FUNCTIONAL STUDIES ON CSIs AND CSPs

FUNCTIONAL SIGNIFICANCE OF NOVEL MOLECULAR MARKERS SPECIFIC FOR *DEINOCOCCUS* AND CHLAMYDIAE SPECIES

BY

F M NAZMUL HASSAN, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the

Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2017) McMaster University

(Biochemistry) Hamilton, Ontario

TITLE: Functional Significance of Novel Molecular Markers Specific for *Deinococcus* and Chlamydiae Species

AUTHOR: F M Nazmul Hassan, B.Sc. (Khulna University)

SUPERVISOR: Professor Radhey S. Gupta

NUMBER OF PAGES: xv, 108

This thesis is dedicated to my mom and dad. I lost my mother when I was a high school student. She always liked to share stories about science and inspired me to do something for the beneficial of human being. Her dream was guided by my father. I was always encouraged by my father to do higher education. I have lost my father this year. It was most critical moment of my life. But, I have continued to do research because of fulfill his dream.

THESIS ABSTRACT

The *Deinococcus* species are highly resistant to oxidation, desiccation, and radiation. Very few characteristics explain these unique features of *Deinococcus* species. This study reports the results of detailed comparative genomics, structural and protein-protein interactions studies on the DNA repair proteins from *Deinococcus* species. Comparative genomics studies have identified a large number of conserved signature indels (CSIs) in the DNA repair proteins that are specific for *Deinococcus* species. In parallel, I have carried out the structural and protein-protein interactions studies of CSIs which are present in nucleotide excision repair (NER), UV damage endonuclease (UvsE)-dependent excision repair (UVER) and homologous recombination (HR) pathways proteins. Comparative genomics studies have identified a 30 amino acid (aa) conserved signature indel (CSI) in the UvrA1 protein and a 1 aa CSI in the UvsE protein that are specific for *Deinococcus* species. I also have identified a 1 aa CSI in the RecA protein that is specific for Deinococcus-Thermus phylum. Structural analysis delineates that these CSIs are located at the surface loop of the protein. Protein-protein interactions analyses reveal that the UvsE protein of *Deinococcus* species is present in a genetic linkage with Zn-ribbon protein. RecA protein is also present in a genetic linkage with two proteins in Deinococcus-Thermus phylum. UvsE and RecA proteins are predicted to be part of same operon with the interacting proteins. The UvrA protein is also present in genetic linkage with disulfide oxidoreductase (DsbA), disulfide bond formation protein (DsbB) and one of our labs identified conserved signature protein (CSP). These four proteins are indicated to be part of the same operon. We have proposed a model to understand the functional

role of these four proteins in together in DNA repair. According to our model, the predicted linkage of the UvrA protein via CSP to the DsbA/DsbB is expected to restore the activity of the NER pathway in *Deinococcus* species upon radiation exposure. Further experimental studies may lead to prove our hypothesis and to understand the functional role of CSI and CSP among *Deinococcus* species.

Similarly, I have carried out the comparative genomics studies in the GyrA and GyrB protein sequences which are supposed to play role in regulation of temporal gene expression during intracellular development. I have identified a number of molecular markers in the form of CSIs in the GyrA and GyrB protein sequences which are specific for all the members of Chlamydiae. Structural analysis indicates that two of these CSIs may play role in enhancing the supercoiling activity of GyrA and GyrB proteins of chlamydial species. I have carried out protein-protein interactions analyses to understand the functional role of other two CSIs of GyrA and GyrB proteins. Protein-protein interactions analyses delineate that the GyrA and GyrB proteins of chlamydial species are present in a unique genetic linkage with thymidylate kinase (TMK), DNA polymerase III subunit δ' (HolB) and one of the chlamydial specific CSP. We have proposed a model to understand the functional role of CSIs, CSP and also to delineate the functional role of these proteins together to chlamydial intracellular development. According to our model, these five proteins together play a role to increase the replication rate during the midcycle of intracellular development. CSP and CSIs are responsible for these unique associations among these proteins. Further experimental studies may lead to prove this hypothesis and will guide to decipher the intricateness of chlamydial intracellular development.

ACKNOWLEDGEMENTS

It is my pleasure to express my gratitude the many people who made this thesis possible. First of all, I would like to thank my supervisor Dr. Radhey S. Gupta for taking me on board his team to work on this challenging project and for his supervision. I still remember, when I started this project I have little idea in this field. But you have given me guidance to build my knowledge in this field and also to enhance my research capacities. Throughout my whole life, I will remember your word "critical thinking is important for research". I always wonder how you conduct research in different biological aspects of living organisms. You are my inspiration to do research. Your mentorship is invaluable and I always feel bless to have a supervisor like you. Your guidance and support for my research and personal life are beyond my expectation when I have joined the lab. Thanks for everything my supervisor.

I would also like to thank my supervisory committee, Dr. Herb E. Schellhorn and Dr. Russell Bishop for their guidance and feedback during committee meetings. Their valuable advice has helped guide my research and given me a more critical understanding of the field.

I also would like to extend my appreciation to colleagues in the Dr. Gupta lab. I would like to provide special thanks Dr. Mobolaji Adeolu and Bijendra Khadka for their continuous support throughout my research.

Finally, I am thankful to my family and friends. I could not able to finish my thesis without your love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

LIST OF FIGURES

CHAPTER 2

LIST OF ABBREVIATIONS

aa……..…Amino Acid

BLAST……Basic Local Alignment Search Tool

BLASTp……..Protein vs Protein BLAST search

CSI……Conserved Signature Indel

CSP……Conserved Signature Protein

dCSP-I……*Deinococcus* specific CSP-I

dCSP-II…...……...*Deinococcus* specific CSP-II

DNA……...Deoxyribonucleic Acid

DsbA……...Disulfide Oxidoreductase

DsbB……Disulfide bond formation protein

E value……Expected value

EB……....Elementary Body

GyrA……DNA gyrase subunit A

GyrB……DNA gyrase subunit B

HPI……...Hour Post Infection

HolB……DNA polymerase III subunit δ'

HR……Homologous Recombination

Indel……..Insertion or Deletion

MR……Mismatch repair

NCBI……..National Center for Biotechnology Information

NER……Nucleotide Excision Repair

PDB……..Protein Data Bank

PSI-BLAST……Position-Specific Iterative BLAST

RB……...Reticulate Body

RNA……...Ribonucleic Acid

STRING……...Search Tool for the Retrieval of Interacting Genes/Proteins

TMK…… Thymidylate Kinase

UvrA……..Excinuclease ABC subunit A

UvrC……...Excinuclease ABC subunit C

UvrD…….DNA helicase UvrD protein

UvsE…….UV damage endonuclease

CHAPTER 1

Introduction

1.1 Background:

Large-scale genome sequencing is providing researchers with enormous wealth of information that can be used to identify molecular traits, shared uniquely by evolutionarily related groups of organisms. The molecular signatures in genes/proteins sequences that have occurred at various stages in evolution are ideally suited for use in evolutionary and biochemical studies (Stackebrandt & Schumann, 2006; Gupta, 2014; Singh & Gupta, 2009; Alnajar et al., 2017; Khadka and Gupta, 2017). One important class of molecular signatures that has been used extensively to understand the evolutionary relationships among different groups of organisms are conserved signature insertions and deletions, i.e., Indels, (CSIs). CSIs are unique types of signature sequences of defined lengths present at specific locations in different genes and proteins in particular group of organisms (Alnajar and Gupta, 2017; Gupta, 2014; Gupta et al., 2015; Gupta, 2016b; Gupta et al., 2016). The useful CSIs are flanked on both sides by conserved residues which ensures that they constitute reliable characteristics and are not resulting from alignment errors/artefacts, CSIs represent highly specific and rare genetic changes and they are of common evolutionary origin and unlikely to arise in different groups due to convergent evolution (Gupta, 2014; Rokas and Holland, 2000). The high conservation of their location in the protein sequences indicates that these genetic changes have high functional significance and are likely under significant selective pressure for retention (Zhi et al., 2012). Earlier work on CSIs provides evidence that the

genetic changes responsible for such indels are necessary for the proper functioning of the proteins in the CSI-containing organisms, and the removal or alteration of the indels lead to cell death (Singh & Gupta, 2009).

In addition to conserved signature indels, comparative genomic studies have identified another class of genome sequence derived molecular signature which are also very useful for evolutionary and biochemical studies. These markers consist of whole proteins, referred to as conserved signature proteins (CSPs), which are present in specific monophyletic clades of microorganism. The genes for these CSPs originated in a common ancestor of the given group of organism and subsequently vertically passed down to its various descendants (Daubin et al., 2003; Lerat et al., 2005; Gao et al., 2006; Gao and Gupta, 2012a; Gupta, 2010; Ho et al., 2016). Overall, like CSIs, CSPs are also valuable molecular signatures for the evolutionary and biochemical studies of different bacterial clades (Gao and Gupta, 2012b; Dutilh et al., 2008). Similar to the CSIs, the CSPs are also predicted to be involved in maintaining the unique attributes of different groups of bacteria. (Danchin, 1999; Roberts, 2004; Fang et al., 2005; Fang et al., 2008; Lorenzini et al., 2010; Chandra and Chater, 2014; Gupta, 2016).

Our lab has identified a large number of CSIs and CSPs for the different groups of microorganisms, specially for the genus *Deinococcus* and the order *Chlamydiales* (Gupta et al., 2015; Ho et al., 2016). Earlier work from our lab has established the importance of CSIs and CSPs in understanding evolutionary relationships among microorganisms and also as novel diagnostic tools However, the functional roles of most of the identitied

CSIs and CSPs are presently not known (Bhandari and Gupta, 2014; Adeolu et al., 2016; Wong et al., 2014; Naushad et al., 2014). Structural analysis of CSIs delineates that most of the CSIs are found to be located at the surface loop of the protein (Alnajar et al., 2017; Khadka and Gupta, 2017). The surface loop regions of proteins are likely involved in mediating protein-protein and protein-ligand interactions (Akiva et al., 2008). If we can predict any protein-protein interactions between the CSIs-containing proteins and any other protein(s) found in the CSIs-containing bacteria, it can prove helpful in understanding the functional role of CSIs. Similarly, it may also be possible to predict the functional role of CSPs by predicting their association with other known functional proteins.

The Deinococcus-Thermus group of bacteria is well established as a distinct phylum, which contains two extremophilic groups of organisms: *Deinococcales* and *Thermales*. This heavily researched phylum has been classified entirely by their branching pattern in the 16S rRNA tree. *Deinococcales* has been divided into 3 distinct genera (viz. *Deinococcus, Deinobacterium,* and *Truepera*). However, most of the species within this order belong to the genus *Deinococcus* and they form a monophyletic group in phylogenetic trees based on 16S rRNA-and other genes/proteins sequences (Omelchenko et al., 2005; Eisen, 1995; Weisburg et al., 1989; Ho et al., 2016b; Eisen, 1995; Makarova et al., 2001b).

The genus *Deinococcus* has distinct properties which differentiate it from other bacteria in the phylum Deinococcus-Thermus (Slade and Radman, 2011). *Deinococcus* spp. are

more than 100-fold resistance to ionizing radiation and more than 20-fold resistance to UV radiation as compared to *Escherichia coli* (Slade et al., 2009). The *Deinococcus* spp. are highly resistant to DNA damage-inducing agents. Ionizing radiation and desiccation are responsible for double strands-breaks (DSBs), single strands-breaks (SSBs), and base damage; UV radiation generates diverse pyrimidine dimers; mitomycin C (MMC) forms DNA interstrand cross-links; and hydrogen peroxide, methyl methane sulfonate (MMS), N-methyl-N-nitro-N-nitrosoguanidine (MNNG), nitrous acid, and hydroxylamine are responsible for numerous base and nucleotide damage (Makarova et al., 2007a; Slade and Radman, 2011; Timmins and Moe, 2016). The following mechanism compensates these diversified DNA damages: direct damage reversal (Slade and Radman, 2011), base and nucleotide excision repair (Agostini et al., 1996; Mun et al., 1994; Sandigursky and Franklin, 1999), mismatch repair (Rayssiguier et al., 1989), and recombinational repair (Carroll, 1996) . DSBs are the most severe form of DNA damage. These types of DNA damages are resolved by two mechanisms: "extended synthesis-dependent strand annealing" (ESDSA) and homologous recombination by crossovers (Zahradka et al., 2006). In *Deinococcus radiodurans*, ionizing radiation induces 10 times more SSBs and many more sites of base damage than DSBs. Extremely efficient exinuclease repair systems are able to repair the base and nucleotide damage with the broad spectrum of substrate specificity (Makarova et al., 2001). According to introduction of chapter 2, very few characteristics are known to be uniquely shared by species of Deinococcus-Thermus and distinctly present in genus *Deinococcus* to explain the radiation, oxidation, and desiccation resistance (Narumi et al., 2004; Tanaka et al., 2004; Sugiman-Marangos et

al., 2016; Xu et al., 2010). Our lab has identified 24 CSIs and 29 CSPs which provide distinguishing characteristics of the phylum "Deinococcus-Thermus phylum" and its main groups. Additional 12 CSIs and 61 CSPs are the molecular characteristics of the genus *Deinococcus,* which provide novel and highly specific means for distinguishing this genus from other groups of bacteria of the Deinococcus-Thermus phylum (Griffiths and Gupta, 2004; Griffiths and Gupta, 2007; Ho et al., 2016). But we do not know the functional role of these CSIs and CSPs. In-silico protein-protein interactions analyses may lead to predict the functional role of these CSIs and CSPs to understand the high tolerance of *Deinococcus* spp. to radiation and other form of DNA damages.

The members of the phylum chlamydiae are Gram-negative obligate intracellular bacteria that infect a broad range of mammalian hosts (Cram et al., 2017). All the bacteria within the phylum Chlamydiae are present in a single class, *Chlamydiia* (Everett et al., 1999; Rurangirwa et al., 1999; Kalayoglu et al., 2001). Our earlier work proposed a division of the class *Chlamydiia* into two orders: *Chlamydiales and Parachlamydiales*. The order *Chlamydiales* is proposed to consist of two families: *Chlamydiaceae* and the closely related *Candidatus* family *Clavichlamydiaceae*. The novel order *Parachlamydiales* is proposed to contains *Parachlamydiaceae*, *Simkaniaceae* and *Waddliaceae* and the *Candidatus* families *Criblamydiaceae, Parilichlamydiaceae, Piscichlamydiaceae.* This was proposed based on the unique molecular characteristics, phylogenetic relationship, and the genetic distance between two clusters within the phylum Chlamydiae (Gupta and Griffiths, 2006; Gupta et al., 2015). The members of all of these families are potential pathogens of human and animals (Corsaro and Greub, 2006). Chlamydial spp. are the

cause for the bacterial sexually transmitted infection, trachoma, community-acquired pneumonia and also risk factor of cardiovascular disease atherosclerotic, cerebrovascular, and chronic lung disease (Gerbase et al., 1998; Schachter, 1978; Kuo et al., 1995; Kuo et al., 1993; Saikku et al., 1988; Balin et al., 1998; Hahn et al., 1991). Pathogenic chlamydial isolates have been identified from a number of animal hosts including birds, cats, rodents, cattle and pigs (Meijer et al., 2002). Though chlamydial spp. have diversified host range and disease pathogenesis, the biphasic intracellular development cycle is the major unifying feature of this group (Moulder, 1991; Hybiske and Stephens, 2007). The phylum Chlamydiae possesses two morphological forms during intracellular development: elementary body (EB) and reticulate body (RB) (Moulder, 1991). The term "elementary body" refers to the small, electron dense, metabolically inert, infectious form of the organism (Matsumoto, 1973). The bacterial histone-like proteins HtcA and HtcB play role in the condensation of DNA during the EB form of chlamydial development cycle (Brickman et al., 1993; Barry, III et al., 1992). EB initiates the infection by their ability to attach and invade susceptible cells. Being the obligate intracellular pathogen, chlamydial spp. depend on a number of effector proteins to destabilize the host cellular pathways, which facilitates the bacterial survival and replications (Valdivia, 2008; Betts et al., 2009). Most of these effector proteins are exported by using type III secretion system (T3SS). T3SS is a complex molecular machine, which is present in Gram-negative bacteria. It works as an injection syringe and dispatches effector proteins and toxins into host cell cytoplasm (Galan and Wolf-Watz, 2006; Valdivia, 2008). A number of cytoplasmic chaperones are involved in the detection, recruitment and secretion of

effector proteins by T3SS (Betts-Hampikian and Fields, 2010). T3SS is located in the inner membrane and it has homology to the flagellar basal body of Gram-negative and Gram-positive bacteria (Kubori et al., 1998). Most of the proteins of T3SS form clusters in bacterial genomes. These clusters are spread throughout the genome in the case of Chlamydiae species (Plano et al., 2001). Several studies have revealed that blocking of T3SS activity leads to the inhibition of chlamydial intracellular development (Duncan et al., 2012; Marshall and Finlay, 2014). The most important stage of early infection is the conversion of EB into the metabolically active form, termed reticulate body (RB) by 6 to 8-hour postinfection (HPI). The RB replicates repeatedly by binary fission during midcycle and later undergoes conversion into EBs which are released to infect neighboring cells (Abdelrahman and Belland, 2005; Wyrick, 2010; Wyrick, 2010). A recent study has indicated that DNA supercoiling is essential for the temporal regulation of genes during intracellular development (Orillard and Tan, 2016). Topoisomerase II protein expression is up-regulated during negative supercoiling of DNA and it is a unique characteristic of the Chlamydiae species (Orillard and Tan, 2016). However, very few studies have been carried out to find out unique molecular characteristics of the Chlamydiae species that could provide insights into their unique and complex developmental or cell division cycle. Earlier work from our lab has identified 32 CSIs and 98 CSPs that are uniquely shared by members of the phylum Chlamydiae, providing novel means for their diagnostics as well as for investigating their novel functional characteristics (Gupta et al., 2015). However, the cellular functions of these CSIs and CSPs are presently not known. Further analysis of these identified CSIs and CSPs

through in-silico protein-protein interactions could provide novel insights in understanding novel functionals aspects of these bacteria.

1.2 Protein-protein interactions based analyses to understand the functional role of proteins:

Proteins in biological systems rarely function alone; they need to collaborate with other proteins in complex interconnected systems referred to as protein complex networks. These interconnected networks contain a variety of interactions and associations among the proteins known as protein-protein interactions (PPIs) (Kuchaiev et al., 2011). PPIs are responsible for cell to cell interactions, metabolic interactions, signalling enzymes and development control.

In a cellular system, proteins generally perform their function as part of a protein complex. Each of the protein complex is responsible for specific tasks in a cell (Pereira-Leal et al., 2004). Thus, in network systems biology, proteins which interact with each other are thought to be involved in the same biological process as members of a protein complex or functional module (Tornow and Mewes, 2003). In network systems biology, protein complexes are a group of proteins which bind or interact with each other at a certain place and time (ex. Replication complex). Functional modules, on the other hand, perform related biological functions but they do not have restrictions on the place or timing of the functions (ex. transcription factors). Understanding these basic properties (Protein complexes, functional modules, clustering, etc.) of PPIs can provide important insights regarding the functioning of a protein within a cell (Jancura et al., 2012; Laraia et al., 2015).

1.3 Methodology for detection of protein-protein interactions (PPIs):

Protein-protein interactions (PPIs) can be detected by utilizing both wet lab and dry lab research approaches. In the experimental or wet lab approach, PPIs can be identified by a variety of methods including tandem affinity purification, affinity chromatography, protein arrays, protein fragment complementation, phage display, yeast two-hybrid and synthetic lethality in addition to a number of other methodologies reviewed recently (Rao et al., 2014). In the dry lab or *in silico* analysis, the following approaches are used to identify PPIs: gene neighborhood, gene fusion, phylogenetic analysis, gene ontology, phylogenetic profile and gene co-expression. Information regarding PPIs are publically available through a number of different web servers and databases, which include both curated experimental PPI data and PPIs networks based on both wet lab and dry lab approaches such as BioGrid, STRING, DIP, MINT, HPID, IntAct and BIND etc. (Rao et al., 2014; Hermjakob et al., 2004). In my research, we have used PPIs data from the STRING database (Szklarczyk et al., 2017). STRING is a widely used, comprehensive database for PPIs. The STRING database contains data regarding the interaction between two proteins based on both experimental data as well as predicted interaction information (Mering et al., 2005). The STRING database has some advantage in comparison with other PPI databases: 1) It also extracts data from other databases 2) It provides a confidence score for each predicted association based on the probability of finding the linked proteins within the same KEGG pathway (Szklarczyk et al., 2015). 3) Version 10.0 of the STRING database, it's most recent iteration, has PPI data for more than 9.6 million proteins, which is significantly larger than the next largest "IntAct database", which

contains information for 91 thousand proteins (http://string-db.org/; http://www.ebi.ac.uk/intact/). 4) Lastly, the STRING database accounts for phylogenetic profiling in its use of co-expression data providing additional insight regarding the likelihood of spurious co-expression (Szklarczyk et al., 2015).

1.4 Research Objectives

Our lab has identified large numbers of CSIs and CSPs for different groups of microorganisms, specially for the genus *Deinococcus* and the phylum Chlamydiae (Gupta et al., 2015; Ho et al., 2016). These two groups of bacteria have distinct properties which differentiate them from other groups of bacteria (Orillard and Tan, 2016; Slade and Radman, 2011). So, my research objective is to gain some insights into the functional roles of the CSIs and CSPs specific for the genus *Deinococcus* and the phylum Chlamydiae using in-silico protein-protein interactions analyses.

The *Deinococcus* spp. are present in the well-established distinct Deinococcus-Thermus phylum which have been characterized by extraordinary tolerance to oxidative stress induced by ionizing radiation (IR), ultraviolet radiation (UV), desiccation and chemical mutagens (Slade and Radman, 2011a). But, very few characteristics are known that are commonly shared by both groups of bacteria and uniquely present in each group of bacteria. Our earlier work has identified a large number molecular signatures in the form of CSIs and CSPs which clearly distinguish the *Deinococcus* spp. (Ho et al., 2016; Griffiths and Gupta, 2007). *Deinococcus* spp. have more efficient DNA repair pathways in comparison to the other groups of Bacteria (Slade et al., 2009). In my thesis, I have

carried out comparative genomics analyses of DNA repair proteins to find out CSIs specific for *Deinococcu*s spp. Furthermore, we carried out structural and in-silico protein-protein interactions analyses of these CSIs to understand their role in oxidative damage protection among *Deinococcus* spp. We have also identified CSPs that are genetically linked to the proteins involved in DNA repair processes.

Similarly, Chlamydial spp. are characterized by a unique intracellular development cycle (Batteiger et al., 2014; Kohlhoff et al., 2014). DNA supercoiling levels are proposed to regulate the temporal expression of chlamydial gene during intracellular development. The promoter of *gyrBA* of chlamydial spp. has been upregulated during negative supercoiling which is distinctive molecular characteristics of chlamydial spp. (Orillard and Tan, 2016). But very few studies have been carried out to understand this unique feature of DNA topoisomerases in Chlamydiae species. In this study, I have carried out comparative genomics analyses to find out unique sequences in the GyrA and GyrB proteins which are specific for the Chlamydiae spp. We have identified several CSIs in the GyrA and GyrB proteins which are unique molecular characteristics of the Chlamydiae spp. I have done structural and in-silico protein-protein interactions analyses on GyrA and GyrB proteins to decipher the functional roles of CSIs present in these proteins.

CHAPTER 2

The utilization of molecular signatures to understand the radiation resistance of *Deinococcus* spp.

2.1 Introduction:

The *Deinococcus* spp. which are part of the phylum Deinococcus-Thermus (Ho et al., 2016; Garrity et al.,2001; Tanaka et al., 2004; de Groot et al., 2009) are characterized by their extraordinary tolerance to ionizing radiation (IR), ultraviolet radiation (UV), desiccation and chemical mutagens. *Deinococcus* spp. are promising research targets for use in the bioremediation of radiation-contaminated sites (Mattimore and Battista, 1996; Battista, 1997; Tanaka et al., 2004; Omelchenko et al., 2005; Makarova et al., 2001; Misra et al., 2006). The *Deinococcus* spp. have efficient antioxidant defense system, contain extensive abundance of genes encoding catabolic enzymes including catalases, peroxidase and superoxide dismutase which play role in removal of H_2O_2 and superoxide radicals from the cells (Markillie, Hradecky, & Wong 1999; Slade and Radman, 2011; Schmier et al., 2017). But all the *Deinococcus* spp. do not encode all these proteins in their genome. *Deinococcus* spp. also contain exceptionally high amounts of manganese (Mn^{2+}) complexes, play role to protect proteins from IR-induced oxidative damage. Increase presence of manganese (Mn^{2+}) complexes has also been observed in some other IR resistant bacteria such as *Lactobacillus plantarum* and *Synechocystis* sp.PCC 6803 (Daly, 2009; Daly, 2012).

