis is still spreading among animal hosts in the other three foci; ecological factors may have an impact on the local Y. pestis population.

THE GENOME OF YERSINIA PESTIS AND ITS BIOLOGICAL LIFESTYLE:

An overview of studies on the genome and evolutionary characteristics of Yersinia pestis, a relatively new disease that diverged from Yersinia pseudotuberculosis at least 5000 years ago. Yersinia pestis is a type of bacteria that has a closed genome and is extremely clonal. A comparative genomic analysis showed that throughout its history, the genome of Y. pestis underwent a great deal of rearrangement and genome degradation. Four of the five main branches of Y. pestis's genealogy appear to have originated from a "big bang" node connected to the Black Death. Despite the fact that Y. pestis's entire genome showed a neutral evolutionary process, the genealogical tree's branch length revealed over dispersion, which was allegedly brought about by a different historical molecular clock that is linked to a demographic effect through alternating cycles of enzootic and epizootic disease in sylvatic plague foci. Recent studies on Black Death and Justinian's plague victims using paleomicrobiology have confirmed that Y. pestis was the source of two past pandemics, but the etiological lineages may no longer exist.

INFECTION WITH YERSINIA PESTIS:

Yersinia pestis, a plague pathogen, was responsible for three deadly pandemics in history that claimed hundreds of millions of lives. Highly invasive Yersinia pestis causes acute septicemia, which is usually deadly to the victim if left untreated. Yersinia pestis employs a number of strategies to circumvent both the innate and adaptive immune systems in order to persist in the host and perpetuate the infection. For instance, infections caused by this organism are biphasic, with a large influx of phagocytes, the production of inflammatory cytokines, and significant tissue destruction during the "proinflammatory" phase that follows an initial "noninflammatory" phase of bacterial replication with minimal inflammation.

INFLUENZA IMMUNIZATION:

Yersinia pestis is the vector-borne illness that causes plague. Spread by fleas from rodent reservoirs, Y. pestis evolved from an intestinal bacterial progenitor approximately 6000 years ago via episodes of genome reduction and gene gain. Recent human outbreaks have brought attention to the seriousness of this paradigm for public health, as well as for our understanding of the evolution of harmful microorganisms. Pathogen-associated molecular patterns (PAMPs), the broad-range protease Pla, the Yersinia outer-membrane proteins (Yops), and iron capture systems are among the complex array of virulence determinants that Y. pestis uses to subvert the human immune system, permitting unrestricted bacterial replication in lymph nodes (bubonic plague) and in lungs (pneumonic plague).

SUPPLIES AND PROCEDURES:

Data Collection:

The primary information regarding the availability of genome sequences of *Yersinia pestis Angola (ypg)* have been gathered from the website www.genome.jp/**kegg**/. The genomic information having complete proteins sequences of complete ORFs have been gathered of strain of the same.

KEGG **Yersinia pestis Angola**

Genome info | Pathway map | Brite hierarchy | Module | Genome browser

Search genes: Go Clear

Genome information

T number	T00635
Org_code	ypg
Name	Yersinia pestis Angola (virulent Pestoides isolate)
Annotation	yes
Taxonomy	TAX: 349746
Lineage	Bacteria; Pseudomonadota; Gammaproteobacteria; Enterobacterales; Yersiniaceae; Yersinia
Brite	KEGG organisms [BR:br08601] KEGG organisms in the NCBI taxonomy [BR:br08610] KEGG organisms in taxonomic ranks [BR:br08611]
Data source	GenBank (Assembly: GCA_000018805.1 Complete Genome) BioProject: 16067
Keywords	Human pathogen, Animal pathogen
Disease	H00297 Plague
Comment	Virulent Pestoides isolate.
Chromosome	Circular
Sequence	GB: CP000901
Length	4504254
Plasmid	pMT-pPCP; Circular
Sequence	GB: CP000900

KEGG HYPOTHETICAL PROTEIN RETRIEVAL: -

Y. Pestis's hypothetical protein sequences were found using limit search on KEGG under the term "ypg hypothetical protein." The genome contains all 898 hypothetical proteins, but we only used 250 of them for our functional annotation analysis.