Previous studies have revealed that *Deinococcus* spp. possess classical prokaryotic DNA repair machinery consisting of the base excision repair (BER), the nucleotide excision repair (NER), the mismatch repair (MMR) and double-strand break (DSB) repair pathways (Timmins and Moe, 2016; Daly, 2009; Slade and Radman, 2011). Most of the

proteins of these pathways are ubiquitous and conserved among other group of bacteria which suggest that DNA repair machinery of *Deinococcus* spp. is either not unique or the novel aspects of DNA repair mechanism have not yet been identified (Slade and Radman, 2011; Shen et al., 2007). However, comparative genomics analyses of their genome indicate that they have efficient genome reconstitution system after exposure to radiation in compare with radio-sensitive bacteria (Slade et al., 2009). The *Deinococcus* spp. encode abundant amount of DNA repair enzymes: a large number of DNA glycosylases involved in BER, two Dps proteins that play the role to DNA protection, two single-stranded DNA binding (SSB) proteins (SSB and DdrB) that contribute to ssDNA intermediates protection and many hypothetical proteins, some of which may be related to DNA repair (Slade and Radman, 2011; Timmins and Moe, 2016). *Deinococcus* spp. also encode a large number of proteins in their genomes which are only specific for this group of bacteria (de Groot et al., 2009). Some of them (e.g. DdrA, DdrC, DdrD, IrrE, PprA) are highly expressed in radiation or desiccation exposure and play role in DNA damage response (Daly, 2009; Vujicic-Zagar et al., 2009; Tanaka et al., 2004; Makarova et al., 2007). Deletion of DdrA and DdrD proteins in combination with PprA leads to 100- and 1000-fold more sensitivity to the lethal effects of UV-induced damage (Selvam et al, 2013).

The *Deionococcus* spp. have unique DNA repair mechanism termed extended synthesis dependent strand annealing (ESDSA). In this repair mechanism, overlapping genomic fragments are first resected in a 5' to 3' fashion to generate 3' single strand DNA tails, then RecA protein converts the long linear repair intermediates into circular genome

replicons. In this way, ESDSA with the help of RecA-dependent homologous recombination (HR) play a role in *Deinococcus* radiation resistance (Slade et al., 2009; Sommer, Bentchikou, & Servant, 2010). In addition, extremely efficient exinuclease repair systems are able to repair the base and nucleotide damage with the broad spectrum of substrate specificity (Slade and Radman, 2011). This group of bacteria possesses two variants of the UvrA protein (UvrA1 and UvrA2) involved in the NER pathway (Tanaka et al., 2005). The UvrA1 protein plays a key role in DNA damage recognition (Jaciuk et al., 2011). This repair process is carried out by UvrB, UvrC, UvrD and completed by DNA polymerase I and Ligase enzyme (Jaciuk et al., 2011; Pakotiprapha et al., 2012). But, the functional role of UvrA2 is not clearly known (Shen et al., 2007). This protein does not possess UvrB interaction domain. But, recent studies indicate that UvrA2 plays a role in DNA repair and resistance to DNA damage inducing agents such as UV and bulky chemicals (Timmins et al., 2009). The *Deinococcus* spp. also have UV damage endonuclease (UvsE)-dependent excision repair (UVER) pathway. Inactivation of UvrA1 and UvsE proteins reduces the ability to remove UV-induced damage from the genomic DNA (Tanaka et al., 2005).

Exposure of *Deinococcus radiodurans* and *Escherichia coli* to ionizing radiation, UV light, dessication and toxic chemicals causes similar amount of DNA damage in both of the species but the amount of protein damage is significantly higher in *Escherichia coli* than *Deinococcus radiodurans* (Krisko and Radman, 2010). Overall the *Deinococcus* spp. show remarkable robustness to radiation due to their efficient proteome protection and cell death by radiation is caused by oxidative damage with resultant loss of maintenance

activities including DNA repair (Krisko and Radman, 2010). Ultimately, the fate of irradiated cells depends on their ability to genome reconstitution; most studies have focused on how damaged DNA is repaired after exposed to radiation (Daly, 2009). However, only few studies have been performed targeting the novel molecular attributes of proteins involved in DNA repair to understand their mechanism of protection against oxidative damage.

We have identified a large number of conserved signature indels (CSIs) in different proteins involved in different cellular functions that are distinctive molecular characteristics of the *Deinococcus* homologs (Ho et al., 2016; Griffiths and Gupta, 2007). CSIs are the special type of signature sequence of that protein which is unique for certain group of bacteria (Gupta, 2014). In addition, we have identified numerous *Deinococcus*-specific proteins in terms of conserved signature proteins (CSPs) (Griffiths and Gupta, 2007; Ho et al., 2016). Among these CSPs few of them such as DdrB, PprA have already been characterized as involved in DNA repair (Ho et al., 2016; Kota et al., 2016; Makarova et al., 2007).

The present work focuses on the sequence features of the proteins involved in DNA repair pathways to identify any characteristics that could prove helpful in understanding the exceptionally high tolerance of *Deinococcus* spp. to oxidative damage. We have identified highly specific molecular signatures in the form of CSIs that clearly distinguish the DNA repair proteins of *Deinococcus* from protein homologs of other bacteria. We have identified CSIs in the three important proteins (UvrA, UvsE and RecA) of NER,

UVER and HR DNA repair pathways. The result of protein modeling clearly reveals that the identified CSIs which are the distinct characteristic of *Deinococcus* homologs are present in the protein surface. The results of protein-protein interactions analyses of these proteins presented here delineate that UvrA1 and UvsE proteins are present in unique genetic linkage with few other proteins involved in protein disulfide bond formation proteins and dCSP-I (Accession: NP_295493), and Zn-ribbon protein. Simultaneously RecA protein is present in genetic linkage with two other proteins which are specific for Deinococcus-Thermus phylum. The importance of these observations regarding the high resistance of *Deinococcus* spp. to radiation is discussed.

2.2 Methods:

2.2.1 Protein-protein interactions (PPIs) analyses to examine the unique association of *Deinococcus* specific DNA repair proteins:

The STRING database (Szklarczyk et al., 2017) has been used to analyze proteins which show unique association with the DNA repair proteins in *Deinococcus* spp. The STRING is a widely used, comprehensive database for PPIs. The STRING database contains data regarding PPIs based on both experimental data as well as predicted interaction information. It predicts interactions based on co-occurrence, gene fusion, co-expression and neighborhood (Mering et al., 2005; Szklarczyk et al., 2015).

2.2.2 Identification of genetic linkage of *Deinococcus* DNA repair proteins:

The *Deinococcus* spp. genomes have examined to identify genetic linkage of DNA repair proteins. tBLASTn search was carried out for DNA repair proteins in the genome of *Deinococcus radiodurans* R1 against all available sequences in the GenBank nucleotide collection (Altschul et al., 1997; White et al., 2004). We looked at the position of the proteins in the graphics view of the genome. It also facilitates to find out the neighborhood proteins of this protein. Intergenic distance between two proteins was calculated manually based on their position in the genome. The direction of the genes in the genome had been determined by visual inspection.

2.2.3 Structural analysis of the CSIs and homology modeling of the DNA repair protein homologs from *Deinococcus* spp.:

Three-dimensional structure model of the DNA repair proteins homologs of *Deinococcus* spp. was developed from their full-length sequence using the homology modeling technique (Sali & Blundell, 1993) . The suitable template for homology modeling was identified by using PSI-Blast search (Altschul et al., 1997) against the Protein Data Bank (PDB) (Bi et al., 2015). Top hits obtained by a PSI-Blast search for the proteins of *Deinococcus* spp. and templates showed sequence identity in the range of 40-63%. The template was selected based on sequence identity and took into consideration other factors such as the probability score, expect value (E-value), analysis of the quality of sequence alignment, and the secondary structure compatibility between target and templates. Finally, the template was selected from crystallized proteins of different

bacterial spp. Initially, 200 models were generated by using modeller v9.14 (Sali & Blundell, 1993) and ranked by using assigned discrete optimized potential scores (DOPE) (Shen & Sali, 2006). Then modrefiner was used to refining the selected models (Xu & Zhang, 2011). Validation of the models of proteins was conducted by using four independent servers: PROSA (Wiederstein & Sippl, 2007; Sippl, 1993), RAMPAGE (Lovell et al., 2003), ERRAT (Colovos & Yeates, 1993), Verify3D (Eisenberg et al., 1993) and QMEAN (Benkert et al.,2008). Then the three-dimensional structure of CSI containing region was examined and compared with the secondary structure of that region to ensure the reliability.

2.3 Results:

2.3.1 Distinguishing characteristics in the DNA repair proteins sequences from *Deinococcus* spp.:

The *Deinococcus* spp. are more than 100-fold resistant to ionizing radiation and 20-fold resistant to UV-radiation as compared to *E. coli*. The UvrA1 and UvsE proteins are the central protein of NER and UVER DNA repair pathways which are responsible for UV-induced damage repair (Tanaka et al., 2005). A multiple sequence alignment of representative DNA repair proteins homologs of NER and UVER pathways from different group of bacteria was carried out to gain insight into the molecular basis of the difference in the UV-radiation and mitomycin C resistance of *Deinococcus* spp. versus other groups of bacteria. Inspection of the sequence alignment has identified a number of CSIs found in the DNA repair proteins that are distinctive characteristics of *Deinococcus*

spp. The first of these CSIs is a 30 aa insertion in the UvrA protein that is commonly shared by all *Deinococcus* homologs (all available homologs without exceptions) (Figure 1A). Secondly, we have identified a 15 aa insertion in the UvrC protein that is also specific for all *Deinococcus* homologs (Figure 1B). Additionally, we have also identified a 6 aa insertion in the UvrD protein that is unique trait of the members of the genus *Deinococcus* (Figure 2A). *Deinococcus* spp. also contain a 1 aa insertion in UvsE protein (all available homologs with few exceptions) (Figure 2B). We have identified 1 aa insertion in RecA protein which is specific for Deionococcus-Thermus phylum (Figure 10). We have also identified a large number of CSIs in other DNA repair proteins (Table 1 and Appendix). These molecular traits are only found in all homologs of these proteins from the member of the *Deinococcus* group of bacteria and absent in all homologs from other groups of bacteria. It provides strong suggestive evidence that these CSIs represent synapomorphic characteristics of the member of *Deinococcus* group, and likely occurred in a common ancestor of the group.

2.3.2 Locations of the CSIs in the structure of the proteins:

In our research, we have carried out the structural analysis of CSIs which are present in the proteins of NER and UVER pathways. Three dimensional structures of UvrD and RecA proteins are available from *Deinococcus* spp. Three-dimensional structure models of other proteins have been developed by homology modelling approach. The structure of these proteins has been visualized by using PyMOL software. The locations of different CSIs in the crystal or model structures of proteins have been mapped and surface

representation of these CSIs in the structure is exhibited in figure (Figure 3). The 6 aa CSI of UvrD protein is located at the helix region of the proteins and surface analysis reveals that this CSI is present on the protein surface. The 1 aa in RecA protein is also located at surface loop (Figure 10). The 30 aa insertion in UvrA1 protein is located at the surface loop in the protein model. Simultaneously, the two other CSIs are also present at the surface loop of protein models. Secondary structure of the CSI region has been predicted by using PSI-PRED server. It has been marked in the figure by purple color (Figure 3).

2.3.3 Unique interactions of DNA repair proteins ubiquitously present in the *Deinococcus* spp.:

Surface loops in proteins are generally involved in mediating novel protein-protein interactions and protein-ligand interactions (Akiva, Itzhaki, & Margalit, 2008). We have analysed our CSI containing proteins to identify unique interactions with other proteins. Unique interactions of CSI-containing bacterial proteins were predicted by analyzing the PPIs data from the STRING database.

Firstly, the UvrA1 protein has an unique interaction with dCSP-I protein (accession ID: NP_295493) among all five available *Deinococcus* spp. in the STRING database as well as with DsbA and DsbB proteins among 4 out of 5 representative species. The STRING database has predicted these associations based on the conserved genomic neighborhood context. Neighborhood analysis for all available genome sequenced *Deinococcus* spp. could lead a deeper insight to predict the functional association of these four proteins. We

have manually inspected the genomic data for available species to find out the neighborhood proteins of UvrA1 protein. Neighborhood analysis reveals that 16 out of 18 CSI containing *Deinococcus spp*. have identically conserved neighborhood patterns in their genome (Figure 4). If two genes coexist in a certain group of microorganisms then these genes should have the same phylogenetic profile. These four proteins should also have same phylogenetic profile. According to Khan I. *et al* (2014), similar phylogenetic profiling proteins should have physical or functional interactions among them (Khan et al., 2014). Overall, neighborhood and phylogenetic profile analyses indicate that DsbA, DsbB and the dCSP-I may carry out UvrA1 protein–related functions in *Deinococcus* spp. Additionally, UvrA is present in genetic linkage with another CSP (dCSP-II) which is only specific for *Deinococcus radiodurans* and *Deinococcus wulumuquiensis.*

In prokaryotic organisms, around 60% of genes are present in the polycistronic operon. An operon consists of a cluster of genes arranged in tandem on the same strand of a genome sharing common promoter and terminator. Genes which are present in the same operon are also functionally related (Dam, Olman, Harris, Su, & Xu, 2007; Moreno-hagelsieb, 2015). So, analysis of UvrA1, DsbA, DsbB and dCSP-I proteins based on the frequency of co-occurrence in the same operon could play the pivotal role to predict the function of a 30 amino acid CSI of UvrA1 and dCSP-I. The DOOR 2.0 database contains computationally predicted operons of the prokaryotic genomes. Operonic data for six *Deinococcus* spp. are available in DOOR 2.0. UvrA1, dCSP-I, DsbA, and DsbB proteins are present in the same operon among 3 out 6 species. UvrA1 and dCSP-I, as well as DsbA and DsbB, proteins are present in two different operons in two other *Deinococcus*

spp. But these two operons are adjacent to each other, and the intergenic distance among them is less than 200 base pairs. These are predicted operons; it could be possible that these four proteins are actually present in the same operon among most of the *Deinococcus* spp. (F. Mao, Dam, Chou, Olman, & Xu, 2009; X. Mao et al., 2014; Moreno-hagelsieb, 2015). In *Deinococcus radiodurans*, dCSP-I and UvrA1 proteins are present in different operons. Overall data suggests that these four proteins should have functional relation among themselves.

Similarly, genetic neighborhood analysis reveals that The UvsE protein is also present in genetic linkage with Zn-ribbon (Zn-R) protein among *Deinococcus* spp. (Figure 6). Predicted operonic data from DOOR.2 database suggests that UvsE and Zn-R proteins are part of the same operon among *Deinococcus* spp.

The RecA protein is also present in a unique genetic linkage with 2'-5' RNA ligase (LigT) and competence inducible protein A (CinA) among Deinococcus-Thermus spp. (Figure 11), and also predicted to be present in the same operon.

2.3.4 Structural analyses of these DNA repair proteins:

Interestingly, we have identified CXXC motifs in the zinc finger motifs of UvrA1 protein using comparative genomics studies. Most of CXXC motifs are conserved through all bacterial species. Structural analysis delineates that these CXXC motifs are located at the surface, and two CXXC motifs of zinc finger III are located close proximity to DNA binding region (Figure 7). Substitution of one of the cysteines of CXXC motifs leads to inactivation of UvrA1 protein (Wang et al., 1994). We also have identified a large

number of CXXC motifs in other different DNA repair proteins of *Deinococcus* spp. (Table 2 and Appendix).

The Secondary Structure and membrane topology of dCSP-I have been predicted by PSIPRED. dCSP-I protein contains five alpha helical regions. It has only one beta strand in its secondary structure (Figure 5). Membrane topology of this protein indicates that it is a transmembrane protein. It has five transmembrane helices.

Comparative genomics analyses have confirmed that UvsE protein contains cysteine residues in the sequences. Structural analysis reveals that two of these cysteine residues are exposed on the surface (Figure 9). The Secondary structure and membrane topology analysis suggest that Zn-ribbon protein is a membrane associated protein (Figure 6).

2.4 Discussion:

The *Deinococcus* spp. differ from other groups of bacteria for their highly resistance pattern to different types of radiations, chemicals, and desiccation. These environmental and chemical mutagens are responsible for double strands-breaks (DSBs), single strands-breaks (SSBs), base and nucleotide damage (Slade & Radman, 2011). The proteins involved different DNA damage repair pathways are pivotal research targets to understand the resistance mechanism of *Deinococcus* spp. Analyses of DNA repair proteins sequences from different organisms carried out in this work have brought important insights in this regard. UV-radiation is substantially responsible for direct and indirect damage to DNA and proteins. UVC radiation causes oxidative damage to proteins through generating reactive species resulting in protein thiol oxidation of

cysteine residues (Chan et al., 2006; Daly, 2009). It has been expected that *Deinococcus* spp. can increase the survival of cell by utilizing their efficient antioxidants system and clear damage proteins using an expanded family of subtilisin-like proteases (Makarova et al., 2001; Ghosal et al., 2005). This work has carried out comparative analyses of DNA repair proteins sequences and identified a high-specific sequence feature in the form of CSI in the protein sequences that clearly differentiate the DNA repair proteins homologs of *Deinococcus* spp. from the DNA repair proteins homologs found in most other organisms.

The findings from this study indicate that most of DNA repairs proteins homologs from *Deinococcus* spp. differ from all other DNA repair proteins homologs by a highly-conserved sequence feature. We have identified three large CSIs in UvrA, UvrC and UvrD proteins. The Crystal structure of UvrD protein from *Deinococcus* spp. is available, Structural analysis reveals that a 6 aa CSI in UvrD protein plays a role to stabilize the binding of adjacent loop with the ssDNA (Stelter et al, 2013). We also have identified a 1 aa CSI in RecA protein which is specific for Deinococcus-thermus phylum. Structural analysis delineates that other CSIs are present in the surface loop of protein. Earlier work on conserved indels provides evidence that the genetic changes responsible for such indels are crucial for the proper functioning of the proteins in the CSI-containing organisms, and the removal of such CSIs has the adverse effect on the proper functioning of the concerned proteins (Singh & Gupta, 2009). The surface loops in protein sequences are considered as determinants of interaction because they constitute highly accessible regions of the protein and they are known to play essential roles in mediating protein-

protein and protein-ligand interactions (Akiva et al., 2008). So, these studies strongly suggest that the described sequence characteristics of DNA repair proteins should play an important role in mediating protein-protein or protein-ligand interactions with other proteins and responsible for making the observed differences in the biochemical characteristics of *Deinococcus* spp.

To understand the functional role of CSIs, we have carried out protein-protein interactions analysis of the CSI containing proteins based on in-silico PPI approaches. We have identified unique interactions in three main DNA repair proteins (RecA, UvsE and UvrA). RecA protein is present in a genetic linkage with CinA and LigT proteins among Deinococcus-Thermus spp. UvrA and UvsE proteins are interacting with proteins which contain functionally active cysteine residues. The UvsE protein has an unique genetic linkage with Zn-R protein among *Deinococcus* spp. Zn-R protein contains a zinc binding motif in their sequences which is very common molecular characteristics among DNA repair proteins. This protein also contains a CXXC motif which has oxidoreductase activity. Secondary structure analysis reveals that this protein is also a transmembrane protein (Fig. 6). The UvsE protein also contains a large number of cysteine residues in their sequences. So, Zn-R protein may play a role to maintain stability of UvsE proteins during oxidative damage. The mechanism has been explained detail in UvrA1 PPI analyses.

The protein-protein interactions analyses completed have led to concur that the UvrA1 protein is present in genetic linkage with dCSP-I, DsbA and DsbB Proteins among

Deinococcus spp.. We have explained the details of these interactions to understand the functional role of CSI, and dCSP-I and also to understand the resistance pattern of *Deinococcus* spp. Finally we have proposed a pathway to understand the functional role of these four proteins in together. This genetic linkage clearly differentiates the UvrA1 homologs of *Deinococcus* spp. from the UvrA1 homologs found in most other organisms in terms of interactions. These four proteins may be present in the same operon among *Deinococcus* spp. The DsbA and DsbB are membrane associated disulfide bond formation protein (Chim, Harmston, Guzman, & Goulding, 2013; Inaba et al., 2006; Lin, Kovalsky, & Grossman, 1997; Makarova et al., 2007). The dCSP-I is a hypothetical protein with unknown function which is only found in *Deinococcus* spp. Secondary structure and membrane topology analyses by PSIPRED server (Buchan et al., 2013; Nugent & Jones, 2009) suggest that dCSP-I is a transmembrane protein (Figure 5). The intergenic distance between dCSP-I and UvrA1 is around -4 to 200 bp in most of *Deinococcus* spp. These types of genetic linkage between UvrA1 and dCSP-I and the highly-specific CSI are unique genomic and molecular characteristics for *Deinococcus* spp. This genetic linkage strongly suggests that UvrA1 protein should have interaction with dCSP-I and it may be mediated by this large CSI in UvrA1.

The Zinc-binding motifs of UvrA1 contain CXXC motifs, and play a critical role in DNA damage detection (Navaratnams, Myless, Strangell, & Sancaren, 1989; Wagner, Moolenaar, & Goosen, 2011; Wang & Mueller, 1994). Cysteines of the CXXC motif form intermolecular disulfide bonds (Shouldice, Walden, Totsika, Schembri, & Martin, 2011). Exposure of the UvrA1 protein to UVC radiation may lead to form a disulfide

bond between cysteine residues through protein thiols oxidation and inactivate the UvrA1 protein (Slade & Radman, 2011). The UvrA1 protein of *Deinococcus* can be fully complemented by the UvrA1 protein of *Escherichia coli.* In that paper, *Deinococcus radiodurans* uvrA-defective strains that contain the Escherichia coli $uvrA^+$ gene have been treated with mitomycin (MM), 4-nitroquinoline-1-oxide (4NQO), and Me_3porsalen plus near-UV exposure to understand their survival. These strains displayed survival to 4NQO identical with that of wild-type *Deinococcus radiodurans*. But they had clearly mentioned that the uvrA of *Escherichia coli* was failed to fully complement uvrA *Deinococcus radiodurans* with respect to survival to MM and Me_3porsalen plus near-UV. But, this work reports that $mtcA^+$ and $mtcB^+$ genes are a single gene and is functionally similar to the UvrA protein of *Escherichia coli*(Agostini et al., 1996). Additionally, we have identified a 30 aa CSI in the UvrA protein of *Deinococcus* spp. Earlier work on CSIs have revealed that CSIs are important for proper functioning of proteins. Complementation of a 1 aa CSI in the proteins leads to inactivation of protein (Singh and Gupta, 2009). So, this 30 aa CSI in UvrA should have functional significance which will be specific for the *Deinococcus* spp. Overall, these results lead to question about the fully complementation of UvrA of *Deinococcus radiodurans* by UvrA protein of *E. coli*. UvrA is also involved in protecting cells from gamma radiation. Inactivation of UvrA proteins increased the sensitivity to gamma radiation (Brena-Valle and Serment-Guerrero, 1998). UvrA also suppresses the illegitimate recombination, which is a major cause of chromosomal aberration, along with duplication, deletion, insertion, and translocation (Hanada et al., 2000). These observations indicate that UvrA protein is not

only play the role in UV-radiation but it is also involved in repair other forms of DNA damage induced by ionizing radiation, bulky and non-bulky chemical agents.