SEARCH FOR CONSERVED SEQUENCES: -

The analysis employed hypothetical proteins that were screened for the existence of functional sections in their sequences, such as superfamily/s regions, conserved domains, and motif searches. Based on these findings, the hypothetical proteins' functions were predicted. The following online tools were used to forecast the function:

A) CONSERVED DOMAIN BLAST:

Conserved Domain BLAST was used to search for and analyze the functionality of the hypothetical protein.The CDD 27036 PSSMs database was utilized to search for conserved sections, with the "low complexity filter" turned on to eliminate any sequences that do not exhibit evolutionary links. The E-value parameter was set to 0.01.

B) INTERPROSCAN: -

The PROSITE pattern search program has been updated with a new feature called Pattern Scan. The InterPro team internally produced the ScanRegExp program, which is dependent on data that is no longer generated (confirm.patterns from Emotif) and is similar to the PROSITE code. As a result, it was decided to switch to ps_scan.pl, the same application as PROSITE, which verifies whether a match is a true positive utilizing evaluator mini-profiles. This leads to an increase in the coverage of True PROSITE matches and a more sensitive predictor of true matches.

InterPro:-

Protein domain and functional site databases are becoming essential tools for predicting the activities of proteins. A number of signature-recognition techniques have developed during the past ten years to handle various sequence analysis issues, leading to the creation of largely distinct and independent databases. Due to the various advantages and disadvantages of their underlying analytic techniques, these resources' optimal diagnostic application varies. Therefore, search strategies should ideally incorporate all of them for optimal results.

The goal of the collaborative project InterPro is to characterize a specific protein family, domain, or functional site in a unique and non-redundant way, thereby offering an integrated layer on top of the most widely used signature databases.PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, PIR superfamily, SUPERFAMILY, Gene3D, and PANTHER databases are all integrated into the InterPro database. The home website for the InterPro project is at http://www.ebi.ac.uk/interpro .

C) Pfam:

Multiple sequence alignments and hidden Markov models (HMMs) are used to represent each of the many protein families that make up the Pfam database. Generally speaking, proteins are made up of one or more domains—functional sections. The wide variety of proteins present in nature are produced by distinct domain combinations. Therefore, information on the function of proteins can be obtained by identifying the domains that exist inside them. Consisting of two parts: Pfam-A and Pfam-B. Pfam-A entrants are carefully selected, premium families. While a significant amount of the sequences in the underlying sequence database are covered by these Pfam-A entries, we additionally build a supplement using the ADDA database to provide a more thorough coverage of known proteins. Pfam-B is the term for these mechanically created entries. When no Pfam-A entries are found, Pfam-B families—despite their poorer quality—can be helpful in finding functionally conserved regions. Clans, which are higher-level groups of connected families, are also produced by Pfam. A group of Pfam-A entries connected by sequence, structure, or profile-HMM similarity is called a clan. Depending on the kind of data provided by each web tool for each hypothetical protein under research, the results of the protein functionality analysis were published in confidential limitations in % for assigning function to the hypothetical protein.

D) CATH:

There are 173 536 domains, 2626 homologous superfamilies, and 1313 fold groups in CATH version 3.5 (Class, Architecture, Topology, Homology; http://www.cathdb.info/). Focusing on structural genomics (SG) structures, we find that CATH v3.5 contains marginally fewer novel folds than prior releases. This finding implies that the majority of folds that are readily accessible for structure determination may now be known. Our functional family (FunFams) sub-classification technique has become more accurate, and we have expanded the CATH sequence domain search facility to include Fun Fam annotations for every domain. The website for CATH has undergone a facelift. We have made enhancements to the way that functional information and conserved sequence characteristics connected to FunFams in every CATH superfamily are shown.

THE SWISS MODEL:

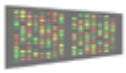
A server for automated comparative modeling of three-dimensional (3D) protein structures is called SWISS-MODEL (http://swissmodel.expasy.org). Since its inception in 1993, it has been at the forefront of automated modeling and is currently the most popular free web-based tool for automated modeling. The server processed 120,000 requests for 3D protein models from users in 2002. Through its Web interface, SWISS-MODEL offers multiple degrees of user participation. For example, in the "first approach mode," a protein's amino acid sequence is the sole input needed to create a 3D model. The server handles all aspects of template selection, alignment, and model construction automatically. The modeling procedure is predicated on a user-defined target-template alignment in the "alignment mode." An integrated sequence-to-structure workstation called Deep View (Switzerland-Pdb Viewer) can perform complex modeling activities in "project mode." Every model is returned via email along with a thorough modeling report. ANOLEA evaluations and What Check analyses are offered voluntarily. In the EVA-CM project, the dependability of SWISS-MODEL is continually assessed. To enhance the effective integration of expert information into a user-friendly server, the SWISS-MODEL server is continuously being developed