The findings in this study indicate that UvrA1 protein of *Deinococcus* spp. is present in genetic linkage with membrane-associated DsbA and DsbB proteins. According to Khairnr N.P. et al (2013), a homolog of DsbA has been found in *D. radiodurans*, which has a role for radiation and cadmium induced oxidative resistance (Khairnar, Joe, Misra, Lim, & Kim, 2013). DsbA and DsbB contain CXXC motif, which have oxidoreductase activity. These both proteins play role in oxidative protein folding by transferring the electrons from unfolded protein to DsbA protein. Then this DsbA protein is oxidized by transfer electrons to the DsbB. Finally, DsbB transfers electrons to ubiquinone and menaquinone and become oxidized (Chim, Harmston, Guzman, & Goulding, 2013). So electron transfers are important for proper functioning of these proteins. When UvrA1 is exposed to oxidative damage then CXXC motif could form disulphide bond and this protein become inactivated. But *Deinococcus* UvrA1 could protect itself from oxidative damage by accepting electrons from DsbB proteins. This association between UvrA1 and DsbB also helps DsbB to remain oxidized. dCSP-I could play the role to make a bridge among these three proteins. Based on structural, functional, topological features of these four proteins, we have proposed a model (Figure 8) to explain the novel functional roles these four proteins play in protecting the critical proteins involved in DNA repair process in *Deinococcus* spp. from oxidative damage. According to our model, the unique interactions/association of DsbA, DsbB, and dCSP-I proteins with the UvrA1 protein

observed in *Deinococcus* spp. provides a novel mechanisms for protection of the UvrA1 against UV-induced oxidative damage (detailed in Figure 8).

A number of elements of this model are supported by experimental observations. In our model, we have indicated that UvrA1 protein transfers to membrane to detect the DNA damage. The UvrA1 protein localizes to the entire chromosome after and before DNA damage in the living cell (Smith, Grossman, & Walker, 2002). This protein translocates to the membrane to detect the DNA damage in *E. coli*. Around 40% of the UvrA was found in the membrane protein fraction (Lin et al., 1997). DNA is found to be associated with the membrane complex among *Deinococcus* spp. (Burrell et al., 1971). So, UvrA protein should transfer to membrane to detect DNA damage among *Deinococcus* spp. But experimental study is required to understand the localization of UvrA1 protein in membrane. The CSI is located at the surface loop of the protein, which is a very flexible region. So, it is very difficult to predict the functional role of this CSI. No experimental study has been conducted to understand the functional association of these four proteins in together. Further experimental studies are expected to validate our proposed models. dCSP-I acts as central protein in our model because it makes the functional linkage among UvrA1, DsbA, and DsbB proteins. Thus studies on deletion or mutation of the gene for dCSP-I protein should be of much importance to understand/confirm the functional role of this protein in the NER pathway and the model proposed here.

Our observation reveals that identified novel molecular traits (CSIs, CSPs) in NER and UVER pathways DNA repair proteins should have a role in radiation, oxidation and

desiccation resistance among *Deinococcus* spp. Furthermore, it indicates that the highly resistant of UvrA1 protein to the oxidative damage may help to restore the NER pathway activity among *Deinococcus* spp., is similar to the result of SOS response of UvrA protein in *E. coli* which is regulated by recA-lexA regulon. Mutations in RecA gene results in the inhibition of SOS response of UvrA protein which causes extremely sensitivity to UV radiation due to defects in recombinational repair and reduced or constitutive levels of NER activity (Kiyosawa, Tanaka, & Matsunaga, 2001). Various oxidoreductase, including thioredoxin (Trxs) and glutaredoxins (Grxs) are involved in the reduction of oxidized cysteine residues to maintain the thiol state. Trxs proteins from all domains of life contain conserved thioredoxin fold and a CXXC catalytic motif (Ezraty et al., 2017). DsbA and DsbB proteins also contain thioredoxin fold and CXXC motif in their structure. So, these proteins could play a role in oxidative damage protection of UvrA1 and provide resistance to this damage. RecA plays significant role in ionizing and UV-radiation resistance of *Deinococcus* spp. *Deinococcus* RecA cannot be fully complimented by *E. coli* (Schlesinger, 2007). We have identified a 1 aa CSI in that protein which is specific for Deinococcus-Thermus spp. (Figure 10). RecA protein is present in a genetic linkage with CinA and CnP proteins which is specific for Deinococcus-Thermus. It is already known that CinA binds with RecA protein and locates it to the cell membrane in *Streptococcus pneumoniae* (Masure et al., 1998). Further studies may provide interesting insight about these unique molecular characteristics of the RecA protein.

Cysteine is important for functional activity of proteins where it has catalytic, regulatory, structure stabilizing, cofactor binding, and other functions. Cysteine residues which are exposed to surface, may easily switch their ability to function as nucleophiles. This electrostatic change may affect the topological and functional ability protein (i.e. interacting with environment, other proteins and charge molecules) (Marino, 2014). We have identified a large number of surface exposed cysteine residues in UvsE protein. Oxidative damage of surface exposed cysteine residues may lead to protein damage. Zn-ribbon protein may also play similar role to protect UvsE protein during oxidative damage.

The proposed models for regulating UV-induced oxidative damage may have implications to further understanding of bioremediation research on *Deinococcus* spp. We have also identified CSIs in other DNA repair proteins which are distinctive molecular characteristics of *Deinococcus* spp. Further analyses of these CSIs can help to decipher the extraordinary tolerance of *Deinococcus* to other types of radiations, mutagens and desiccation.

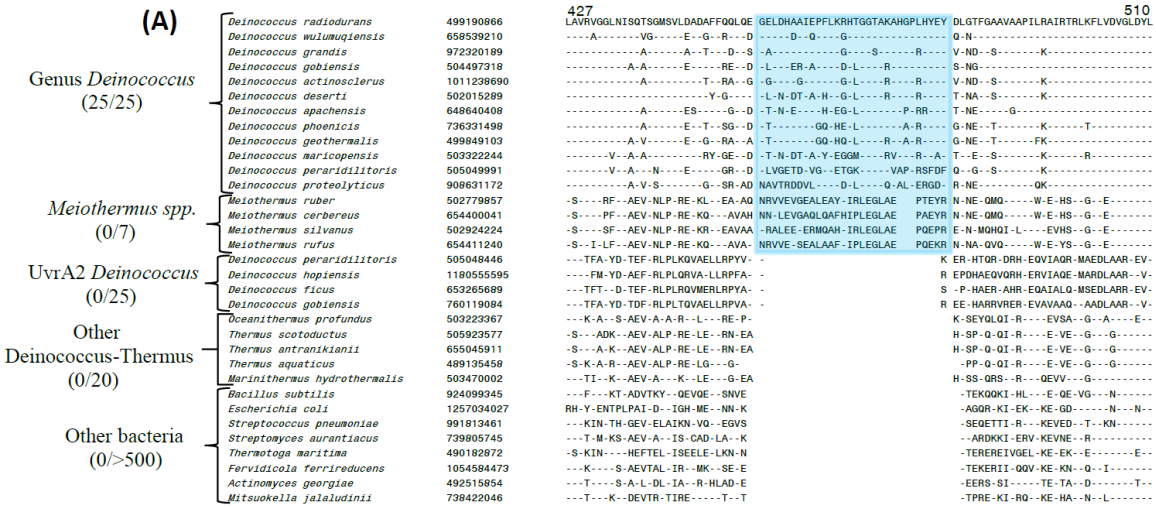


Figure 1A: Partial sequence alignment of excinuclease ABC subunit A (UvrA1) protein showing a 30 amino acid insertion that is uniquely shared by all *Deinococcus* group of bacteria. The dashes (-) in the alignment file indicates the sequence identity with the top line. *Meiothermus* spp. have an insertion in this position but there is a 3 amino acid deletion in the insertion region. The region is not also conserved. UvrA2 protein of *Deinococcus* spp. does not contain this insertion. So, this 30 amino acid insertion is a distinctive characteristic of the member of *Deinococcus* spp.

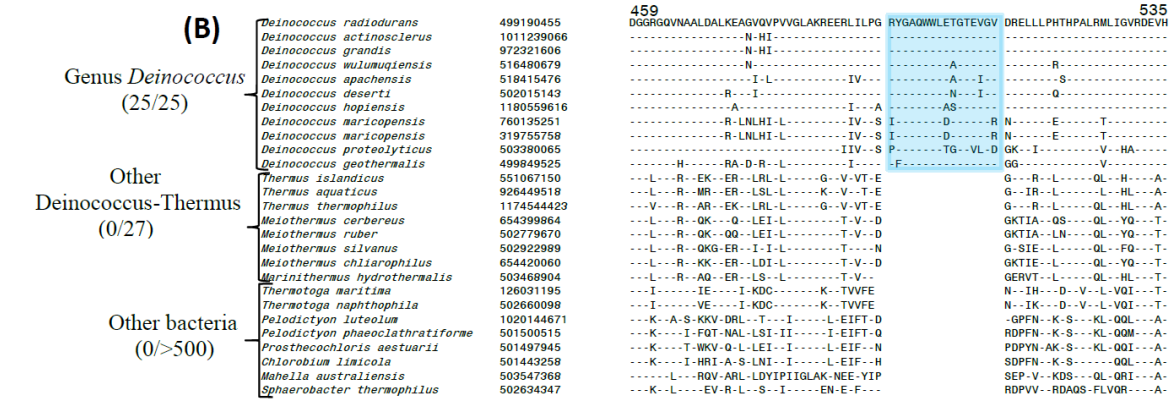


Figure 1B: Partial sequence alignment of conserved region of excinuclease ABC subunit C (UvrC) protein showing a 16 amino acid insertion that is specific for all *Deinococcus* group of bacteria.

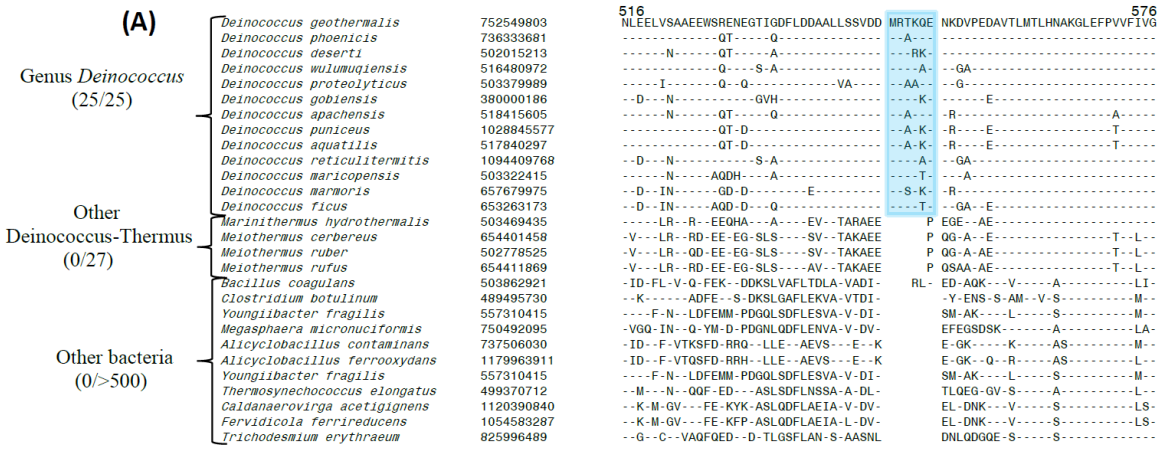


Figure 2A: Partial sequence alignment of conserved region of DNA helicase UvrD protein showing a 6 amino acid insertion that is uniquely shared by all *Deinococcus* group of bacteria.

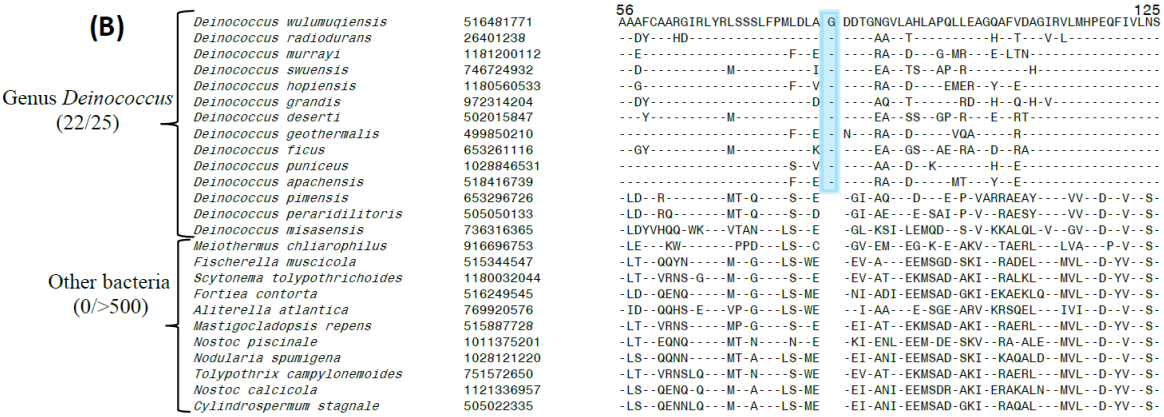


Figure 2B: Partial sequence alignment of conserved region of UV damage endonuclease (UvsE) protein showing a 1 amino acid insertion that is uniquely shared by most of the *Deinococcus* group of bacteria. Three *Deinococcus* spp. do not contain this insertion which is marked by purple color.

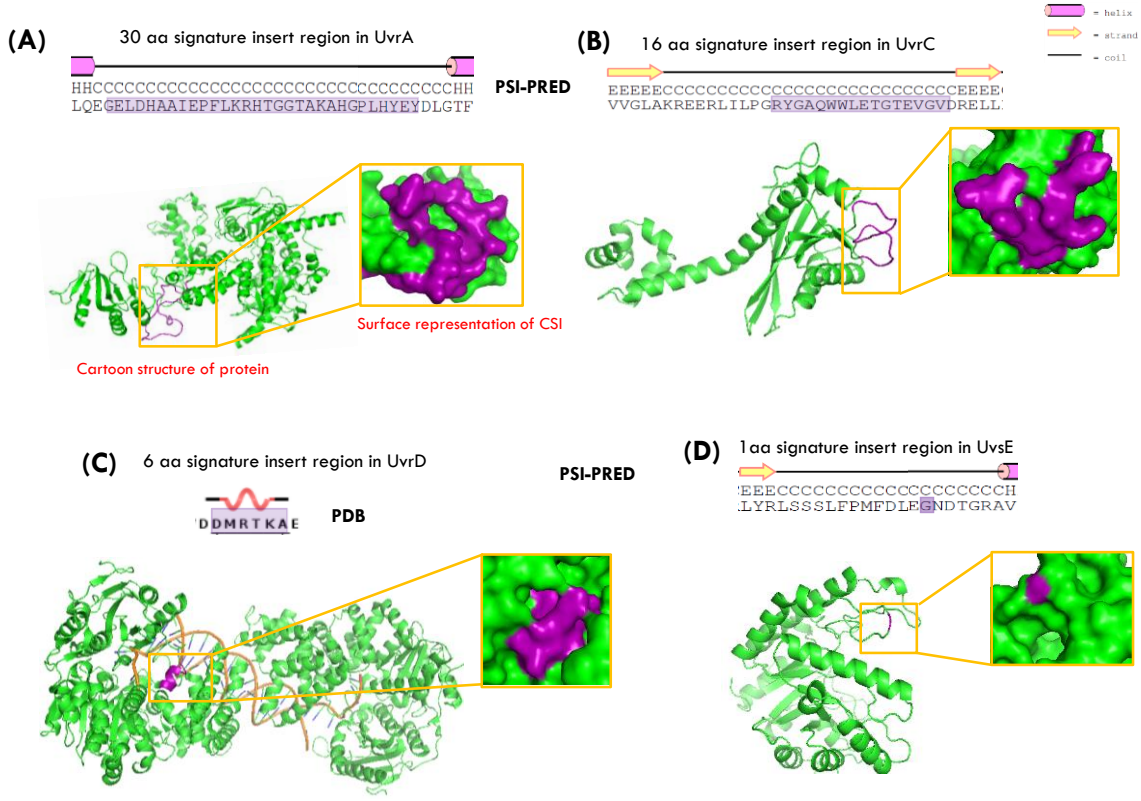


Figure 3: A) Location of a 30 amino acid CSI in modeled UvrA1 protein of *Deinococcus radiodurans*. This CSI is located at loop region of UvrA1 protein. Surface location analysis delineates that this CSI is present in surface loop of protein. B) Location of a 15 amino acid CSI in modeled UvrC protein. Structural analysis reveals that this CSI is located at surface loop region in the modeled UvrC. C) Carton and surface representation of a 6 amino acid CSI in crystalized UvrD protein (PDB id: 4C2T). This CSI is located at helix region of protein. Surface representation reveals that this CSI is located at surface. D) Structural analysis of 1 aa CSI in UvsE protein. This CSI is located at surface loop region in modeled UvsE. Secondary structure of CSIs region has been predicted by PSI-PRED server. The Templates for homology modeling of UvrA, UvrC and UvsE of *Deinococcus radiodurans* have been selected from following species sequentially: *Geobacillus* sp. Y412MC52 (PDB id: 3UWX), *Thermotoga maritime* (PDB id: 2NRT) and *Sulfolobus acidocaldarius* (PDB id: 3TC3). All CSI regions are marked by purple color.

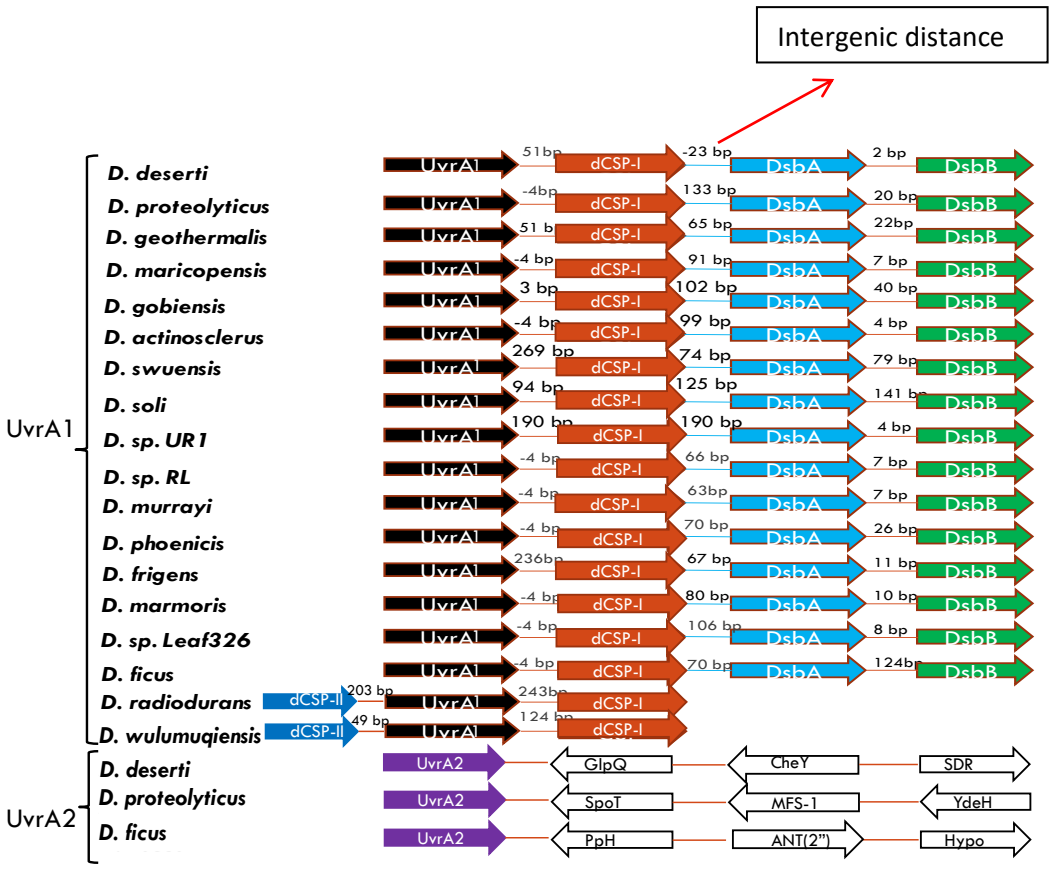


Figure 4: A diagrammatic representation of genomic neighborhood analysis result of UvrA1 protein for representative *Deinococcus* spp. Among all *Deinococcus* spp. (except *Deinococcus radiodurans* and *Deinococcus wulumuqiensis)*, dCSP-I, DsbA and DsbB proteins are present in the neighbourhood to UvrA1. In *Deinococcus radiodurans* and *Deinococcus wulumuqiensis*, UvrA1 protein is present as a neighbourhood of dCSP-I. These four proteins are also present in the same direction in the *Deinococcus* genome. Intergenic distances between proteins are also shown in the diagram. Intergenic distances between proteins are less than 200 base pairs in most of the species. So, all four proteins should have same phylogenetic profile and present in the same operon. UvrA2 protein of *Deinococcus* spp. does not show this pattern of neighbourhood. In *Deinococcus radiodurans* and *Deinococcus wulumuqiensis*, UvrA1 protein is present as a neighbourhood of another CSP (dCSP-II) which is only specific for these bacteria.

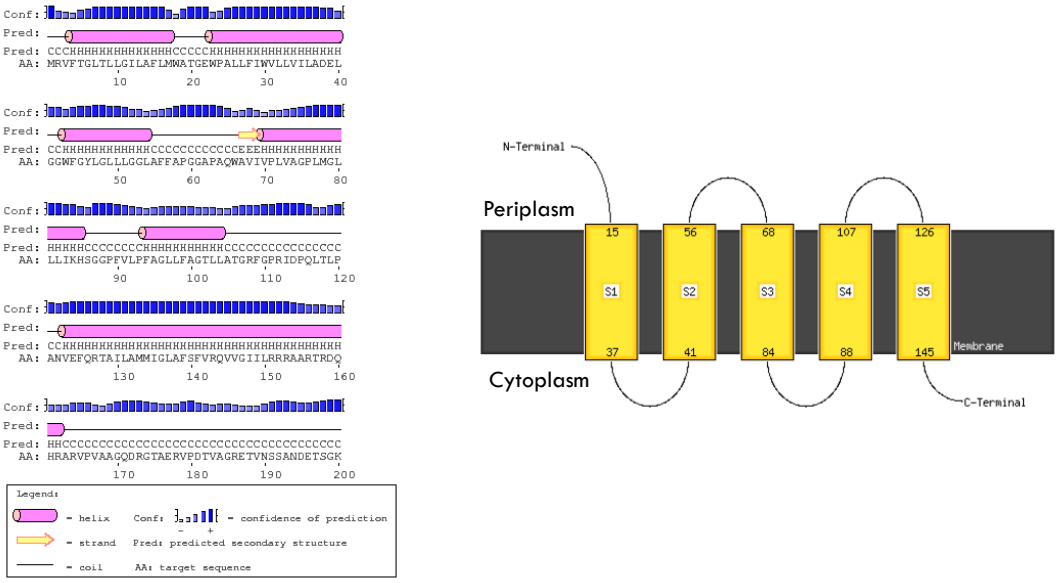


Figure 5: Diagrammatic representation of (A) predicted secondary structure of dCSP-I (Accession ID: NP_295493). Secondary structure delineates that most of the residues are present in helix region of the protein. It is marked by magenta color. It has single beta strand region which is marked by yellow color. Black lines represent the coiled region of the protein. (B) Membrane topology of dCSP-I. Membrane topology has shown that it has five transmembrane helices, two loop regions are present in the periplasm and other two loops are present in the cytoplasm. N-terminal and C-terminal regions are positioned periplasmic and cytoplasmic region respectively.