In light of the following guidelines, the Parameter of Confidence limit was established at 100%, 75%, 50%, 25%, and 0%.-

1. The confidence level would be 100% if all four of the instruments indicated the same functions.
2. The confidence level would be 75% if the three instruments provided indicated the same functions.
3. The confidence level would be 50% if the two instruments provided indicated the same functions.
4. The confidence level would be 25% if the four tools provided indicated the various functions.
5. The confidence level would be zero percent if the provided tool shows no functions.

METABOLIC INTERACTIONS OF HYPOTHETICAL PROTEINS:

The network of each hypothetical protein with other proteins was understood in the form of a protein interaction network, which could be further implemented to understand the precise role of these hypothetical proteins in the metabolism of Yersinia pestis. This allowed for the connection of the hypothetical proteins present in Yersinia pestis with the likely metabolic activity. The required online server uses STRING for each hypothetical protein's connecting network. A database of anticipated and known protein interactions is called STRING. The four sources of the interactions are as follows: direct (physical) and indirect (functional) associations

Genomic Context	High-throughput Experiments	(Conserved) Coexpression	Previous Knowledge
			

STRING conveys information between these creatures when necessary and quantitatively integrates interaction data for a large number of organisms from many sources. As of right now, 5'214'234 proteins from 1133 species are covered in the database. You can access the website at http://string-db.org/.

RESULTS:

FUNCTIONAL ANNOTATION TO THE HYPOTHETICAL PROTEINS:

The complete genome survey led to the finding that nearly 25% of proteins in the *Y. pestis* have been categorized as Hypothetical proteins. These proteins are which the function is yet to define and their existence remains obscure. But several methodologies have been designed to find out functionalities in the given amino acid sequence by the virtue of comparative proteomics which are armed with the potential web tools and well-developed searching strategies which can search essential enzymatic conserved domains within the protein sequence against known protein sequences which harbors these enzymatic conserved domains. These Conserved Domain Databases armed with the tools helped in the analysis of hypothetical proteins and nearly 250 hypothetical proteins have been analyzed for the presence of such enzymatic conserved domains which can assign predicted functions.

The search for motifs in the sequence of hypothetical protein have been done with CDD-BLAST, CATH, INTERPROSCAN and PFAM, all these web tools has accepted the FASTA format sequence of hypothetical protein and analyzed the hidden enzymatic conserved domains in all 898 hypothetical proteins. The study has classified out of 250 hypothetical proteins, enzyme function into four confidence level as follows with 100% are 3 , 75% are 6 , 50% are 7, 25% are 6 proteins total 22 enzyme coding have been found from 250 hypothetical proteins used in the study when compared with different web tools and represented in Table 01 and Table 02.

Table 2 Enzyme function with confidence level			
100%	75%	50%	25%
3	6	7	6

Similarly, The study has classified out of 250 hypothetical proteins, non-enzyme function into four confidence level as follows with 100% are 15 , 75% are 29, 50% are 24 , 25% are 23 proteins total 91 non enzyme coding have been found from 250 hypothetical proteins used in the study when compared with different web tools and represented in Table 3 and Table 4.

Table 4 Non enzyme functions with confidence level			
100%	75%	50%	25%
15	29	24	23

These proteins shown the conserved domain of the enzyme functions which was searched by the conserved domain programs using the matrix of pattern search. The combined results of the four programs suggested that even though, the genome have been marked with the more than 25 % of the genes for the hypothetical proteins by using the designed strategy of conserved domain search 250 hypothetical proteins showed the defined presence of conserved enzyme function sequence. Based on the % confidence it has been noteworthy that proteins with 100% confidence could be selected further for the *In Vivo* analysis which will assist in determining the fate of these well annotate hypothetical proteins in the metabolism of the bacterium.