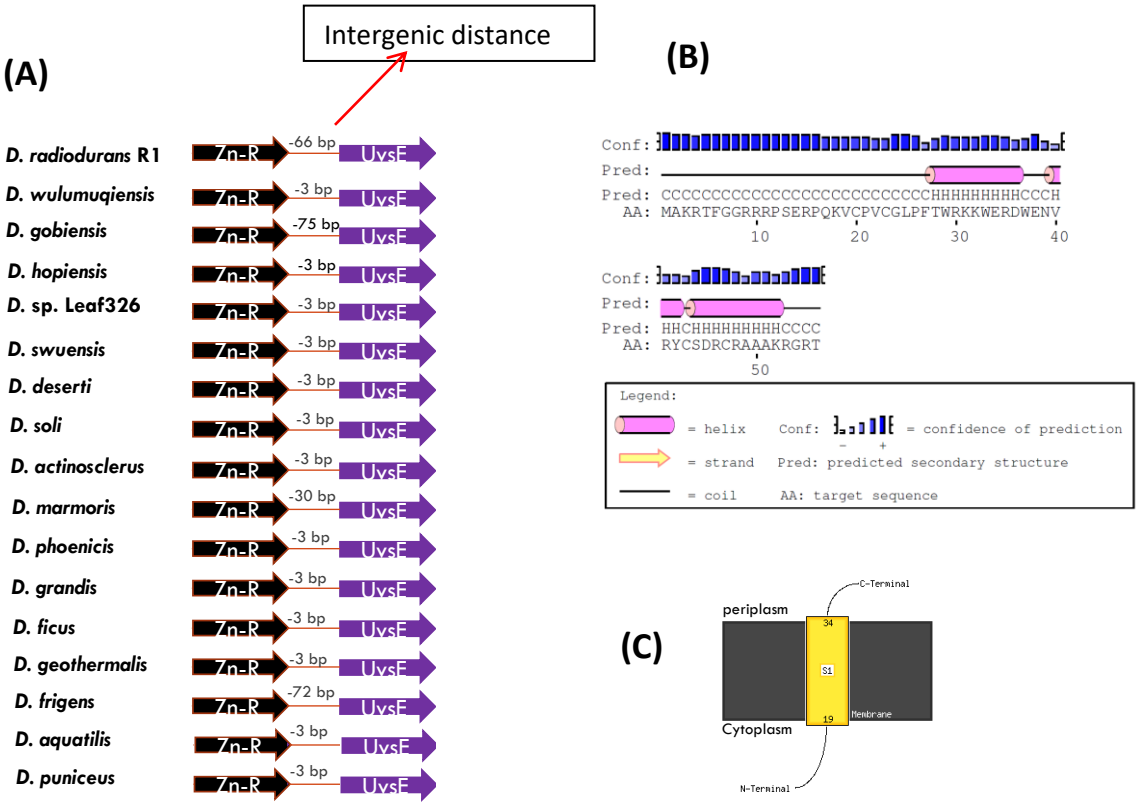


Figure 6: A) Diagrammatic representation of genomic neighborhood analysis result ofUV-endonuclease protein UvsE in *Deinococcus* spp. In all of the *Deinococcus* spp. UvsE protein is present in the neighbourhood to Zn-ribbon protein (Zn-R). These two proteins are also present in the same direction in the *Deinococcus* genome. Intergenic distances between proteins are also shown in the diagram. Intergenic distances between these proteins are less than 200 base pairs in most of the species. So, all four proteins should have same phylogenetic profile and present in the same operon. B) Predicted secondary structure of Zn-ribbon protein (Accession ID: AFD24462.1). Secondary structure reveals that half of the residues are present in helix region of the protein and other half of the residues are present in coiled region. C) Predicted membrane topology of Zn-ribbon protein. This protein contains a transmembrane region. N-terminal and C-terminal region are positioned cytoplasm and periplasm respectively.

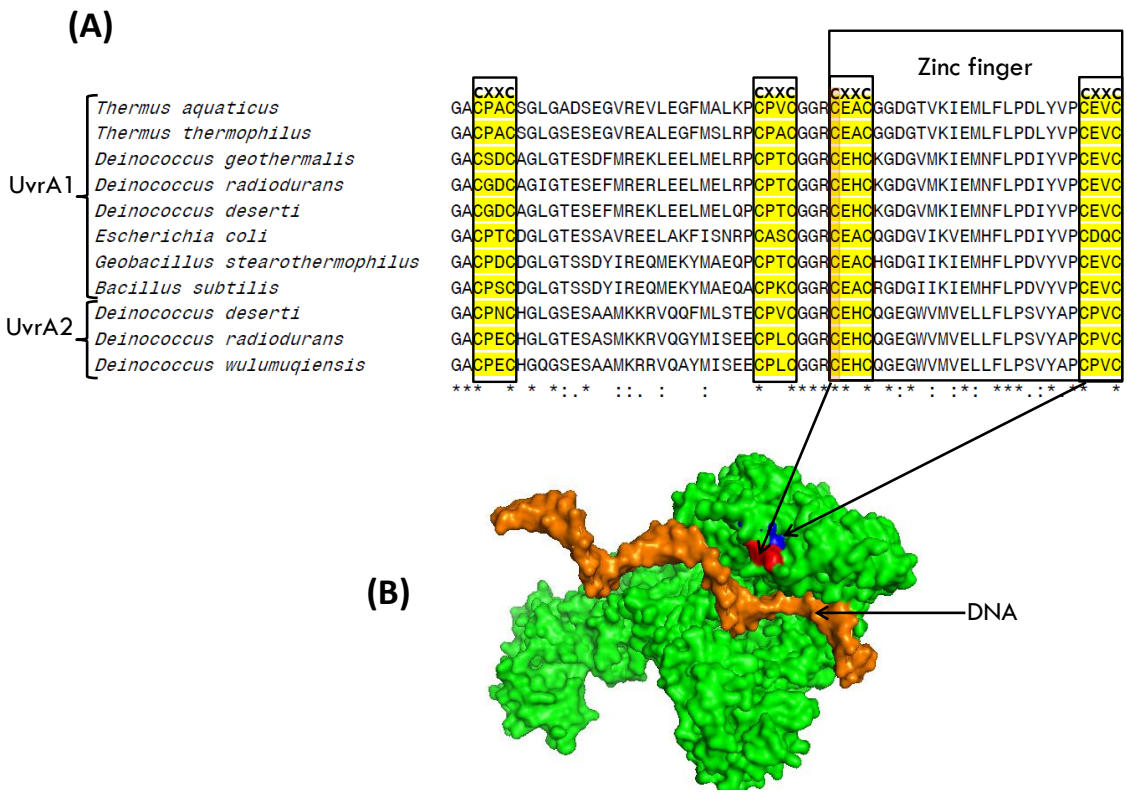


Figure 7: A) Partial sequence alignment of UvrA protein. Zinc-finger motifs of UvrA protein contain CXXC motifs. Four CXXC motifs are conserved through all bacterial species. Purple boxed cysteine of CXXC motif has been substituted with other amino acids experimentally and found significant for the function of UvrA protein. B) Structural analysis of CXXC motif of UvrA protein. CXXC motifs of Zinc finger III motifs are present in the surface. Two CXXC motifs are present in close proximity to DNA in UvrA-DNA binding complex. Two CXXC motifs have been marked by red and blue colour respectively.

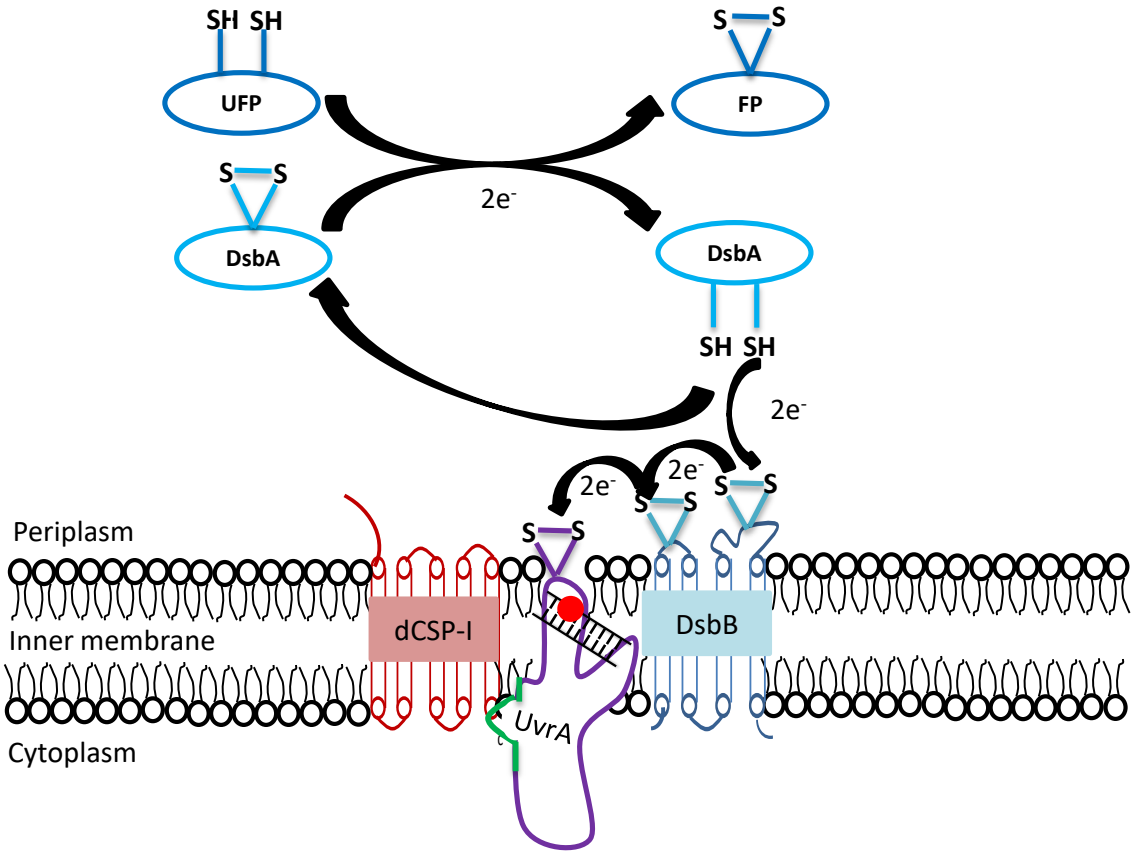


Figure 8: Proposed model too understand the functional role of UvrA1, dCSP-I, DsbA and DsbB in together in DNA repair. Oxidized DsbA transfers its active site disulfide to an unfolded protein substrate; the substrate cysteines are thus oxidized, and the protein becomes folded. The active site cysteines of DsbA become reduced and are re-oxidized by transferring electron to DsbB protein. During the oxidative damage, UvrA protein becomes inactivated by forming the disulphide bond between cysteine residues of CXXC motifs and can not continue DNA repair. *Deinococcus* specific dCSP-I by interacting with CSI of UvrA1 protein acts as a functional bridge among UvrA1, DsbA, and DsbB. This oxidized UvrA become functional by accepting electron from DsbB. Now, UvrA1 protein can detect the DNA damage and continue the NER pathway. The DsbB protein becomes oxidized and play role in the activation of DsbA protein which is responsible for fold formation in the substrate protein. The CSI region of UvrA protein has been marked by green color.

(A)

```
Deinococcus reticulitermitis          PALGLVCMTVGPELRF AADYCAARGIRLYRLSS ENDERAYGVSDLFPVCEATGTP
Deinococcus radiodurans R1            PQLGLVCLTVGPEVRF AADYCAAHDIRLYRLSS ENDERAYSPAELLPICEATGTP
Deinococcus wulumuqiensis             PQLGLVCLTVGPEVRF AAAFCAARGIRLYRLSS ENDERAYSPAQLLPICEATGTP
Deinococcus aquatilis                 PAYGLVCMTVGPEVRF AAAFCAARGIRLYRLSS ENDERAYGVDDLWPVCQATGVP
Deinococcus swuensis                  PAYGLVCLTVGPEVRF AADFCAARGIRLYRMSS ENDERAYGPADLLPVCEATGTP
Deinococcus deserti VCD115            PAYGLVCLTVGPEVRF AAAYCAARGIRLYRMSS ENDERAYGPQDLLPVCEATGTP
Deinococcu smurrayi                   PAYGLVCMTVGPEVRF AAEFCAARGIRLYRLSS ENDERAYGPQDLLPVCEGTGVP
Deinococcus phoenicis                 PAYGLVCLTVGPEVRF AAAFCAARGIRLYRLSS ENDERAYGPADLLPVCEATGTP
Deinococcus geothermalis DSM1130      PAYGLVCMTAGPEVRF AAAFCAARGIRLYRLSS ENDERAYGPAELLPVCEGTGVP
Chloroflexi bacterium 54-19           PRLGLVCITNSEAVRF ALTFCIENKIRLYRLSS ENDEHAYGAWEILEICQAAHVP
Nodularia spumigena CENA596           PYLGLVCVTISKQVRF ALSFCQQNNIRLYRMTS ENDEYAYSSEEILAVCEQTGIP
Nostoc punctiforme PCC73102           PSLGLVCITSDKQVRF ALSFCQQNKIQLYRMSS ENDEYAYSADEILAVCQQAGVP
Scytonema millei VB511283             PHLGLVCITFSKEVRF ALTFCVRNSLRLYRMTS ENDEHAYSASEILEVCKRAEIP
Hapalosiphon sp.MRB220                PYLGLVCITFSKQVRF ALTFCQQYNIRLYRMSS ENDEYAYSASEILEVCQRTGVP
Fischerella sp.PCC9431                PYLGLVCITFSKQVRF ALTFCQQYNIRLYRMSS ENDEYAYSASEILEVCQRTGVP
Stanieria cyanosphaera PCC7437        PVLGLVCITASSEVRF AIAFCQREKIRLYRMSS ENDEHTYSAAEILEICQATKVP
Microcoleus vaginatus FGP-2           PELGLVCITTSDAVRF AIDFCSANQIKLYRMTS ENDEYAYSSSEILEVCLDTGVP
Phormidium ambiguum IAMM-71           PQLGLVCITASDKVRY AIAFCAANNINLYRVTS ENDEYAYSASEIFDICCAANIP
                                      *  ****:* .  :*: *  :*    :.***::* **** :*.  ::  :*  :  *
```

(B)

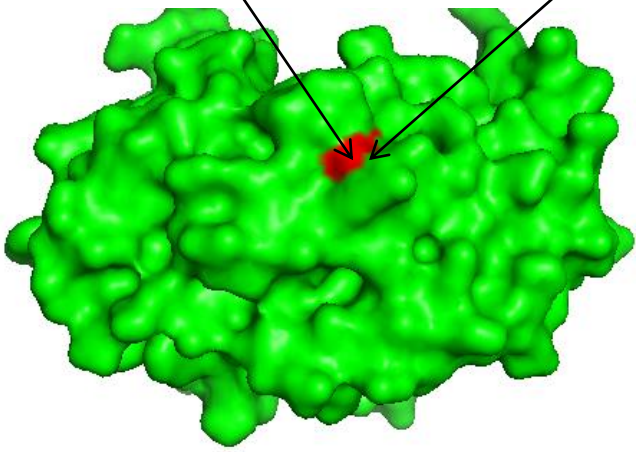
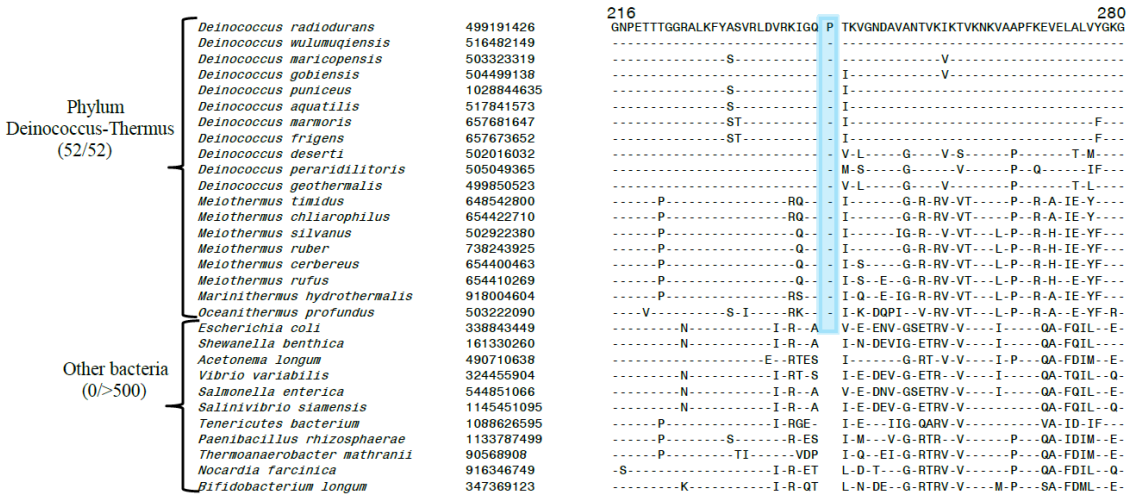


Figure 9: A) Partial sequence alignment of UvsE protein. UvsE protein contains three conserved cysteine residue in their sequence. Cysteine residues are marked by yellow color. B) Location analysis of these cysteine residues in the UvsE protein. Among three cysteine residues, two of them are located at protein surface which have been marked by red color.



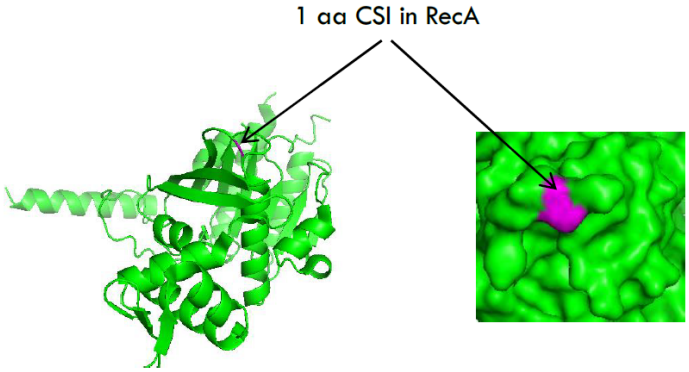


Fig: Location of 1 aa CSI in RecA protein

Figure 10: Partial sequence alignment of conserved region of DNA recombination protein RecA showing a 1 amino acid insertion that is specific for all bacteria of Deinococcus-Thermus phylum. Crystal structure of RecA protein is available from *Deinococcus radiodurans* (PDB id: 1XP8). Structural analysis reveals that 1 aa CSI in RecA protein is located at surface loop.

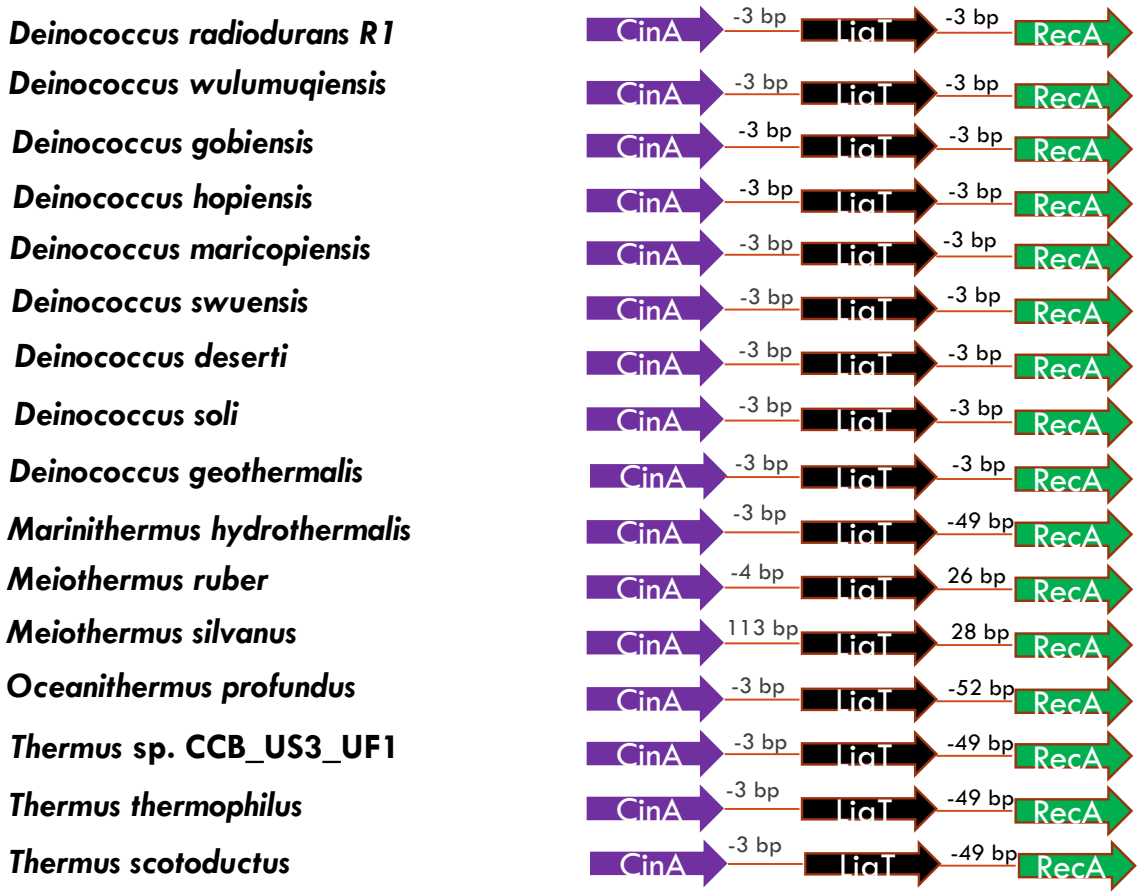


Figure 11: Diagrammatic representation of genomic neighborhood analysis result of DNA recombination protein RecA for representative Deinococcus-Thermus spp. RecA protein is present in a genetic linkage with LigT and CinA proteins throughout all Deinococcus-Thermus spp. The intergenic distance among them is less than 200 bp and is present in same direction. So, RecA, LigT and CinA proteins should present in same operon among Deinococcus-Thermus phylum.

Table 1: Summary of CSIs found in DNA repair proteins of *Deinococcus* spp.

Protein name	Pathway	Protein ID	Figure(s)	Indel size	Indel [c] position
8-Oxoguanine DNA glycosylase (MutY)[a]	BER	DR2285	Supp. Fig.1	4 aa ins	24-82
Formamidopyrimidine and 8-oxoguanine DNA glycosylase (MutM)	BER	DR0493	Supp. Fig.2	2 aa del	168-225
Endonuclease III (Nth)	BER	DR0928	Supp. Fig.3	2 aa ins	144-228
UvrA1	NER	DR1771	Fig.1A	30 aa ins	421-514
UvrC	NER	DR1354	Fig.1B	16 aa ins	459-535
UvsE	UVER	DR1819	Fig.2B	1 aa ins	48-125
DNA recombination protein (RecA)[b]	HR	DR2340	Supp. Fig.4	1 aa ins	216-280
DNA helicase recombination protein (RecR) [b]	HR	DR0198	Supp. Fig.5	2 aa del	104-164
DNA helicase (RecQ)	HR	DR1289	Supp. Fig.6	2 aa del	453-510
Helicase RecD protein	HR	DR1902	Supp. Fig.7A	2 aa del	376-427
Helicase RecD protein	HR	DR1902	Supp. Fig.7B	2 aa del	426-493
crossover junction endodeoxyribonuclease RuvC protein	HR	DR0440	Supp. Fig.8	2 aa del	82-147
DNA helicase RecG protein	HR	DR1916	Supp. Fig.9	1 aa ins	14-73
DNA polymerase I (PolA)	MP	DR1707	Supp. Fig.10	2 aa ins	191-257
DNA polymerase III, alpha subunit(DnaE)	MP	DR0507	Supp. Fig.11	65 aa ins	315-491
DNA polymerase III, alpha subunit(DnaE)	MP	DR0507	Supp. Fig.12	2 aa ins	75-131
DNA ligase (LigA)	MP	DR2069	Supp. Fig.13	3 aa ins	101-169
UvrD	MP	DR1775	Fig.2B	16 aa ins	516-576
Gyrase A(GyrA)	MP	DR1913	SupP.Fig.14	1 aa ins	265-341
Gyrase B(GyrB)	MP	DR0906	Supp. Fig.15	1 aa ins	27-99

BER- Base Excision Repair; NER-Nucleotide Excision Repair; HR- Homologous Recombination;
UVER-UV damage endonuclease (UvsE)-dependent excision repair; MR- Mismatch repair; MP-Multiple pathways.

[a] The indel is specific for *Deinococcus* and *Hyphomonas* spp..

[b] The indel is specific for Deinococcus-Thermus.

[c] the idel position indicates the region of the protein containing the CSI.

Table 2: CXXC motif in DNA repair proteins of *Deinococcus* spp.