PROTEIN STRUCTURE PREDICTION AND RAMACHANDRAN PLOT ANALYSIS:

In the present study total nine enzyme coding hypothetical protein searched for template successfully as shown in Table 5 , while the predicted enzyme models showcased in fig 1 , where formed 3D model and its Ramachandran plot for structural quality has been showcased.

Similarly, in the present study total forty four non enzyme coding hypothetical protein searched for template successfully as shown in Table 6, while the predicted non enzyme models showcased in fig 2, where formed 3D model and its Ramachandran plot for structural quality has been showcased.

Metabolic Linking by STRING:

Metabolic linking of hypothetical proteins is the key step in the functional biology which helps us in directing the proteins with particular activity of the protein in cellular function. STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations available in the organism. STRING results suggested that these hypothetical proteins have been linked with several known proteins involved in fatty acid synthesis, nucleotide synthesis and energy metabolism and of *Yersinia pestis*. These protein-protein interactions definitely helped to understand the role of each protein in the *Y. Pestis* life cycle. These metabolic linking could be evaluated in wet lab analysis in the coming time to decipher its truth in metabolism. Fig. 1.

	A	B	C	D	E	F	G
1	Table 5 Protein struture prediction of enzyme coding hypothetical proteins based on template homology						
2	KEGG number	Template	GMQE	QSQE	Identity	Method	Oligo State
3	YpAngola_A0374	Elongation factor P hydroxylase	0.94	-	88.27	AlphaFold v2	monomer ✓
4	YpAngola_A0623	Gamma carbonic anhydrase-like	0.98	-	75	AlphaFold v2	monomer ✓
5	YpAngola_A0642	YedF	0.8	-	90.91	NMR	monomer ✓
6	YpAngola_A0657	Citrate lyase beta subunit-like	0.94	-	100	AlphaFold v2	monomer ✓
7	YpAngola_A0658	Cysteine protease StiP-like	0.95	-	100	AlphaFold v2	monomer ✓
8	YpAngola_A0659	Trehalose phosphatase-like	0.93	-	98.52	AlphaFold v2	monomer ✓
9	YpAngola_A0816	Inosine/xanthosine triphosphatase	0.84	0.74	57.65		homo-dimer ✓
10	YpAngola_A0971	HicA mRNA interferase family	0.93	-	100	X-ray, 2.1Å	hetero-tetramer ⚠
11	YpAngola_A1240	Type IV methyl-directed restriction enzyme EcoKMcrBC	0.85	-	95.43	AlphaFold v2	monomer ✓

Table 6 Protein struture prediction of non enzyme coding hypothetical proteins based on template homology

KEGG number	Interprosean	GMQE	QSQE	Identity	Method	Oligo State
YpAngola_A0082	LTXXQ motif family protein	0.76	-	100	AlphaFold v2	monomer ✓
YpAngola_A0102	Cell division protein ZapB	0.9	0.74	75.95	X-ray, 2.8Å	homo-dimer ✓
YpAngola_A0169	Type VI secretion system, RhsGE-associated Vgr family subset	0.9	-	79.16	AlphaFold v2	monomer ✓
YpAngola_A0171	Type VI secretion system effector Hcp	0.95	-	92.02	AlphaFold v2	monomer ✓
YpAngola_A0174	Type VI secretion system TssK	0.89	-	98.89	AlphaFold v2	monomer ✓
YpAngola_A0176	Type VI secretion system sheath protein TssB1	0.82	-	79.88	AlphaFold v2	monomer ✓
YpAngola_A0265	Putative tight adherence pilin protein F	0.86	-	100	AlphaFold v2	monomer ✓
YpAngola_A0295	SH3 domain protein	0.8	-	91.26	AlphaFold v2	monomer ✓
YpAngola_A0374	Elongation factor P hydroxylase	0.94	-	88.27	AlphaFold v2	monomer ✓
YpAngola_A0406	Type VI secretion system effector Hcp	0.93	-	99.37	AlphaFold v2	monomer ✓
YpAngola_A0411	DUF1407/YfgJ-like	0.92	-	98.65	AlphaFold v2	monomer ✓
YpAngola_A0419	Cytoskeleton protein RodZ	0.72	-	98.55	AlphaFold v2	monomer ✓
YpAngola_A0429	ISC system FeS cluster assembly, IscX	0.95	-	77.27	NMR	monomer ✓
YpAngola_A0490	Conserved hypothetical protein CHP03034	0.88	-	89.73	AlphaFold v2	monomer ✓
YpAngola_A0609	Alternative ribosome-rescue factor A	0.85	-	78.79	AlphaFold v2	monomer ✓
YpAngola_A0623	Gamma carbonic anhydrase-like	0.98	-	75	AlphaFold v2	monomer ✓
YpAngola_A0653	Intracellular heme transport protein HutX-like	0.88	0.77	58.28	X-ray, 2.0Å	homo-dimer ✓
YpAngola_A0656	Probable tellurium resistance transcriptional regulator TerW	0.88	-	100	AlphaFold v2	monomer ✓
YpAngola_A0657	Citrate lyase beta subunit-like	0.94	-	100	AlphaFold v2	monomer ✓
YpAngola_A0658	Cysteine protease StiP-like	0.95	-	100	AlphaFold v2	monomer ✓
YpAngola_A0659	Trehalose phosphatase-like	0.93	-	98.52	AlphaFold v2	monomer ✓
YpAngola_A0661	ATP-grasp family	0.95	-	100	AlphaFold v2	monomer ✓
YpAngola_A0703	tRNA threonylcarbamoyl adenosine modification protein TsaE	0.95	-	73.72	AlphaFold v2	monomer ✓
YpAngola_A0758	Type VI secretion, TssG-like	0.86	-	99.69	AlphaFold v2	monomer ✓
YpAngola_A0759	Type VI secretion system TssF	0.89	-	100	AlphaFold v2	monomer ✓
YpAngola_A0761	Type VI secretion system effector Hcp	0.91	-	84.88	AlphaFold v2	monomer ✓
YpAngola_A0762	Type VI secretion system TssC-like	0.88	-	82.13	AlphaFold v2	monomer ✓
YpAngola_A0816	Inosine/xanthosine triphosphatase	0.84	0.74	57.65	X-ray, 2.3Å	homo-dimer ✓
YpAngola_A0943	FaeA-like protein	0.92	-	100	AlphaFold v2	monomer ✓
YpAngola_A0971	HicA mRNA interferase family	0.93	-	100	X-ray, 2.1Å	hetero-tetramer ⚠
YpAngola_A1022	Chaperone lipoprotein, PulS/OutS-like	0.92	-	98.15	AlphaFold v2	monomer ✓
YpAngola_A1046	Cell division protein ZapD	0.91	0.8	69.39	X-ray, 2.8Å	homo-dimer ✓
YpAngola_A1056	Polyketide cyclase/dehydrase	0.97	-	100	AlphaFold v2	monomer ✓
YpAngola_A1109	Modulator protein MzrA	0.79	-	100	AlphaFold v2	monomer ✓
YpAngola_A1112	Inner membrane protein YqjD/ElaB	0.72	-	78.22	AlphaFold v2	monomer ✓
YpAngola_A1121	UPF0102 protein YraN-like	0.92	-	100	AlphaFold v2	monomer ✓
YpAngola_A1136	Z-ring associated protein G-like	0.69	-	90.98	AlphaFold v2	monomer ✓
YpAngola_A1161	Periplasmic lysozyme inhibitor, I-type	0.84	-	100	AlphaFold v2	monomer ✓
YpAngola_A1170	Ribosome-associated, YjgA	0.89	-	97.8	AlphaFold v2	monomer ✓
YpAngola_A1180	Insecticidal toxin complex/plasmid virulence protein	0.97	0.66	96.38	EM	homo-pentamer ✓
YpAngola_A1193	Conserved hypothetical protein CHP02099	0.8	-	81.32	AlphaFold v2	monomer ✓
YpAngola_A1240	Type IV methyl-directed restriction enzyme EcoKMcrBC	0.85	-	95.43	AlphaFold v2	monomer ✓
YpAngola_A1374	RnfH protein	0.84	-	100	AlphaFold v2	monomer ✓
YpAngola_A1431	DNA-binding protein VF530-like	0.84	-	71.6	AlphaFold v2	monomer ✓

DISCUSSION:

The Bioinformatics based study of ORF's data available for the *Yersinia Pestis* has given the opportunity to elucidate enzyme functions in the hypothetical proteins for linking them to particular metabolic pathway. Study aimed at the metabolic linking and enzyme function finding in the 898 hypothetical proteins of *Y. Pestis* . Where we analyzed 250 hypothetical proteins have been computationally predicted to harbor enzymatic domains which could be further linked with various metabolic pathways. In those, number of proteins having confidence limit at least 50% which represented some better existence probability of them. The functionality finding in all available hypothetical proteins of *Y . Pestis* will remain the first priority in the future by all the research community for which our study provided leads to study. Similar studies were conducted by researcher to report probability of enzyme coding ability of hypothetical proteins in *Shigella flexneri*, *Bacillus anthracis*, *H. influenzae* and *Helicobacter pylori* [70-73].