Protein name	Pathway	Protein ID	CXXC motif in the sequence
8-Oxoguanine DNA glycosylase (MutY)	BER	DR2285	Yes(1)
Formamidopyrimidine and 8-oxoguanine DNA glycosylase (MutM)	BER	DR0493	Yes(2)
Endonuclease III (Nth)	BER	DR0928	Yes(1)
UvrA1	NER	DR1771	Yes(5)
DNA or RNA helicase of superfamily II (Rad25)	NER	DRA0131	Yes(2)
RecO	HR	DR0819	Yes(2)
RecR	HR	DR0198	Yes(2)
RecQ	HR	DR1289	Yes(1)
SbcC	HR	DR1922	Yes(1)
RadA	HR	DR1105	Yes(2)
DNA polymerase III subunit gamma/tau	MP	DR2410	Yes(1)
DNA ligase (LigA)	MP	DR2069	Yes(1)

BER- Base Excision Repair; NER-Nucleotide Excision Repair; HR- Homologous Recombination;
UVER-UV damage endonuclease (UvsE)-dependent excision repair; MR- Mismatch repair; MP-Multiple pathways.

CHAPTER 3

Studies on Chlamydiae-specific CSIs and CSPs to understand their possible cellular functions

3.1 Introduction:

The members of the phylum Chlamydiae are intracellular bacteria that cause a variety of diseases by infecting eukaryotic cells (Brunham et al., 1990; Carabeo et al., 2003; CDC, 2014). Chlamydial spp. are responsible for the bacterial sexually transmitted infection, trachoma, community-acquired pneumonia and also risk factor for cardiovascular disease (Kuo et al., 1993; Kuo et al., 1995; Batteiger et al., 2014; Kohlhoff et al., 2014). Although members of the phylum Chlamydiae produce infections that have different clinical manifestations, the main features of the intracellular lifestyle characterized by biphasic development cycle are similar for all chlamydial species and strains (Hybiske and Stephens, 2007; Moulder, 1991). The infectious elementary body (EB) initiates the infection by their ability to attach and invade susceptible cells. The most critical stage of infection is the conversion of EB into the metabolically active form, termed reticulate body (RB) by 6 to 8 hour postinfection (HPI). DNA is present in the relaxed form in RB that facilitates the replication of RB repeatedly by binary fission during midcycle. After 8 or 10 rounds of division, RBs convert into EB. These EBs are released to infect neighboring cells (Wyrick, 2010; Abdelrahman and Belland, 2005; Batteiger et al., 2014).

This chlamydial lifestyle is controlled by the expression of genes during the development cycle in three main temporal classes (Abdelrahman and Belland, 2005; Nicholson et al., 2003; Belland et al., 2003). Early genes are transcribed within a few hour of infection through an unknown mechanism, and most of these genes play the role in stress response,

translational machinery and DNA-binding (Rosario and Tan, 2012). The expression of late genes is essential to determine whether an individual RB continues to divide or terminally differentiates into EB. The transcription of late genes is negatively regulated by the transcription factor called EUO, which binds with the late genes promoters and may prevent the transcription by RNA polymerase (Yu and Tan, 2003; Yu et al., 2006; Rosario et al., 2014; Rosario and Tan, 2015; Rosario and Tan, 2012). Majority of chlamydial genes are transcribed in midcycle during the RB replication. Midcycle genes are positively regulated by negative DNA supercoiling among Chlamydial spp. because promoters of midcycle genes are highly transcribed in more supercoiled templates. Simultaneously, a subset of early genes also has supercoiling dependents promoters. These genes also upregulate in midcycle of Chlamydial development.(Niehus et al., 2008; Cheng and Tan, 2012; Orillard and Tan, 2016).

DNA supercoiling is considered as the global mechanism for regulation of gene expression which is believed to represent the highest level in the hierarchy of prokaryotic gene regulation (Dorman, 1991; Laponogov et al., 2013). A recent study has revealed that temporal expression of chlamydial genes during intracellular infection is expected to be regulated by changes in DNA supercoil level (Orillard and Tan, 2016). DNA topoisomerase enzymes play role in controlling the DNA supercoiling level in bacterial cells (Laponogov et al., 2013). DNA Gyrase is a heterotetrameric protein composed of GyrA and GyrB subunits which are unique in their ability to introduce negative supercoils into DNA thereby controlling chromosome supercoiling (Gellert et al., 1976; Reece and Maxwell, 1991; Maxwell and Lawson, 2003). This mechanism promotes

replication fork advancement and allows global regulation of gene expression. This negative supercoiling has been relaxed by Topoisomerase I. Topoisomerase IV is involved in decatenation and relaxation [(Bush et al., 2015; Sissi and Palumbo, 2010; Champoux, 2001; Viard and de la Tour, 2007; Dorman and Berlin, 2008). Chlamydial spp. encode these three topoisomerases that are involved in regulating supercoiling levels. The genes of gyrase (gyrA and gyrB) are co-transcribed the same operon and this is also the case for topoisomerase IV (parC and parE) (Orillard and Tan, 2016). The promoter of topoisomerase genes are also transcribed at higher level during supercoiling among chlamydial spp. (Huang, 1996). One of the most surprising characteristics of chlamydial topoisomerases is that the promoter of gyrBA is up-regulated by higher levels of DNA supercoiling, which has not been observed for any other bacterial spp. (Huang, 1996; Orillard and Tan, 2016). Therefore, this is a distinctive characteristic of chlamydial spp. However, No study has been performed targeting the novel molecular attributes of Topoisomerase II proteins involved negative DNA supercoiling to understand the role during intracellular infection.

The advent of large-scale genome sequencing is providing researchers with tremendous wealth of information that can be used to identify molecular traits for different groups of prokaryotic and eukaryotic organisms. Our recent work has identified a large number of conserved signature indels (CSIs) in a variety of proteins involved in different cellular functions that are unique molecular characteristics of the chlamydial homologs (Gupta et al., 2015). Additionally, our comparative analyses also have identified chlamydial-

specific proteins in terms of conserved signature proteins (CSPs) (Gupta and Griffiths, 2006; Gupta et al., 2015).

The work in this chapter focuses on the genomic and sequence features of GyrA and GyrB proteins to identify any molecular characteristics that may helpful to understand the role of DNA gyrase enzymes in chlamydial intracellular development. This study has led to the identification of several highly specific molecular signatures in the form of CSIs that clearly distinguish the DNA Gyrase proteins of chlamydial spp. from protein homologs of other bacteria. Structural analysis delineates that most of the CSIs, which are distinct characteristics of Chlamydial spp., are present on the protein surface. The results of protein-protein interactions analyses provide evidence that GyrA and GyrB proteins are present in a distinctive genetic linkage with thymidylate kinase (TMK), DNA polymerase III subunit δ'(HolB) and one of the chlamydial specific CSP (Accession: ARZ55828.1) in the chlamydiae species. The importance of these observations is discussed to understand the unique intracellular infection of chlamydial spp.

3.2 Methods:

3.2.1 Protein-protein interactions (PPIs) analyses to examine the unique association of Chlamydiae GyrA and GyrB proteins:

The STRING database (Szklarczyk et al., 2017) was used to predict unique association of GyrA and GyrB proteins with other proteins that are specific for the phylum Chlamydiae. The STRING is a widely used, comprehensive database for PPIs. The STRING database contains both experimental and predicted PPIs data. It predicts PPIs based on co-

occurrence, gene fusion, co-expression and neighborhood (von Mering et al., 2005; Szklarczyk et al., 2015).

3.2.2 Identification of genetic linkage in Chlamydiae GyrA and GyrB proteins:

The chlamydial spp. genomes have been inspected to identify the genetic linkage of GyrA and GyrB proteins. We have analyzed the graphics view of the Chlamydiae genomes to find out the neighborhood proteins of these two proteins. Intergenic distance between two proteins was mapped manually based on their position in the genome. The direction of the genes in the genome has been identified by visual inspection.

3.2.3 Structural analysis of the CSIs and homology modeling of the GyrA and GyrB homologs from the phylum Chlamydiae:

Three-dimensional structure of the C-terminal domain of GyrA and N-terminal ATPase domain of GyrB proteins homologs of Chlamydial spp. was modeled from their full-length sequence using the homology modeling technique (Sali and Blundell, 1993). The PSIPRED v3.3 web server was used to perform the secondary structure analyses of the selected GyrA and GyrB homologs sequence (Buchan et al., 2013). To find the suitable template for homology modeling, PSI-Blast (Altschul et al., 1997) search was used against the Protein Data Bank (PDB) (Rose et al., 2015). To select the template; the sequence identity between template and target proteins, the probability score, expect value (E-value), analysis of the quality of sequence alignment, and the secondary structure compatibility between target and templates have taken into consideration. Initially, 200 models were generated by using modeller v9.14 (Sali and Blundell, 1993)

and ranked by using assigned discrete optimized potential scores (DOPE) (Shen and Sali, 2006). The selected models have been refined by the modrefiner (Xu and Zhang, 2011). Validation of the models of GyrA and GyrB proteins was performed by using four independent servers: PROSA (Wiederstein and Sippl, 2007), RAMPAGE (Lovell et al., 2003), ERRAT (Colovos and Yeates, 1993), Verify3D (Eisenberg et al., 1997) and QMEAN (Benkert et al., 2008). Then the structure results of the CSIs regions was inspected and compared with the results of PSIPRED analysis to ensure the reliability. The PyMoL was used to carry out the superimposition of validated models with templates to evaluate the structure and location of identified CSIs in the structure of GyrA and GyrB.

3.3 Results

3.3.1 Distinguishing characteristics in the GyrA and GyrB proteins sequences of chlamydial spp.:

Negative DNA supercoiling is involved in reduced winding of the DNA helix in the bacterial cells that favors these DNA reactions which depend on strand separation, i.e, transcriptions, transpositions, replication, and recombination (Dorman, 1991). This negative DNA supercoiling is regulated by GyrA and GyrB proteins (Reece and Maxwell, 1991). The sequences of these two proteins were examined to gain insights into the molecular basis of unique characteristics of chlamydial topoisomerase II enzymes. The results of our sequence analyses have identified a number of CSIs in GyrA and GyrB proteins that are distinctive characteristics of the chlamydial spp. Among these CSIs, the

first CSI is a 1 aa insertion in GyrA which is commonly shared by all chlamydial spp. (Figure 12A). The GyrA protein contains another 1 aa CSI that is also a unique molecular characteristic for chlamydial spp. (Figure 12B). Multiple sequence alignments of GyrB protein have identified two CSIs which are comprised of 6 and 4 aa inserts in the conserved region of that protein (Figure 13A and Figure 13B). The 6 aa CSI of GyrB has been identified in our previous work; however, no structural and functional analyses of this CSI have been carried out by us (Gupta and Griffiths, 2006). All of these CSIs are flanked by conserved residues on both sides and are ubiquitously present in all Chlamydial homologs of GyrA and GyrB. Except for Chlamydiae these CSIs are not present in any other bacterial homologs in the top 500 blast hit.

3.3.2 Locations of the CSIs in the structure of the GyrA and GyrB proteins and their possible significance:

The GyrA protein is composed of two domains: N-terminal domain (NTD) and C-terminal domain (CTD) (Lanz et al., 2014). The GyrB protein also possesses two domains: N-terminal ATPase domain and Mg^{2+}-binding-TOPRIM domain (Brino et al., 2000). The insertions in the GyrA protein are present in C-terminal domain. Both of the insertions in the GyrB protein are located in N-terminal ATPase domain.

The locations of different CSIs in the model structures of proteins have been mapped and surface representations of these CSIs in the structure are exhibited in Figure 15. Secondary structure elements are also displayed above the primary structure. We have modeled the structure of GyrA-CTD and GyrB-N-terminal ATPase domains to predict

the location of CSIs. The first 1 aa CSI in GyrA is present in a helix region, and surface analysis reveals that this CSI is located at protein surface. Another 1 aa CSI in GyrA is located in a surface loop of the protein (Figure 15). The 6 aa and 4 aa CSIs in GyrB protein are also located in surface loops in the three-dimensional cartoon structure of protein (Figure 15).

The first 1 aa CSI in GyrA protein is located in the β-CA loop. This β-CA loop is important for folding of GyrA-CTD and imparts superhelicity on the DNA as it binds with this loop. It has been proposed that the abundance of positively charged arginine and lysine on the loop enhance the DNA contact (Lanz et al., 2014). Interestingly, this CSI contains arginine or lysine amino acid. So, this CSI may play a role in increasing the strength of DNA contact with GyrA-CTD among chlamydial spp. which may also enhance superhelical activity of this protein among this group of bacteria. This CSI is located in helix at a position which plays a role in interaction between blades 1 and 6. When we modeled the GyrA-CTD region by deleting this CSI, no helix region was formed near the interaction side between blade 1 and blade 6 (Figure 16). Overall, this CSI may play a role in protein folding and the superhical activity among chlamydial spp.

The 6 aa CSI in GyrB protein is also present in a loop region of the protein (Figure 17). This loop forms a channel which facilitates the nucleotide binding and also play role in dimerization (Brino et al., 2000). Structural analysis reveals that chlamydial spp. have an extended loop due to the presence of this 6 aa CSI in the protein (Figure 17). So, this loop

may play a more efficient role in nucleotide binding among chlamydial spp. which facilitates the dimerization of GyrB during supercoiling.

Another 1 aa CSI in GyrA and 4 aa CSI in GyrB are located at surface loop in the structure (Figure 15) Surface located loop is usually necessitated in mediating protein-protein and protein-ligand interactions (Akiva et al., 2008). So, these two CSI may responsible for unique protein-protein interaction among chlamydial spp.

3.3.3 Unique interactions of GyrA and GyrB proteins ubiquitously present in the chlamydial spp.

We have analyzed CSI containing proteins to identify any unique interactions that are specific for the chlamydial spp. We have predicted the unique interaction by using protein-protein interactions (PPIs) data from STRING database. Our PPIs analyses reveal that GyrA and GyrB proteins are associated with each other among all chlamydial spp. These two proteins also have unique genetic linkage with thymidylate kinase (TMK), DNA polymerase III, subunit δ' (HolB) and one of the hypothetical protein. We have analysed this hypothetical protein to understand if this is a chlamydial specific CSP or not. Our analyses have confirmed that this hypothetical protein is a chlamydial specific CSP that has been identified by earlier work of our lab at 2006 (Figure 18) (Gupta and Griffiths, 2006). This CSP is still specific for the phylum Chlamydiae though a large number of genomic data is now available in compare to 2006. This provides clear evidence about the realiability and specificity of CSPs. These associations have been predicted based on the conserved genome neighborhood context. Neighborhood analysis

of all available sequenced genome species will provide clear hints to predict the functional association of these five proteins. Neighborhood analysis reveals that the genes GyrA and GyrB proteins are present in genetic linkage with TMK, HolB and one of our identified CSP among all available sequenced chlamydial spp. (Figure 19). When two or more genes are present in a certain group of microorganisms, they should have similar phylogenetic profile. Proteins that have similar phylogenetic profile should have carried out functional or physical interactions among them (Khan et al., 2014). So, these five proteins together should carry out unique function among chlamydial spp.

All of these proteins are present in the same direction throughout the all chlamydial genomes. We have analysed these proteins to understand their operonic arrangement in the genome. We have predicted the operonic arrangement of these proteins by using the DOOR 2.0 database, which contains computationally predicted operons of the prokaryotic genomes. GyrA, GyrB, CSP, TMK and HolB proteins are present in the same polycistronic operon among 5 out of 7 chlamydial spp. In *Waddlia chondrophila* and *Simkania negevensis*; GyrA , GyrB, TMK and HolB are present in same operon but the CSP is present in the adjacent operon. The intergenic distance between two operons is less than 30 bp and they are in the same direction (Mao et al., 2014; Mao et al., 2009; Moreno-Hagelsieb, 2015). These five proteins may be present in same operon among all chlamydial spp. because these are predicted operons. Overall data indicates that these five proteins should have functional relation among themselves and could be unique characteristics of chlamydial spp.

3.4 Discussion:

The members of the phylum Chlamydiae encode three DNA topoisomerases that have been characterized for controlled DNA supercoiling level. All three topoisomerases have shown supercoiling responsiveness. Topoisomerase I and topoisomerase IV expression are controlled by negative feedback during supercoiling, which is typical of other bacteria. The promoter of gyrBA is upregulated by DNA supercoiling among chlamydial spp., which is unique molecular characteristic of topoisomerase II enzymes among chlamydial spp. (Orillard and Tan, 2016). Negative supercoiling is also considered to be a global regulator of genes during intracellular infection among chlamydial (Maxwell and Lawson, 2003; Dorman and Berlin, 2008). In this work, we have analyzed the GyrA and GyrB sequences from different organisms to bring important insights in this regard. This work has carried out comparative analyses of GyrA and GyrB proteins sequences and identified a high-specific sequence feature in the form of CSI in the GyrA and GyrB sequences. These CSIs clearly differentiate these proteins homologs of chlamydial spp. from the GyrA and GyrB homologs found in most other organisms. We also have inspected the genomic arrangement of these proteins among chlamydial spp. and other groups of bacteria.

This study reports that GyrA and GyrB homologs from chlamydial spp. differ from all other GyrA and GyrB homologs by highly-conserved sequence features. These results strongly suggest that these highly-conserved sequence characteristics should have the role in the biochemical differences in GyrA and GyrB proteins of the phylum Chlamydiae.

Structural analysis reveals that 3 out of 4 CSIs are located at surface loop of the proteins. One of the CSI of GyrA is present in a helix region of the protein. This CSI is also located at the protein surface. Earlier works on conserved signature indels provide significant evidence that CSIs play essential for the proper functioning of the proteins in the CSI-containing organisms, and the removal of the CSIs leads to loss of concerned protein biological activity or complete inactivation of the protein (Singh and Gupta, 2009). Structural analysis has indicated that 1 aa CSI of GyrA in the 517-584 sequence position is present in β-CA loop. Overall structural, sequence and nature of amino acid in the CSI position suggests that this CSI may play role in DNA binding with CTD-GyrA and also strength the interaction between blade 1 and blade 6 of GyrA protein among the Chlamydiae. The 6 aa CSI of GyrB is also present in a loop that facilitates the nucleotide binding. Our analysis strongly suggests that both of these CSIs should play role in more efficient functioning of GyrA and GyrB protein among chlamydial spp.

Another 1 aa CSI of GyrA in the 576-626 sequence position and 4 aa CSI of GyrB are located at surface loop. The surface loop constitutes highly accessible regions of the protein, are considered as determinants of interaction and these loops are responsible for mediating protein-protein and protein-ligand interactions (Akiva et al., 2008). So, these studies strongly indicate that the highly-conserved sequence characteristics of GyrA and GyrB proteins may be involved in protein- ligand and protein-protein interactions with other proteins, which may be associated with the observed differences in the biochemical characteristics of the phylum Chlamydiae. The protein-protein interactions analyses have led to concur that CSI containing GyrA and GyrB proteins are present in a unique genetic

linkage with a hypothetical protein, Tmk and HolB proteins among *Chamydiales*. We have carried out further analyses of these genetically linked proteins to understand the functional association with GyrA and GyrB proteins. Our comparative genomics studies reveal that this hypothetical protein is a Chlamydiae specific CSP. We have identified a 8 aa deletion in HolB proteins which is unique molecular characteristics of *Chlamydiales* (Figure 14B). Structural analysis of HolB protein reveals that this deletion is present in zinc module of this protein. HolB protein is one of the key players of the clamp-loading complex which is involved in the removal of γ complex from the β subunit after the clamp is placed on DNA. It is expected that zinc module of HolB is responsible for interaction with DNA. This interaction of the zinc module with DNA might introduce a mechanism for coupling DNA binding to ATP hydrolysis. So, this zinc module is critical for the proper functioning of HolB protein (Guenther et al., 1997; Podobnik et al., 2003). It may have different mechanism to function properly among *c*hlamydial spp. Tmk is considered as one of the most attractive antibacterial targets as it is responsible for catalyzing components of DNA synthesis. The TMK protein is present in the junction of de novo and salvage pathway of thymidine triphosphate (dTTP) synthesis. In addition with this, it is the last specific enzymes of dTTP synthesis (Kotaka et al., 2006). So, this enzyme plays a critical role of DNA synthesis. We have also identified 2 aa insertion in the TMK protein sequence which is specific for *Chlamydia* spp. (Figure 14A). This CSI may have some unique functional role for this group of bacteria. Overall, GyrA, GyrB, HolB and Tmk proteins are associated with DNA synthesis; however the functional role of CSP is unknown. The Operonic analysis suggests that these five proteins should also

present in the same polycistronic operon among chlamydial spp. These five proteins may carry out unique functions together among chlamydial spp. During Intracellular development of the phylum Chlamydiae, the compacted EB DNA is relaxed and signals for DNA, RNA, and protein synthesis (Wyrick, 2010). The midcycle genes that are transcribed during RB replication, are upregulated by negative supercoiling. Negative supercoiling also upregulates the gyrBA promoters. This upregulation is only unique to the members of the phylum Chlamydiae. Our study suggests that these five proteins should be transcribed from the same promoter. So, HolB, TMK and CSP gene expression also should be upregulated during negative supercoiling. It has been already known that genes that are involved in DNA replication are upregulated at active stage of Chlamydial intracellular development (Wyrick, 2010). We have proposed a model to understand the functional role of these five proteins together in chlamydial intracellular development (Figure 20).

When infectious EB converts into metabolically active RB, the compacted DNA becomes relaxed. The negative supercoiling levels peak at this stage of chlamydial development. The expression of gyrBA promoter upregulates by positive feedback control. The expression of CSP, GyrA, GyrB, holB and TMK should be upregulated because these five proteins are transcribed by the same promoter. The CSP may provide evolutionary advantage among Chlamydiae spp., because promoter for these five proteins are located at upstream region of CSP. It may also act as a linker among these proteins during the replication. GyrA and GyrB proteins enhance the negative supercoiling which increase the replication rate. To increase the replication rate, the expression of DNA polymerase

III holoenzymes is important. HolB is one of the holoenzyme of clamp-loader complex. TMK protein is responsible for the components of DNA synthesis. So, these five proteins together may play role in increasing replication rate in midcycle genes. GyrA and GyrB may enhance the functional role of HolB protein because it is lacking the zinc module among the members of the phylum Chlamydiae. It has been proposed that DNA supercoiling also induces the expression of late midcycle topoisomerase I and topoisomerase IV, which leads to unbalance the gyrase activity. It down-regulates the expression of gyrBA promoters and also the expression of HolB and TMk at the late cycle of chlamydial development. So, replication rate becomes lower during the conversion of RB to EB. This model provides molecular insight of chlamydial intracellular development. The GyrA and GyrB proteins with the association of a CSP are regulating the functional activity of TMK and HolB proteins. The association of CSP with GyrA and GyrB may be mediated by 1 aa CSI of GyrA in the 576-626 sequence position and 4 aa CSI of GyrB. Structural analyses of GyrA and GyrB proteins indicate that GyrA and GyrB proteins of chlamydial spp. should carry out more supercoiling activity due to presence of highly conserved sequence features.

A number of elements of this model need to proof by further experiment. First of all, these five proteins are present in the same operon among the members of the phylum Chlamydiae based on predicted operonic arrangement by DOOR.2 database. A recent study has been revealed that the promoter of gyrA and gyrA are present the upstream site of CSP in *Chlamydia tractomatis* (Orillard and Tan, 2016). It does not indicate anything about downstream part of this operon. So, further studies are required to understand their

operonic arrangement. Though, phylogenetic profiles, direction of the genes, neighborhood and intergenic distance analyses already clearly indicate that these five should have some functional or physical interactions among the chlamydial spp. There is no experimental study carried out to understand the functional association of these five proteins together. Further experimental studies are expected to validate our proposed models. To understand the functional role of CSP, we could carry out knock out study. The role of topoisomerase I and topoisomerase IV in reducing the negative supercoiling at late cycling of intracellular development has been proposed in a recent study. It is already known that both enzymes have a capacity to relax DNA. In addition to this, temporal correlation between topoisomerase I and topoisomerase IV expression and chlamydial DNA relaxation indicates that these two enzymes may involve unbalancing the gyrase activity.

This work provides a clear hint to understanding the role of topoisomerase II enzymes in chlamydial intracellular development. We have identified unique molecular characteristics known as CSI in the GyrA and GyrB proteins specific for the members of the phylum Chlamydiae. Two of these CSIs should play role in increasing supercoiling activity during the intracellular development. We also have identified unique genetic linkage of GyrA and GyrB with a CSP, HolB and TMK proteins and proposed a model to understand their functional or physical association. Our proposed model is supported by several studies (Orillard and Tan, 2016; Wyrick, 2010). Our identified CSIs and proposed model may have implications to identify drug targets for the treatment of chlamydial infections.