CONCLUSION:

Study concluded that, available web tools linked with rich protein databases information proved to be beneficial and resultant filtered out some of the important enzyme coding hypothetical proteins out of whole pool of *Y. Pestis* proteomics and obtained data in future may assist in relational linking of today hypothetical proteins in tomorrow's functional proteins by engaging the work on *Yersinia Pestis* hypothetical genes and its cloning along with its expression study in model *E. coli* in order to aim the validation of predicted enzyme function.

REFERENCES:

- Brady, M. F., Yarrarapu, S. N. S., & Anjum, F. (2023). Yersinia Pseudotuberculosis. In StatPearls. StatPearls Publishing.
- Qi, Z., Cui, Y., Zhang, Q., & Yang, R. (2016). Taxonomy of Yersinia pestis. *Advances in experimental medicine and biology*, *918*, 35–78. https://doi.org/10.1007/978-94-024-0890-4_3
- Barbieri, R., Signoli, M., Chevé, D., Costedoat, C., Tzortzis, S., Aboudharam, G., Raoult, D., & Drancourt, M. (2020). Yersinia pestis: the Natural History of Plague. *Clinical microbiology reviews*, *34*(1), e00044-19. https://doi.org/10.1128/CMR.00044-19
- Cui, Y., & Song, Y. (2016). Genome and Evolution of Yersinia pestis. *Advances in experimental medicine and biology*, *918*, 171–192. https://doi.org/10.1007/978-94-024-0890-4_6
- He, Z., Wei, B., Zhang, Y., Liu, J., Xi, J., Ciren, D., Qi, T., Liang, J., Duan, R., Qin, S., Lv, D., Chen, Y., Xiao, M., Fan, R., Song, Z., Jing, H., & Wang, X. (2021). Distribution and Characteristics of Human Plague Cases and Yersinia pestis Isolates from 4 Marmota Plague Foci, China, 1950-2019. Emerging infectious diseases, 27(10), 2544–2553. https://doi.org/10.3201/eid2710.202239

- Cui, Y., & Song, Y. (2016). Genome and Evolution of Yersinia pestis. *Advances in experimental medicine and biology*, *918*, 171–192. https://doi.org/10.1007/978-94-024-0890-4_6
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. 2003 Jul 1;31(13):3381-5. doi: 10.1093/nar/gkg520. PMID: 12824332; PMCID: PMC168927.
- Bi Y. (2016). Immunology of Yersinia pestis Infection. *Advances in experimental medicine and biology*, *918*, 273–292. https://doi.org/10.1007/978-94-024-0890-4_10
- Demeure, C. E., Dussurget, O., Mas Fiol, G., Le Guern, A. S., Savin, C., & Pizarro-Cerdá, J. (2019). Yersinia pestis and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. Genes and immunity, 20(5), 357–370. https://doi.org/10.1038/s41435-019-0065-0
- Gore D. In silico Prediction of Structure and Enzymatic Activity for Hypothetical Proteins of Shigella flexneri. Biofrontiers. 2009;1, Issue.2: 1-10.
- Gore D and Raut A. Computational Function and Structural Annotations for Hypothetical proteins of Bacillus anthracis. Biofrontiers, (2009) 1:1:27-36.
- Dogra P and Gore D. Prediction of Enzymatic Function and Structure of *H. influenzae* Hypothetical Proteins - An In silico Approach. International Journal of Soft Computing and Bioinformatics .2010;1: 2 :67-77.
- Gore G, Denge P and Amrute M. Homology Modeling and Enzyme Function Prediction in the Hypothetical Proteins of Helicobacter pylori - an Insilico Approach. Biomirror. 2010; 1-5/ bm-1111251610.