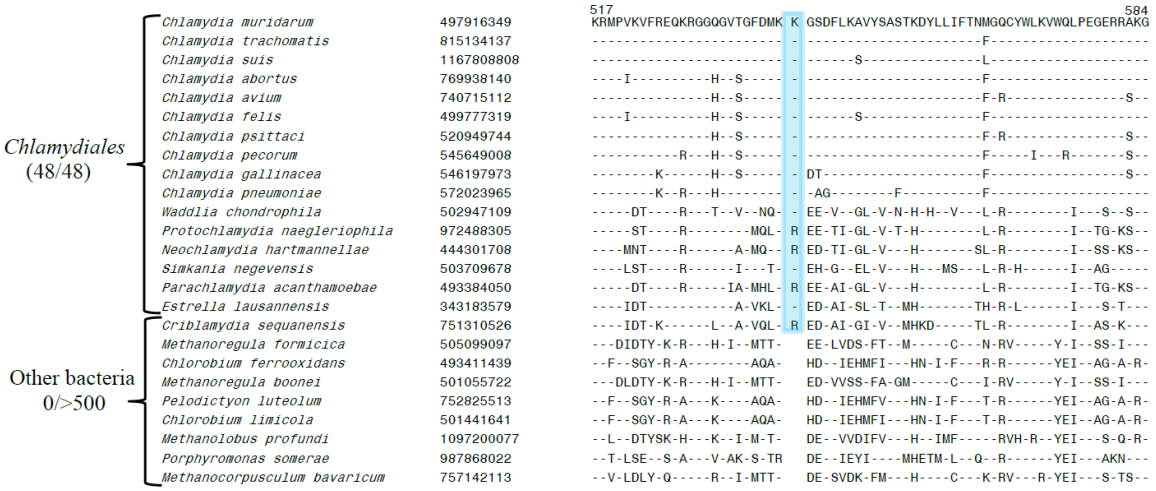


Figure 12A: Partial sequence alignment of conserved region of DNA gyrase subunit A (GyrA) protein showing a 1 amino acid insertion that is uniquely shared by all *Chlamydiales*. Dashes in the sequence alignment indicate identity to the amino acid on the top line. The accession numbers of the sequences and the position of this region in the *Chlamydia muridarum* are shown. Only sequence information for representative species is presented.

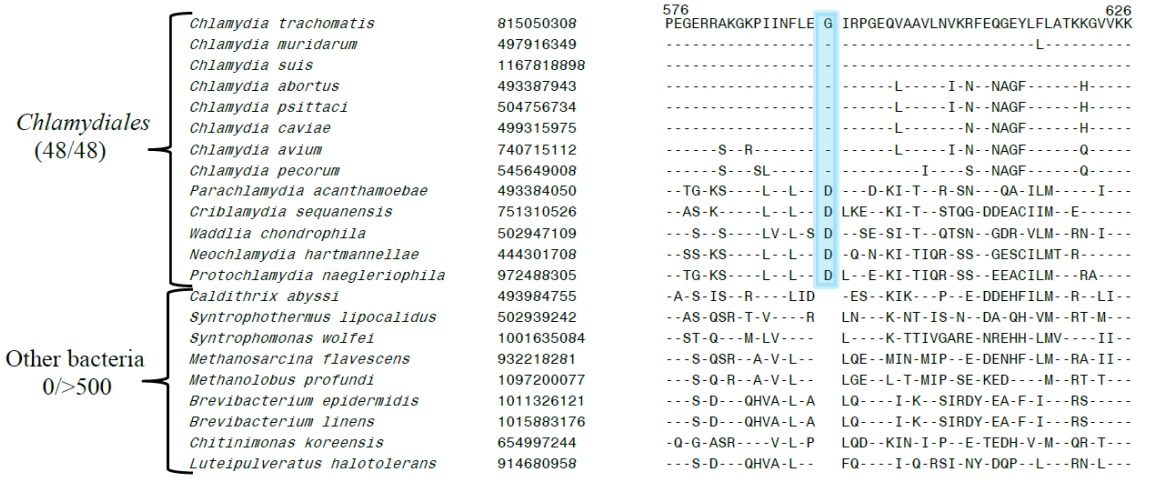


Figure 12B: Partial sequence alignment of conserved region of DNA gyrase subunit A (GyrA) protein showing a 1 amino acid insertion that is specific for all *Chlamydiales*.

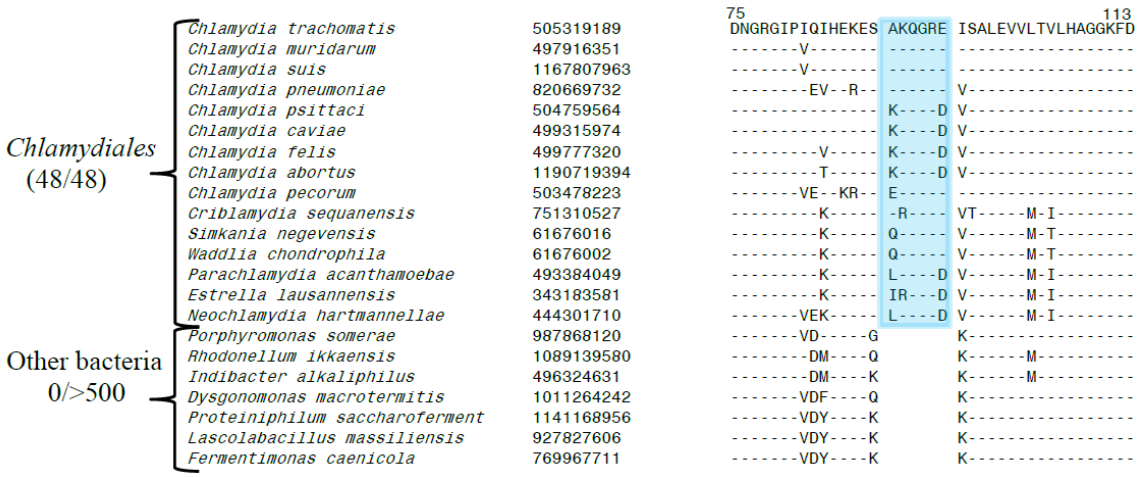


Figure 13A: Partial sequence alignment of conserved region of DNA gyrase subunit B (GyrB) protein showing a 6 amino acid insertion that is a distinctive characteristic from homologs of *Chlamydiales*.

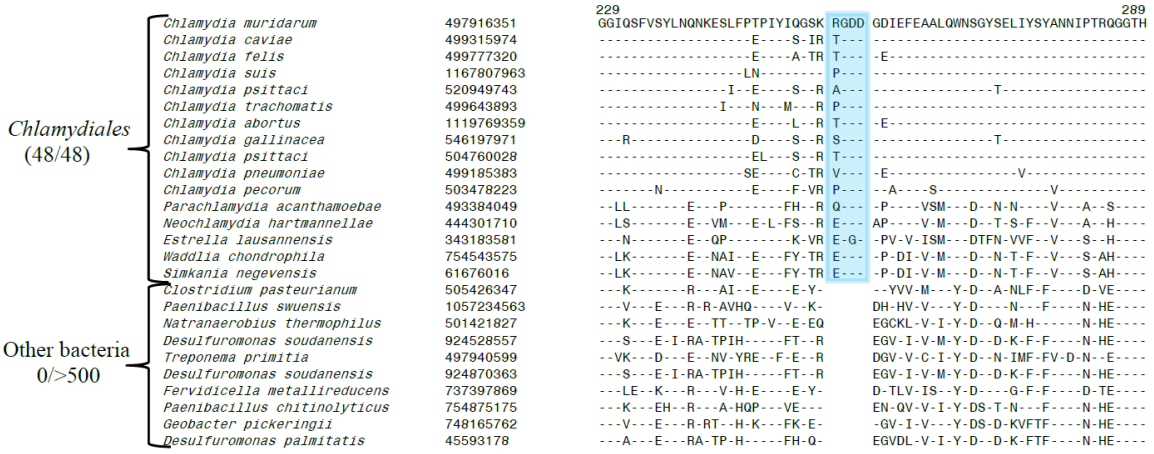


Figure 13B: Partial sequence alignment of conserved region of DNA gyrase subunit B (GyrB) protein showing a 4 amino acid insertion that is uniquely shared by all *Chlamydiales*.

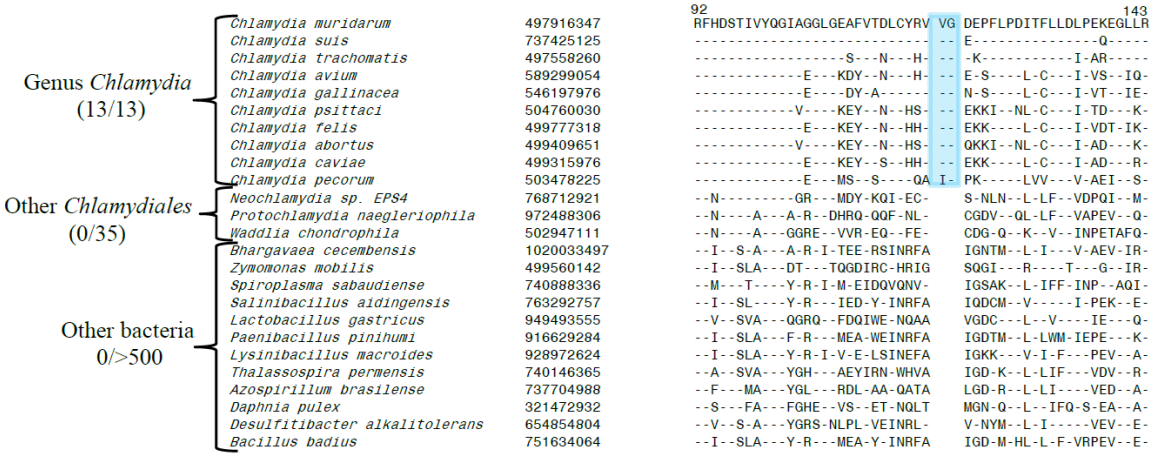


Figure 14A: Partial sequence alignment of conserved region of thymidylate kinase (TMK) protein showing a 2 amino acid insertion that is uniquely shared by all *Chlamydia* spp.

Figure 14B: Partial sequence alignment of conserved region of DNA polymerase III subunit δ' (HolB) protein showing a 8 amino acid deletion that is the distinctive molecular characteristics of *Chlamydiales* spp..

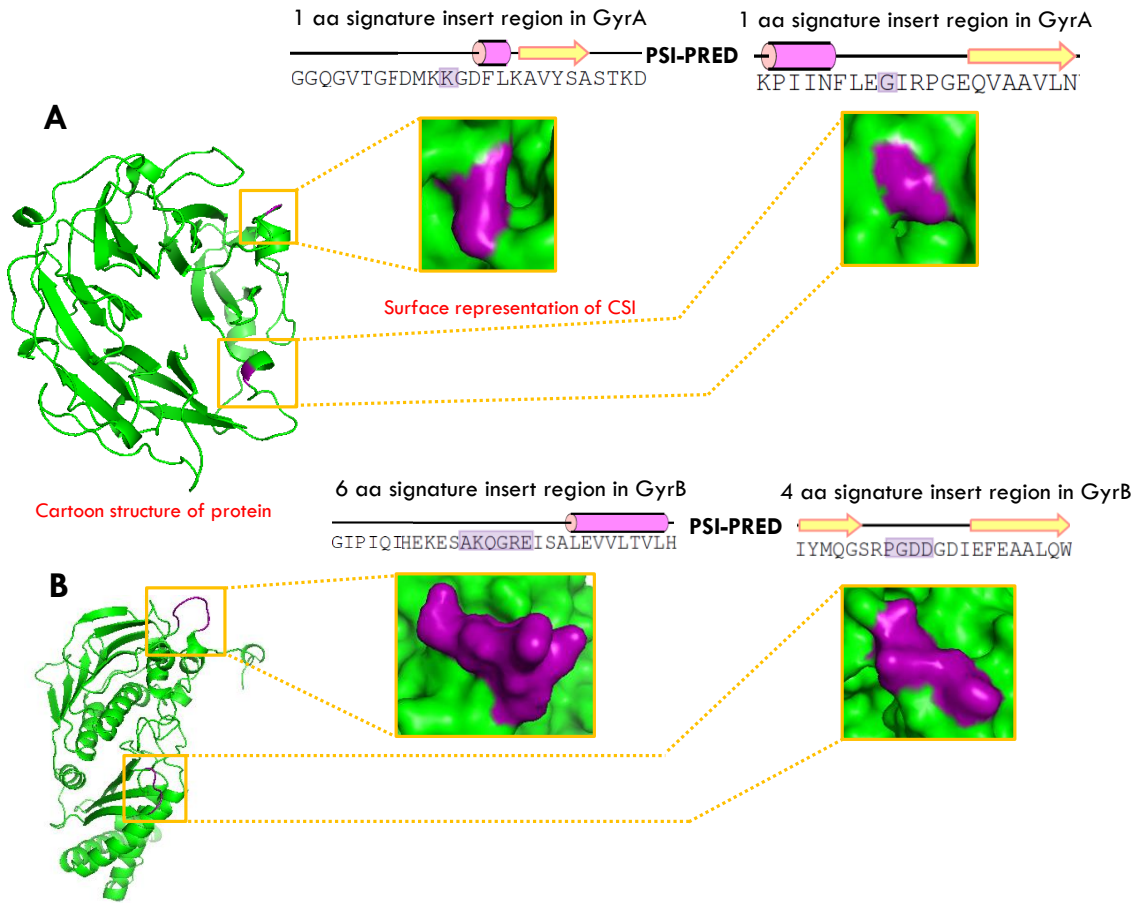


Figure 15: A) Location of two 1 aa CSIs in the modelled structure of GyrA protein. One of the CSI is present in helix region of the protein and another one is located at loop region. Surface location analysis delineates that both of these CSIs are located at the surface of protein. B) Location of 6 aa and 4 aa CSIs in modeled GyrB protein. Structural analysis indicates that these CSIs are located at surface loop region in the modeled GyrB. Secondary structure of CSIs regions has been predicted by PSI-PRED server. The templates for homology modeling of the C-terminal domain of GyrA and the N-terminal ATPase domain of GyrB from *Chlamydia trachomatis* have been selected from following species sequentially: *Mycobacterium tuberculosis* (PDB id: 3UC1), *Escherichia coli* (PDB id: 1EI1). All CSIs regions are mapped in the sequence by purple color.

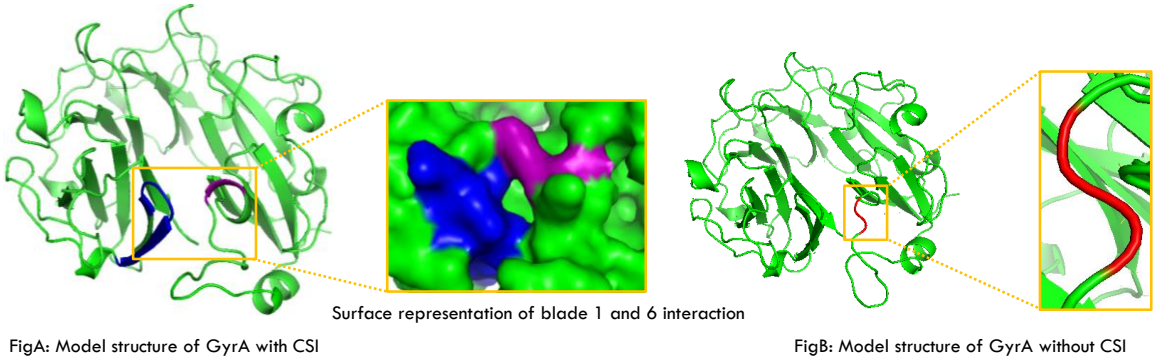


Figure 16: A) Structural and Functional analyses of 1 aa CSI in the 517–584 sequence region of GyrA protein. This CSI is located at the β-CA loop of blade 1. Surface structural analysis of this CSI indicates that this CSI may strength the interactions between blade 1 and blade 6. B) Homology modelled structure of GyrA protein without CSI. Interestingly, modeled GyrA protein without CSI does not contain the helix region which has intimate contact with blade 6. This result strongly suggests that this CSI should have role in maintain topological activity of GyrA protein of *Chlamydiales* spp.. Blade 6 has marked by blue color and 1 aa CSI region has marked by purple color. In CSI lacking modeled GyrA, red colored region is the residues beside the CSI.

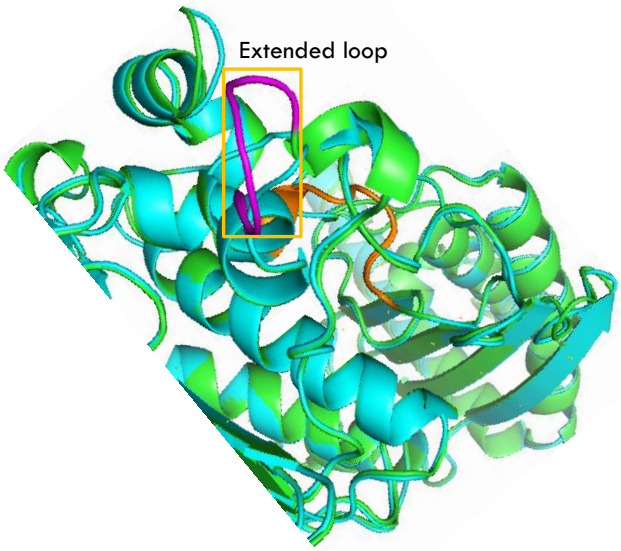


Figure 17: Structural alignment of modeled 6 aa CSI containing GyrB protein with CSI lacking GyrB protein. Structural analysis reveals that This CSI which is marked by purple color in the figure is responsible for making extended loop which may facilitate more nucleotide binding at nucleotide binding site. Nucleotide binding site has been marked by orange color.

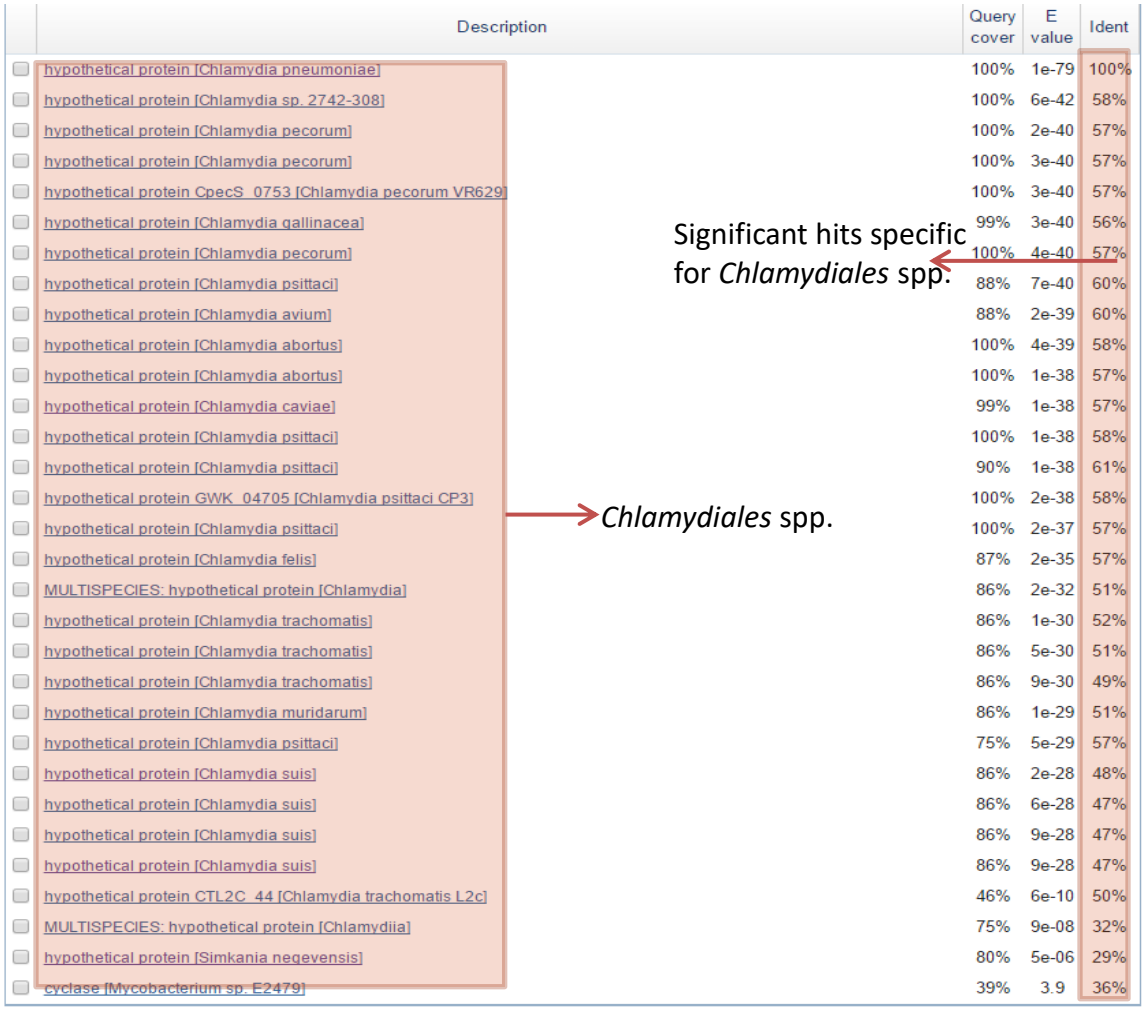


Figure 18: Diagrammatic representation of the *Chlamydiales* specific CSP. Blastp search was carried out against in the genome of *Chlamydia trachomatis* against all available sequences. Blast search results delineates that this CSP is unique for *Chlamydiales* spp. Though one out-group bacteria was found in blast search. But expected value (E-value) of other hits is so much lower from the E-value for *Chlamydiales* spp.

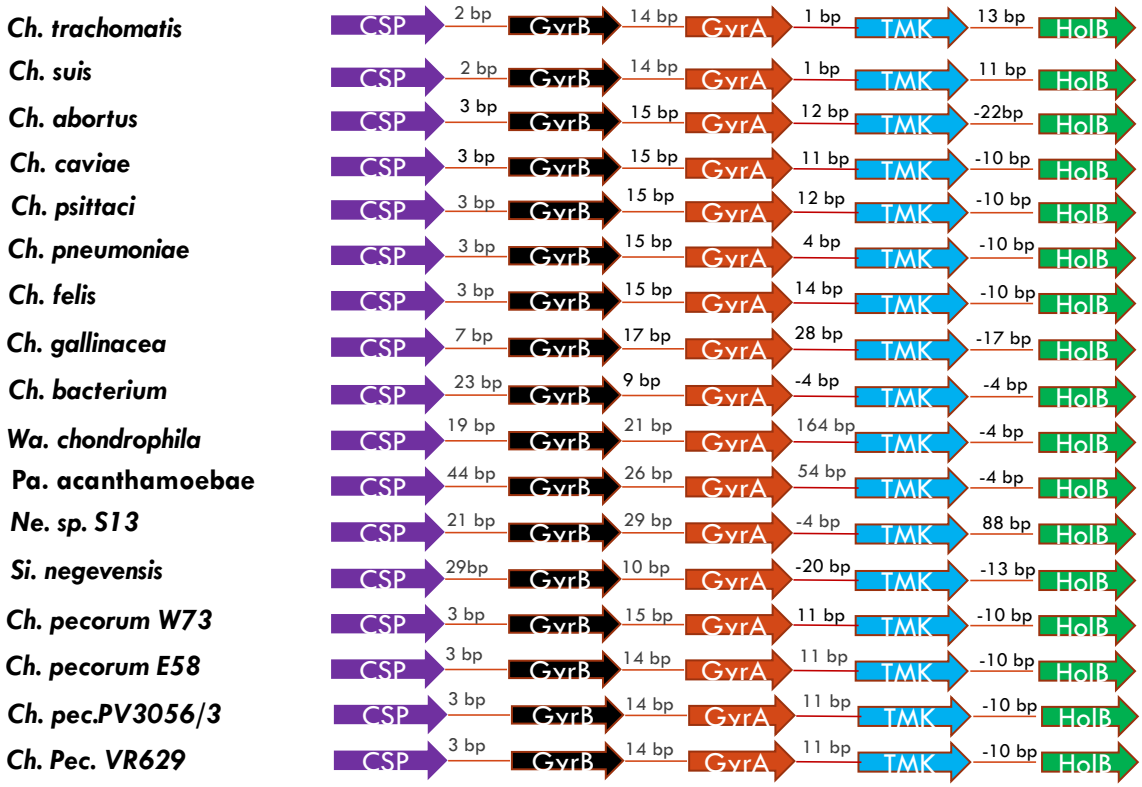


Figure 19: A diagrammatic representation of genomic neighborhood analysis result of GyrA and GyrB proteins for representative *Chlamydiales* spp. GyrA and GyrB proteins are present genetic linkage with a chlamydial specific CSP, TMK, and HolB protein proteins among all *Chlamydiales* spp. These five proteins are also present in the same direction in the *Chlamydiales* genome. Intergenic distances between proteins are also shown in the diagram. Intergenic distances between proteins are less than 200 base pairs in all the species. So, all five proteins should have same phylogenetic profile and present in the same operon.

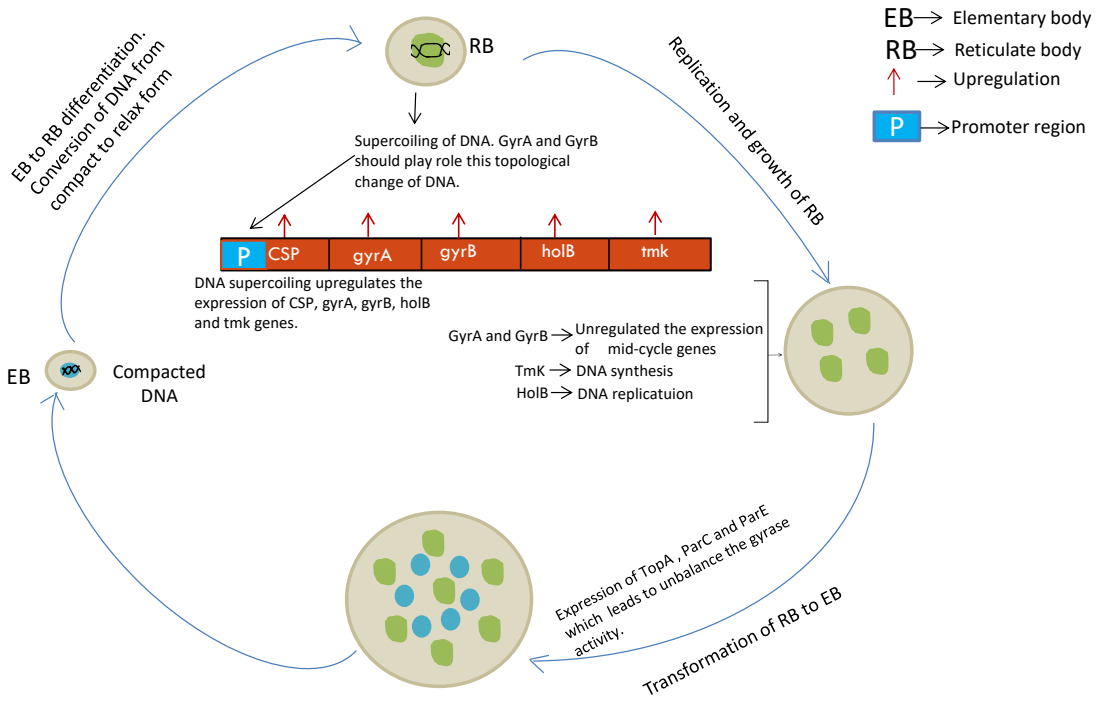


Figure 20: Proposed model for chlamydial intracellular development. Firstly, elementary body (EB) converts into the reticulate body (RB). The DNA becomes relaxed and supercoiled. GyrA and GyrB should involve in this topological change of the DNA. DNA supercoiling upregulates the gyrA, gyrB CSP, holB and tmk genes. CSP, holB and tmk genes are also present in the same operon with gyrA and gyrB. So, these five proteins together play role in replication and growth of RB during mid-cycle of intracellular development. topA, parC and parE genes are expressed the late mid-cycle which leads to unbalance the gyrase activity. RB becomes converts to EB and starts another development cycle. CSP is playing role to make functional bridge among these proteins.

CHAPTER 4

Conclusions

4.1 The impact of protein-protein interactions analyses to understand the functional role of CSIs and CSPs:

CSIs are rare genetic changes found in the conserved regions of proteins which are found to be essential for the bacterial growth. So, CSI should be carried out unique function for the certain group of bacteria (Gupta, 2014; Gupta, 1998; Singh and Gupta, 2009). To understand the functional role of CSI, it is important to map the location of CSI in the protein structure. But three-dimensional structures of most of the proteins have not crystalized yet, we have developed three dimensional model of protein using homology modeling approach (Sali and Blundell, 1993; Eswar et al., 2007). It constructs an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental 3D structure of a related homologous protein (the "template"). Suitable templates for homology modeling were identified by a PSI-Blast search (Altschul et al., 1997).The locations of CSIs are mapped in the modeled protein structures. Structural analysis of CSIs reveals that most of these CSIs are found to present at the surface loop region of the proteins (Gupta et al., 2017; Alnajar et al., 2017). Surface loop region of the protein is likely to mediate protein-protein and protein-ligand interactions (Akiva et al., 2008). CSP is unique protein that is present in certain group of microorganism (Griffiths and Gupta, 2007; Gupta and Griffiths, 2006; Ho et al., 2016). The CSPs which are specific for *Deinococcus* spp. include a number of unique proteins with known functional roles in the radiation resistance and DNA repair mechanisms of *Deinococcus* spp. such as the DNA damage repair protein PprA (Accession: NP_285669) (Narumi et al., 2004; Tanaka et al., 2004) and the single-stranded DNA-binding protein DdrB (Accession:

NP_293796) (Sugiman-Marangos et al., 2016; Xu et al., 2010). The work presented in this thesis has carried out comparative genomics analysis to identify the molecular marker in term of CSIs and CSPs and have analyzed these distinctive molecular characteristics using different in-silico protein-protein interactions analyses approach to understand their functional role.

4.2 The utilization of molecular signatures to understand the radiation resistance of *Deinococcus* spp.:

The *Deinococcus* spp. are present in the phylum Deionococcus-Thermus which has been characterized by extreme resistance to oxidative damage induced by radiation and desiccation (Slade and Radman, 2011). According to Krisko, A. and Radman, M. (2010), *Deinococcus* spp. are showing remarkable robustness due to their efficient proteome protection (but not DNA protection) and cell death by radiation is caused by oxidative damage with the resultant loss of maintenance activities including DNA repair (Krisko and Radman, 2010). The *Deinococcus* spp. should contain unique DNA repair enzymes that could provide the evolutionary advantage of DNA repair mechanism in this group bacteria (Slade and Radman, 2011). In this work, we have carried out comparative genomics analysis of DNA repair proteins to find out CSIs that are distinctive molecular characteristics for DNA repair proteins of *Deinococcus* spp. We have identified a large number of CSIs in the DNA repair proteins of *Deinococcus* spp. Interestingly, we have identified a large 30 aa CSI in the UvrA protein of *Deinococcus* spp. Protein-protein interactions analyses completed has led to concur that UvrA protein interacts with dCSP-

I, DsbB and DsbA proteins among *Deinococcus* spp.. This genetic linkage of UvrA protein with these three proteins is only specific for *Deinococcus* spp. Furthermore, we have carried out the functional, structural, and cellular localization analyses of these genetically linked proteins to decipher their functional role together in DNA repair. Finally, we have proposed a model to understand the functional role of the CSI and CSP. According to our model, *Deinococcus*-specific CSP by interacting with the CSI in UvrA may act as a functional bridge among UvrA and the DsbA and DsbB proteins. The predicted linkage of the UvrA protein via dCSP-1 to the DsbA/DsbB is expected to protect the UvrA protein from UV-induced oxidative damage, which facilitates the UvrA protein to restore its function in *Deinococcus* spp. upon radiation exposure. Experimental studies are required to validate the few elements of this model. CSP is playing role to make the functional linkage among UvrA1, DsbA, DsbB proteins. Inactivation or inhibition of dCSP-I protein may guide to understand the functional role of the model in NER pathways. Experimental studies could also bring new insights to understand the functional association among UvrA, DsbA and DsbB proteins.

Our observation indicates that identified novel molecular traits (CSIs, CSPs) in DNA repair proteins may involve in oxidative stress resistance induced by radiation, desiccation and chemical mutagens among *Deinococcus* spp. Further functional analysis of other CSIs may provide valuable information about the functional role of DNA repair proteins during oxidative damage.

4.3 Studies on Chlamydiae-specific CSIs and CSPs to understand their possible cellular functions:

Chlamydial spp. possess a unique, biphasic development cycle that may unite the various species belonging to the phylum Chlamydiae, since they possess such varying hosts and disease development (Ward, 1983; Abdelrahman and Belland, 2005; Zhong et al., 2001). Interestingly, chlamydial encode small genome of 1-2 Mb yet interact in a complex manner upon host interaction (Sisko et al., 2006b). Identification of chlamydial specific CSIs and CSPs may serve as the useful tool in gaining the better understanding of the pathogenesis in this group of bacteria (Sisko et al., 2006a; Griffiths et al., 2005). It has been proposed that negative supercoiling could play the role for the temporal expression of genes during the intracellular development (Orillard and Tan, 2016). In this study, we have carried out the comparative genomics study of GyrA and GyrB proteins to identify the unique sequence features which will be specific for chlamydial. We have identified several CSIs in these proteins which are distinctive molecular characteristics for all chlamydial spp. The structural and functional analyses reveal that two of these CSIs should play role to enhance the supercoiling activity among this group of bacteria. We have carried out the in-silico protein-protein interactions analyses to understand the functional role of other two CSIs. Protein-protein interactions analyses reveal that GyrA and GyrB proteins are present in a unique genetic linkage with three different proteins. One of the interacting proteins is a chlamydial specific CSP. To understand the functional role of these proteins in the intracellular development, we have proposed a model. According to our model, GyrA and GyrB proteins by the association with Tmk, HolB,

and CSP involve in increasing the replication rate of the mid-cycle genes. Both of these CSIs could play role to interact with these genetically linked proteins during intracellular development.

A number of elements need experimental supports to validate this model. Operonic analysis is performed based on the operonic arrangement by DOOR.2 database. The promoter of gyrA and gyrB genes is located at the upstream site of CSP in *Chlamydia tractomatis* (Orillard and Tan, 2016). But this study does not indicate anything about the downstream part of this operon. So, further studies are essential to understand their operonic arrangement. The functional role of CSP is unknown. This research indicates that this protein should have functionally related to DNA synthesis, but further studies are required. CSP is the central protein of our proposed model. We could knock out the CSP to understand the functional role of models. Furthermore, we have proposed the functional role of two of the CSIs based on structural analysis. The mutational analysis could be carried out to decipher the functional role of the CSI regions.

This work provides molecular insights about the unique role of topoisomerase II enzymes in chlamydial spp. Our proposed model is supported by several studies (Orillard and Tan, 2016;Wyrick, 2010) which may have implications to decipher the intricateness of chlamydial intracellular development.

Concluding Remark:

In summation, the in-silico protein-protein interactions analyses could be one of the approaches to predict the functional role of CSIs and CSPs. But this approach could not

be used to predict the functional role of all CSIs and CSPs specific for different groups of bacteria. In my research, I have analyzed a large number of CSIs and CSPs for different groups of bacteria. But I have found very few genetic linkages between CSIs containing protein or CSPs with other proteins which are only specific for CSIs and CSPs containing bacteria. Finally, this approach may guide to predict the functional role of CSIs and CSPs if we could identify unique genetic linkage between CSI containing protein, CSPs with other proteins for the specific group of bacteria.

Reference List

Abdelrahman,Y.M. and Belland,R.J. (2005). The chlamydial developmental cycle. FEMS Microbiol. Rev. *29*, 949-959.

Achouak,W., and Armengaud,J. (2009). Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. PLoS Genet. *5*, e1000434.

Adeolu,M., Alnajar,S., Naushad,S., and Gupta,S. (2016). Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. Int J Syst Evol. Microbiol. *66*, 5575-5599.

Agostini,H.J., Carroll,J.D., and Minton,K.W. (1996). Identification and characterization of uvrA, a DNA repair gene of *Deinococcus radiodurans*. J Bacteriol *178*, 6759-6765

Akiva,E., Itzhaki,Z., and Margalit,H. (2008). Built-in loops allow versatility in domain-domain interactions: lessons from self-interacting domains. Proc. Natl. Acad. Sci. U. S A *105*, 13292-13297.

Alnajar,S. and Gupta,R.S. (2017). Phylogenomics and comparative genomic studies delineate six main clades within the family *Enterobacteriaceae* and support the reclassification of several polyphyletic members of the family. Infect. Genet. Evol. *54*, 108-127.

Alnajar,S., Khadka,B., and Gupta,R.S. (2017). Ribonucleotide Reductases from Bifidobacteria Contain Multiple Conserved Indels Distinguishing Them from All Other Organisms: In Silico Analysis of the Possible Role of a 43 aa Bifidobacteria-Specific Insert in the Class III RNR Homolog. Front Microbiol. *8*, 1409.

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W., and Lipman,D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs Nucleic Acids Res. *25*, 3389-3402

Balin,B.J., Gerard,H.C., Arking,E.J., Appelt,D.M., Branigan,P.J., Abrams,J.T., Whittum-Hudson,J.A., and Hudson,A.P. (1998). Identification and localization of *Chlamydia pneumoniae* in the Alzheimer's brain. Med. Microbiol. Immunol. *187*, 23-42.

Barry,C.E., III, Hayes,S.F., and Hackstadt,T. (1992). Nucleoid condensation in *Escherichia coli* that express a chlamydial histone homolog. Science *256*, 377-379.

Batteiger,B.E., Wan,R., Williams,J.A., He,L., Ma,A., Fortenberry,J.D., and Dean,D. (2014). Novel *Chlamydia trachomatis* strains in heterosexual sex partners, Indianapolis, Indiana, USA. Emerg. Infect. Dis. *20*, 1841-1847.

Battista,J.R. (1997). Against all odds: the survival strategies of *Deinococcus radiodurans*. Annu. Rev. Microbiol. *51*, 203-224.

Belland,R.J., Zhong,G., Crane,D.D., Hogan,D., Sturdevant,D., Sharma,J., Beatty,W.L., and Caldwell,H.D. (2003). Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. Proc. Natl. Acad. Sci. U. S A *100*, 8478-8483.

Benkert,P., Tosatto,S.C., and Schomburg,D. (2008). QMEAN: A comprehensive scoring function for model quality assessment.Proteins *71*, 261-277.

Bentchikou,E., Servant,P., Coste,G., and Sommer,S. (2010). A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in *Deinococcus radiodurans*. PLoS Genet. *6*, e1000774.

Betts,H.J., Wolf,K., and Fields,K.A. (2009). Effector protein modulation of host cells: examples in the Chlamydia spp. arsenal. Curr. Opin. Microbiol. *12*, 81-87.

Betts-Hampikian,H.J. and Fields,K.A. (2010). The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. Front Microbiol. *1*, 114.

Bhandari,V. and Gupta,R.S. (2014). Molecular signatures for the phylum (class) Thermotogae and a proposal for its division into three orders (*Thermotogales*, *Kosmotogales* ord. nov. and *Petrotogales* ord. nov.) containing four families (*Thermotogaceae*, *Fervidobacteriaceae* fam. nov., *Kosmotogaceae* fam. nov. and *Petrotogaceae* fam. nov.) and a new genus *Pseudothermotoga* gen. nov. with five new combinations. Antonie Van Leeuwenhoek *105*, 143-168.

Brickman,T.J., Barry,C.E., III, and Hackstadt,T. (1993). Molecular cloning and expression of hctB encoding a strain-variant chlamydial histone-like protein with DNA-binding activity. J Bacteriol *175*, 4274-4281.

Brunham,R.C., Laga,M., Simonsen,J.N., Cameron,D.W., Peeling,R., McDowell,J., Pamba,H., Ndinya-Achola,J.O., Maitha,G., and Plummer,F.A. (1990). The prevalence of *Chlamydia trachomatis* infection among mothers of children with trachoma. Am. J Epidemiol. *132*, 946-952.

Brino,L., Urzhumtsev,A., Mousli,M., Bronner,C., Mitschler,A., Oudet,P., and Moras,D. (2000). Dimerization of *Escherichia coli* DNA-gyrase B provides a structural mechanism for activating the ATPase catalytic center. J Biol. Chem. *275*, 9468-9475.

Brena-Valle,M. and Serment-Guerrero,J. (1998). SOS induction by gamma-radiation in Escherichia coli strains defective in repair and/or recombination mechanisms. Mutagenesis *13*, 637-641.

Burrell,A.D., Feldschreiber,P., and Dean,C.J. (1971). DNA-membrane association and the repair of double breaks in x-irradiated Micrococcus radiodurans. Biochim. Biophys. Acta *247*, 38-53.

Buchan,D.W., Minneci,F., Nugent,T.C., Bryson,K., and Jones,D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. *41*, W349-W357.

Bush,N.G., Evans-Roberts,K., and Maxwell,A. (2015). DNA Topoisomerases. EcoSal. Plus. *6*.

Carabeo,R.A., Mead,D.J., and Hackstadt,T. (2003). Golgi-dependent transport of cholesterol to the *Chlamydia trachomatis* inclusion. Proc. Natl. Acad. Sci. U. S A *100*, 6771-6776.

Carroll,D. (1996). Homologous genetic recombination in Xenopus: mechanism and implications for gene manipulation. Prog. Nucleic Acid Res. Mol. Biol. *54*, 101-125.

CDC. Summary of notifiable diseases – United States, 2012. MMWR. 2014; 61:1–121. [PubMed: 25233134]

Champoux,J.J. (2001). DNA topoisomerases: structure, function, and mechanism. Annu. Rev. Biochem. *70*, 369-413.

Chan,H.L., Gaffney,P.R., Waterfield,M.D., Anderle,H., Peter,M.H., Schwarz,H.P., Turecek,P.L., and Timms,J.F. (2006). Proteomic analysis of UVC irradiation-induced damage of plasma proteins: Serum amyloid P component as a major target of photolysis.FEBS Lett. *580*, 3229-3236.

Chandra,G. and Chater,K.F. (2014). Developmental biology of Streptomyces from the perspective of 100 actinobacterial genome sequences. FEMS Microbiol. Rev. *38*, 345-379

Cheng,E. and Tan,M. (2012). Differential effects of DNA supercoiling on *Chlamydia* early promoters correlate with expression patterns in midcycle. J Bacteriol *194*, 3109-3115.

Chim,N., Harmston,C.A., Guzman,D.J., and Goulding,C.W. (2013). Structural and biochemical characterization of the essential DsbA-like disulfide bond forming protein from *Mycobacterium tuberculosis*. BMC Struct. Biol *13*, 23.

Colovos,C. and Yeates,T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. *2*, 1511-1519.

Corsaro,D. and Greub,G. (2006). Pathogenic potential of novel *Chlamydiae* and diagnostic approaches to infections due to these obligate intracellular bacteria. Clin. Microbiol. Rev. *19*, 283-297.

Cram,E.D., Rockey,D.D., and Dolan,B.P. (2017). *Chlamydia* spp. development is differentially altered by treatment with the LpxC inhibitor LPC-011. BMC. Microbiol. *17*, 98.

Daly,M.J. (2009). A new perspective on radiation resistance based on *Deinococcus radiodurans*. Nat. Rev. Microbiol *7*, 237-245.

Daly MJ (2012) Death by protein damage in irradiated cells. DNA repair, Vol.11 (1):12–21

Dam,P., Olman,V., Harris,K., Su,Z., and Xu,Y. (2007). Operon prediction using both genome-specific and general genomic information. Nucleic Acids Res. *35*, 288-298.
Daubin,V., Lerat,E., and Perriere,G. (2003). The source of laterally transferred genes in bacterial genomes. Genome Biol. *4*, R57.

Danchin,A. (1999). From protein sequence to function. Curr. Opin. Struct. Biol. *9*, 363-367.

de Groot,A., Dulermo,R., Ortet,P., Blanchard,L., Guerin,P., Fernandez,B., Vacherie,B., Dossat,C., Jolivet,E., Siguier,P., Chandler,M., Barakat,M., Dedieu,A., Barbe,V., Heulin,T., Sommer,S.,
Dorman,C.J. (1991). DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. Infect. Immun. *59*, 745-749.

Dorman CJ, Berlin: Regulation of transcription in bacteria by DNA supercoiling. Bacterial Physiology. Edited by: El-Sharoud W. 2008, Heidelberg: Springer

Duncan,M.C., Linington,R.G., and Auerbuch,V. (2012). Chemical inhibitors of the type three secretion system: disarming bacterial pathogens. Antimicrob. Agents Chemother. *56*, 5433-5441.

Dutilh,B.E., Snel,B., Ettema,T.J., and Huynen,M.A. (2008). Signature genes as a phylogenomic tool. Mol. Biol. Evol. *25*, 1659-1667.

Eisenberg,D., Luthy,R., and Bowie,J.U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol. *277*, 396-404.

Eisen,J.A. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J Mol. Evol. *41*, 1105-1123.

Ezraty,B., Gennaris,A., Barras,F., and Collet,J.F. (2017). Oxidative stress, protein damage and repair in bacteria. Nat. Rev. Microbiol. *15*, 385-396

Eswar,N., Webb,B., Marti-Renom,M.A., Madhusudhan,M.S., Eramian,D., Shen,M.Y., Pieper,U., and Sali,A. (2007). Comparative protein structure modeling using MODELLER. Curr. Protoc. Protein Sci. *Chapter 2*, Unit.

Everett,K.D., Bush,R.M., and Andersen,A.A. (1999). Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. Int J Syst Bacteriol *49 Pt 2*, 415-440.

Fang,G., Rocha,E., and Danchin,A. (2005). How essential are nonessential genes? Mol. Biol. Evol. *22*, 2147-2156.

Fang,H., Yu,Y.L., Wang,X.G., Chu,X.Q., Pan,X.D., and Yang,X.E. (2008). Effects of repeated applications of chlorpyrifos on its persistence and soil microbial functional diversity and development of its degradation capability. Bull. Environ. Contam Toxicol. *81*, 397-400.

Galan,J.E. and Wolf-Watz,H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. Nature *444*, 567-573.

Gao,B. and Gupta,R.S. (2012a). Microbial systematics in the post-genomics era. Antonie Van Leeuwenhoek *101*, 45-54.

Gao,B. and Gupta,R.S. (2012b). Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. Microbiol. Mol. Biol. Rev. *76*, 66-112.

Gao,B., Paramanathan,R., and Gupta,R.S. (2006). Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. Antonie Van Leeuwenhoek *90*, 69-91.

Garrity GM, Holt JG (2001) Phylum BIV. “Deinococcus–Thermus”. In: Boone D, Castenholz R (eds) Bergey’s manual of systematic bacteriology, 2nd edn, vol 1. Springer, New York, pp 395-420

Gerbase,A.C., Rowley,J.T., and Mertens,T.E. (1998). Global epidemiology of sexually transmitted diseases. Lancet *351 Suppl 3*, 2-4.

Gellert,M., Mizuuchi,K., O'Dea,M.H., and Nash,H.A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. Proc. Natl. Acad. Sci. U. S A *73*, 3872-3876.

Ghosal,D., Omelchenko,M.V., Gaidamakova,E.K., Matrosova,V.Y., Vasilenko,A., Venkateswaran,A., Zhai,M., Kostandarithes,H.M., Brim,H., Makarova,K.S., Wackett,L.P., Fredrickson,J.K., and Daly,M.J. (2005). How radiation kills cells: survival of *Deinococcus radiodurans* and *Shewanella oneidensis* under oxidative stress. FEMS Microbiol Rev. *29*, 361-375.

Griffiths,E. and Gupta,R.S. (2007). Identification of signature proteins that are distinctive of the Deinococcus-Thermus phylum. Int Microbiol *10*, 201-208.

Griffiths,E., Petrich,A.K., and Gupta,R.S. (2005). Conserved indels in essential proteins that are distinctive characteristics of *Chlamydiales* and provide novel means for their identification. Microbiology *151*, 2647-2657

Griffiths,E. and Gupta,R.S. (2004). Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the deinococcus-thermus phylum. J Bacteriol *186*, 3097-3107.

Guenther,B., Onrust,R., Sali,A., O'Donnell,M., and Kuriyan,J. (1997). Crystal structure of the delta' subunit of the clamp-loader complex of *E. coli* DNA polymerase III. Cell *91*, 335-345.

Gupta,R.S. (1998). Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol. Mol. Biol. Rev. *62*, 1435-1491.

Gupta,R.S. (2010). Molecular signatures for the main phyla of photosynthetic bacteria and their subgroups. Photosynth. Res. *104*, 357-372.

Gupta, R.S. (2014) Identification of Conserved Indels that are Useful for Classification and Evolutionary Studies. In: Michael Goodfellow, Iain Sutcliffe and Jongsik Chun, editors, Methods in Microbiology, Oxford: Academic Press. Vol. 41:153-182.

Gupta,R.S. (2016). Molecular signatures that are distinctive characteristics of the vertebrates and chordates and supporting a grouping of vertebrates with the tunicates. Mol. Phylogenet. Evol. *94*, 383-391.

Gupta,R.S. (2016). Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. FEMS Microbiol. Rev. *40*, 520-553.

Gupta,R.S. and Griffiths,E. (2006). Chlamydiae-specific proteins and indels: novel tools for studies. Trends Microbiol. *14*, 527-535.

Gupta,R.S., Nanda,A., and Khadka,B. (2017). Novel molecular, structural and evolutionary characteristics of the phosphoketolases from bifidobacteria and Coriobacteriales. PLoS One *12*, e0172176.

Gupta,R.S., Naushad,S., Chokshi,C., Griffiths,E., and Adeolu,M. (2015). A phylogenomic and molecular markers based analysis of the phylum Chlamydiae: proposal to divide the class *Chlamydiia* into two orders, *Chlamydiales* and *Parachlamydiales* ord. nov., and emended description of the class *Chlamydiia*. Antonie Van Leeuwenhoek *108*, 765-781.

Gupta,R.S., Naushad,S., Fabros,R., and Adeolu,M. (2016). A phylogenomic reappraisal of family-level divisions within the class *Halobacteria*: proposal to divide the order *Halobacteriales* into the families *Halobacteriaceae*, *Haloarculaceae* fam. nov., and *Halococcaceae* fam. nov., and the order *Haloferacales* into the families, *Haloferacaceae* and *Halorubraceae* fam nov. Antonie Van Leeuwenhoek *109*, 565-587.

Gupta,R.S. (2014). Identification of conserved indels that are useful for classification and evolutionary studies. Methods in Microbiology *41*, 153-182.

Hahn,D.L., Dodge,R.W., and Golubjatnikov,R. (1991). Association of *Chlamydia pneumoniae* (strain TWAR) infection with wheezing, asthmatic bronchitis, and adult-onset asthma. JAMA *266*, 225-230.

Hanada,K., Iwasaki,M., Ihashi,S., and Ikeda,H. (2000). UvrA and UvrB suppress illegitimate recombination: synergistic action with RecQ helicase. Proc. Natl. Acad. Sci. U. S A *97*, 5989-5994.

Hermjakob,H., Montecchi-Palazzi,L., Lewington,C., Mudali,S., Kerrien,S., Orchard,S., Vingron,M., Roechert,B., Roepstorff,P., Valencia,A., Margalit,H., Armstrong,J., Bairoch,A., Cesareni,G., Sherman,D., and Apweiler,R. (2004). IntAct: an open source molecular interaction database. Nucleic Acids Res. *32*, D452-D455.

Ho,J., Adeolu,M., Khadka,B., and Gupta,R.S. (2016). Identification of distinctive molecular traits that are characteristic of the phylum "Deinococcus-Thermus" and distinguish its main constituent groups. Syst Appl Microbiol *39*, 453-463.

Huang,W.M. (1996). Bacterial diversity based on type II DNA topoisomerase genes. Annu. Rev. Genet. *30*, 79-107.

Hybiske,K. and Stephens,R.S. (2007). Mechanisms of host cell exit by the intracellular bacterium Chlamydia. Proc. Natl. Acad. Sci. U. S A *104*, 11430-11435.

Inaba,K., Murakami,S., Suzuki,M., Nakagawa,A., Yamashita,E., Okada,K., and Ito,K. (2006). Crystal structure of the DsbB-DsbA complex reveals a mechanism of disulfide bond generation. Cell *127*, 789-801.

Jaciuk,M., Nowak,E., Skowronek,K., Tanska,A., and Nowotny,M. (2011). Structure of UvrA nucleotide excision repair protein in complex with modified DNA. Nat. Struct. Mol Biol *18*, 191-197.

Jancura,P., Mavridou,E., Carrillo-de Santa,P.E., and Marchiori,E. (2012). A methodology for detecting the orthology signal in a PPI network at a functional complex level. BMC. Bioinformatics *13 Suppl 10*, S18.

Kalayoglu,M.V., Perkins,B.N., and Byrne,G.I. (2001). *Chlamydia pneumoniae*-infected monocytes exhibit increased adherence to human aortic endothelial cells. Microbes. Infect. *3*, 963-969.

Khadka,B. and Gupta,R.S. (2017). Identification of a conserved 8 aa insert in the PIP5K protein in the *Saccharomycetaceae* family of fungi and the molecular dynamics simulations and structural analysis to investigate its potential functional role. Proteins *85*, 1454-1467.

Khairnar,N.P., Joe,M.H., Misra,H.S., Lim,S.Y., and Kim,D.H. (2013). FrnE, a cadmium-inducible protein in *Deinococcus radiodurans*, is characterized as a disulfide isomerase chaperone in vitro and for its role in oxidative stress tolerance in vivo. J Bacteriol *195*, 2880-2886.

Khan,I., Chen,Y., Dong,T., Hong,X., Takeuchi,R., Mori,H., and Kihara,D. (2014). Genome-scale identification and characterization of moonlighting proteins. Biol Direct *9*, 30.

Kiyosawa,K., Tanaka,M., Matsunaga,T., Nikaido,O., and Yamamoto,K. (2001). Amplified UvrA protein can ameliorate the ultraviolet sensitivity of an *Escherichia coli* recA mutant. Mutat. Res. *487*, 149-156.

Kohlhoff,S.A., Huband,M.D., and Hammerschlag,M.R. (2014). In vitro activity of AZD0914, a novel DNA gyrase inhibitor, against *Chlamydia trachomatis* and *Chlamydia pneumoniae*. Antimicrob. Agents Chemother. *58*, 7595-7596.

Kota,S. and Misra,H.S. (2006). PprA: A protein implicated in radioresistance of *Deinococcus radiodurans* stimulates catalase activity in *Escherichia coli*. Appl. Microbiol. Biotechnol. *72*, 790-796.

Kota,S., Rajpurohit,Y.S., Charaka,V.K., Satoh,K., Narumi,I., and Misra,H.S. (2016). DNA Gyrase of *Deinococcus radiodurans* is characterized as Type II bacterial topoisomerase and its activity is differentially regulated by PprA in vitro 2. Extremophiles. *20*, 195-205.

Krisko,A. and Radman,M. (2010). Protein damage and death by radiation in *Escherichia coli* and *Deinococcus radiodurans*. Proc. Natl. Acad. Sci. U. S A *107*, 14373-14377.

Krisko,A. and Radman,M. (2013). Biology of extreme radiation resistance: the way of *Deinococcus radiodurans*. Cold Spring Harb. Perspect. Biol *5*.

Kubori,T., Matsushima,Y., Nakamura,D., Uralil,J., Lara-Tejero,M., Sukhan,A., Galan,J.E., and Aizawa,S.I. (1998). Supramolecular structure of the Salmonella typhimurium type III protein secretion system. Science *280*, 602-605.

Kuchaiev,O., Stevanovic,A., Hayes,W., and Przulj,N. (2011). GraphCrunch 2: Software tool for network modeling, alignment and clustering. BMC. Bioinformatics *12*, 24.

Kuo,C.C., Gown,A.M., Benditt,E.P., and Grayston,J.T. (1993). Detection of *Chlamydia pneumoniae* in aortic lesions of atherosclerosis by immunocytochemical stain. Arterioscler. Thromb. *13*, 1501-1504.

Kuo,C.C., Grayston,J.T., Campbell,L.A., Goo,Y.A., Wissler,R.W., and Benditt,E.P. (1995). *Chlamydia pneumoniae* (TWAR) in coronary arteries of young adults (15-34 years old). Proc. Natl. Acad. Sci. U. S A *92*, 6911-6914.

Kiyosawa,K., Tanaka,M., Matsunaga,T., Nikaido,O., and Yamamoto,K. (2001). Amplified UvrA protein can ameliorate the ultraviolet sensitivity of an *Escherichia coli* recA mutant
1. Mutat. Res. *487*, 149-156.

Lanz,M.A., Farhat,M., and Klostermeier,D. (2014). The acidic C-terminal tail of the GyrA subunit moderates the DNA supercoiling activity of *Bacillus subtilis* gyrase. J Biol. Chem. *289*, 12275-12285.

Laponogov,I., Veselkov,D.A., Crevel,I.M., Pan,X.S., Fisher,L.M., and Sanderson,M.R. (2013). Structure of an 'open' clamp type II topoisomerase-DNA complex provides a mechanism for DNA capture and transport. Nucleic Acids Res. *41*, 9911-9923.

Laraia,L., McKenzie,G., Spring,D.R., Venkitaraman,A.R., and Huggins,D.J. (2015). Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions. Chem. Biol. *22*, 689-703.

Lerat,E., Daubin,V., Ochman,H., and Moran,N.A. (2005). Evolutionary origins of genomic repertoires in bacteria. PLoS Biol. *3*, e130.

Lorenzini,E., Singer,A., Singh,B., Lam,R., Skarina,T., Chirgadze,N.Y., Savchenko,A., and Gupta,R.S. (2010). Structure and protein-protein interaction studies on Chlamydia trachomatis protein CT670 (YscO Homolog). J Bacteriol *192*, 2746-2756.

Lin,C.G., Kovalsky,O., and Grossman,L. (1997). DNA damage-dependent recruitment of nucleotide excision repair and transcription proteins to *Escherichia coli* inner membranes. Nucleic Acids Res. *25*, 3151-3158.

Lovell,S.C., Davis,I.W., Arendall,W.B., III, de Bakker,P.I., Word,J.M., Prisant,M.G., Richardson,J.S., and Richardson,D.C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins *50*, 437-450.

Makarova,K.S., Aravind,L., Wolf,Y.I., Tatusov,R.L., Minton,K.W., Koonin,E.V., and Daly,M.J. (2001a). Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. Microbiol Mol Biol Rev. *65*, 44-79.

Makarova,K.S., Omelchenko,M.V., Gaidamakova,E.K., Matrosova,V.Y., Vasilenko,A., Zhai,M., Lapidus,A., Copeland,A., Kim,E., Land,M., Mavrommatis,K., Pitluck,S., Richardson,P.M., Detter,C., Brettin,T., Saunders,E., Lai,B., Ravel,B., Kemner,K.M., Wolf,Y.I., Sorokin,A., Gerasimova,A.V., Gelfand,M.S., Fredrickson,J.K., Koonin,E.V., and Daly,M.J. (2007). *Deinococcus geothermalis*: the pool of extreme radiation resistance genes shrinks . PLoS One *2*, e955.

Marino,S.M. (2014). Protein flexibility and cysteine reactivity: influence of mobility on the H-bond network and effects on pKa prediction. Protein J *33*, 323-336.

Marshall,N.C. and Finlay,B.B. (2014). Targeting the type III secretion system to treat bacterial infections. Expert. Opin. Ther. Targets. *18*, 137-152.

Matsumoto,A. (1973). Fine structures of cell envelopes of *Chlamydia* organisms as revealed by freeze-etching and negative staining techniques. J Bacteriol *116*, 1355-1363.

Mattimore,V. and Battista,J.R. (1996). Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. J Bacteriol *178*, 633-637.

Markillie, L. M., V. S., Hradecky, P. & Wong, K. K. (1999) Targeted Mutagenesis by Duplication Insertion in the Radioresistant Bacterium *Deinococcus radiodurans:* Radiation Sensitivities of Catalase (katA) and Superoxide Dismutase (sodA) MutantsJ. Bacteriol. 181, 666–669.

Mao,F., Dam,P., Chou,J., Olman,V., and Xu,Y. (2009). DOOR: a database for prokaryotic operons. Nucleic Acids Res. *37*, D459-D463.

Mao,X., Ma,Q., Zhou,C., Chen,X., Zhang,H., Yang,J., Mao,F., Lai,W., and Xu,Y. (2014). DOOR 2.0: presenting operons and their functions through dynamic and integrated views
1. Nucleic Acids Res. *42*, D654-D659.

Masure,H.R., Pearce,B.J., Shio,H., and Spellerberg,B. (1998). Membrane targeting of RecA during genetic transformation. Mol. Microbiol. *27*, 845-852.

Maxwell,A. and Lawson,D.M. (2003). The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. Curr. Top. Med. Chem. *3*, 283-303.

Misra,H.S., Khairnar,N.P., Kota,S., Shrivastava,S., Joshi,V.P., and Apte,S.K. (2006). An exonuclease I-sensitive DNA repair pathway in *Deinococcus radiodurans*: a major determinant of radiation resistance. Mol Microbiol *59*, 1308-1316.

Meijer, A. and Ossewaarde, J.M. (2002) Description of wider diversity within the order *Chlamydiales* than currently classified. In: Chlamydial Infections – Proceedings of the 10th International Symposium on Human Chlamydial Infections. (Schachter, J et al., Eds.), pp. 13–16. International Chlamydial Symposium, San Francisco. (PAGE 7)

Moreno-Hagelsieb,G. (2015). The power of operon rearrangements for predicting functional associations. Comput. Struct. Biotechnol J *13*, 402-406.

Moulder,J.W. (1991). Interaction of Chlamydiae and host cells in vitro. Microbiol. Rev. *55*, 143-190.

Mun,C., Del Rowe,J., Sandigursky,M., Minton,K.W., and Franklin,W.A. (1994). DNA deoxyribophosphodiesterase and an activity that cleaves DNA containing thymine glycol adducts in Deinococcus radiodurans. Radiat. Res. *138*, 282-285.

Narumi,I., Satoh,K., Cui,S., Funayama,T., Kitayama,S., and Watanabe,H. (2004). PprA: a novel protein from *Deinococcus radiodurans* that stimulates DNA ligation. Mol. Microbiol. *54*, 278-285.

Naushad,H.S., Lee,B., and Gupta,R.S. (2014). Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya*, *Pectobacterium* and *Brenneria*. Int J Syst Evol. Microbiol. *64*, 366-383.

Navaratnam,S., Myles,G.M., Strange,R.W., and Sancar,A. (1989). Evidence from extended X-ray absorption fine structure and site-specific mutagenesis for zinc fingers in UvrA protein of *Escherichia coli*. J Biol Chem. *264*, 16067-16071.

Nicholson,T.L., Olinger,L., Chong,K., Schoolnik,G., and Stephens,R.S. (2003). Global stage-specific gene regulation during the developmental cycle of *Chlamydia trachomatis*. J Bacteriol *185*, 3179-3189.

Niehus,E., Cheng,E., and Tan,M. (2008). DNA supercoiling-dependent gene regulation in *Chlamydia*. J Bacteriol *190*, 6419-6427.

Nugent,T. and Jones,D.T. (2009). Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics *10*, 159.

Omelchenko,M.V., Wolf,Y.I., Gaidamakova,E.K., Matrosova,V.Y., Vasilenko,A., Zhai,M., Daly,M.J., Koonin,E.V., and Makarova,K.S. (2005). Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance. BMC. Evol. Biol. *5*, 57.

Orillard,E. and Tan,M. (2016). Functional analysis of three topoisomerases that regulate DNA supercoiling levels in Chlamydia. Mol. Microbiol. *99*, 484-496.

Pakotiprapha,D., Samuels,M., Shen,K., Hu,J.H., and Jeruzalmi,D. (2012). Structure and mechanism of the UvrA-UvrB DNA damage sensor. Nat. Struct. Mol Biol *19*, 291-298.

Pereira-Leal,J.B., Enright,A.J., and Ouzounis,C.A. (2004). Detection of functional modules from protein interaction networks. Proteins *54*, 49-57.

Plano,G.V., Day,J.B., and Ferracci,F. (2001). Type III export: new uses for an old pathway. Mol. Microbiol. *40*, 284-293.

Podobnik,M., Weitze,T.F., O'Donnell,M., and Kuriyan,J. (2003). Nucleotide-induced conformational changes in an isolated *Escherichia coli* DNA polymerase III clamp loader subunit. Structure *11*, 253-263.

Rao,V.S., Srinivas,K., Sujini,G.N., and Kumar,G.N. (2014). Protein-protein interaction detection: methods and analysis. Int J Proteomics *2014*, 147648.

Rayssiguier,C., Thaler,D.S., and Radman,M. (1989). The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. Nature *342*, 396-401.

Reece,R.J. and Maxwell,A. (1991). Probing the limits of the DNA breakage-reunion domain of the *Escherichia coli* DNA gyrase A protein. J Biol. Chem. *266*, 3540-3546.

Rokas,A. and Holland,P.W. (2000). Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. *15*, 454-459.

Rosario,C.J., Hanson,B.R., and Tan,M. (2014). The transcriptional repressor EUO regulates both subsets of *Chlamydia* late genes. Mol. Microbiol. *94*, 888-897.

Rosario,C.J. and Tan,M. (2012). The early gene product EUO is a transcriptional repressor that selectively regulates promoters of *Chlamydia* late genes. Mol. Microbiol. *84*, 1097-1107.

Rosario,C.J. and Tan,M. (2015). Regulation of *Chlamydia* Gene Expression by Tandem Promoters with Different Temporal Patterns. J Bacteriol *198*, 363-369.

Rose,P.W., Prlic,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J., Young,J., Zardecki,C., Berman,H.M., Bourne,P.E., and Burley,S.K. (2015). The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res. *43*, D345-D356.

Rurangirwa,F.R., Dilbeck,P.M., Crawford,T.B., McGuire,T.C., and McElwain,T.F. (1999). Analysis of the 16S rRNA gene of micro-organism WSU 86-1044 from an aborted bovine foetus reveals that it is a member of the order *Chlamydiales*: proposal of *Waddliaceae* fam. nov., *Waddlia chondrophila* gen. nov., sp. nov. Int J Syst Bacteriol *49 Pt 2*, 577-581.

Saikku,P., Leinonen,M., Mattila,K., Ekman,M.R., Nieminen,M.S., Makela,P.H., Huttunen,J.K., and Valtonen,V. (1988). Serological evidence of an association of a novel *Chlamydia*, TWAR, with chronic coronary heart disease and acute myocardial infarction. Lancet *2*, 983-986.

Sali,A. and Blundell,T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol *234*, 779-815.

Sancar,A., Franklin,K.A., and Sancar,G.B. (1984). *Escherichia coli* DNA photolyase stimulates uvrABC excision nuclease in vitro. Proc. Natl. Acad. Sci. U. S A *81*, 7397-7401.

Sandigursky,M. and Franklin,W.A. (1999). Thermostable uracil-DNA glycosylase from Thermotoga maritima a member of a novel class of DNA repair enzymes. Curr. Biol. *9*, 531-534.

Schachter,J. (1978). Chlamydial infections (second of three parts). N. Engl. J Med. *298*, 490-495.

Schlesinger,D.J. (2007). Role of RecA in DNA damage repair in Deinococcus radiodurans. FEMS Microbiol. Lett. *274*, 342-347.

Schmier,B.J., Chen,X., Wolin,S., and Shuman,S. (2017). Deletion of the rnl gene encoding a nick-sealing RNA ligase sensitizes *Deinococcus radiodurans* to ionizing radiation. Nucleic Acids Res. *45*, 3812-3821.

Selvam, K. et al (2013) DdrA, DdrD, and PprA: Components of UV and Mitomycin C Resistance in *Deinococcus radiodurans* R1.Plos one, Vol.8:e69007.

Shen,M.Y. and Sali,A. (2006). Statistical potential for assessment and prediction of protein structures. Protein Sci. *15*, 2507-2524.

Shen,C.H., Chiang,Y.C., Hsu,C.H., and Yang,M.K. (2007). Identification and characterization of two uvrA genes of *Xanthomonas axonopodis pathovar citri*. Mol Genet. Genomics *277*, 149-160.

Shouldice,S.R., Heras,B., Walden,P.M., Totsika,M., Schembri,M.A., and Martin,J.L. (2011). Structure and function of DsbA, a key bacterial oxidative folding catalyst. Antioxid. Redox. Signal. *14*, 1729-1760.

Singh,B. and Gupta,R.S. (2009). Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. Mol Genet. Genomics *281*, 361-373.

Sinha,R.P. and Hader,D.P. (2002). UV-induced DNA damage and repair: a review. Photochem. Photobiol. Sci. *1*, 225-236.
Sippl,M.J. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins *17*, 355-362.

Sisko,J.L., Spaeth,K., Kumar,Y., and Valdivia,R.H. (2006). Multifunctional analysis of *Chlamydia*-specific genes in a yeast expression system. Mol. Microbiol. *60*, 51-66

Sissi,C. and Palumbo,M. (2010). In front of and behind the replication fork: bacterial type IIA topoisomerases. Cell Mol. Life Sci. *67*, 2001-2024.

Slade,D., Lindner,A.B., Paul,G., and Radman,M. (2009). Recombination and replication in DNA repair of heavily irradiated Deinococcus radiodurans 2. Cell *136*, 1044-1055.

Slade,D. and Radman,M. (2011). Oxidative stress resistance in *Deinococcus radiodurans* . Microbiol Mol Biol Rev. *75*, 133-191.

Smith,B.T., Grossman,A.D., and Walker,G.C. (2002). Localization of UvrA and effect of DNA damage on the chromosome of *Bacillus subtilis*. J Bacteriol *184*, 488-493.

Szklarczyk,D., Franceschini,A., Wyder,S., Forslund,K., Heller,D., Huerta-Cepas,J., Simonovic,M., Roth,A., Santos,A., Tsafou,K.P., Kuhn,M., Bork,P., Jensen,L.J., and von

Mering,C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447-D452.

Szklarczyk,D., Morris,J.H., Cook,H., Kuhn,M., Wyder,S., Simonovic,M., Santos,A., Doncheva,N.T., Roth,A., Bork,P., Jensen,L.J., and von Mering,C. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45*, D362-D368.

Stackebrandt, E., & Schumann, P. (2006). Introduction to the taxonomy of actinobacteria The Prokaryotes (3 ed., pp. 297-321).

Sugiman-Marangos,S.N., Weiss,Y.M., and Junop,M.S. (2016). Mechanism for accurate, protein-assisted DNA annealing by *Deinococcus radiodurans* DdrB. Proc. Natl. Acad. Sci. U. S A *113*, 4308-4313.

Szklarczyk,D., Franceschini,A., Wyder,S., Forslund,K., Heller,D., Huerta-Cepas,J., Simonovic,M., Roth,A., Santos,A., Tsafou,K.P., Kuhn,M., Bork,P., Jensen,L.J., and von Mering,C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447-D452.

Tanaka,M., Earl,A.M., Howell,H.A., Park,M.J., Eisen,J.A., Peterson,S.N., and Battista,J.R. (2004). Analysis of *Deinococcus radiodurans*'s transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance 1. Genetics *168*, 21-33.

Tanaka,M., Narumi,I., Funayama,T., Kikuchi,M., Watanabe,H., Matsunaga,T., Nikaido,O., and Yamamoto,K. (2005). Characterization of pathways dependent on the uvsE, uvrA1, or uvrA2 gene product for UV resistance in *Deinococcus radiodurans*. J Bacteriol *187*, 3693-3697.

Timmins,J., Gordon,E., Caria,S., Leonard,G., Acajjaoui,S., Kuo,M.S., Monchois,V., and McSweeney,S. (2009). Structural and mutational analyses of *Deinococcus radiodurans* UvrA2 provide insight into DNA binding and damage recognition by UvrAs. Structure. *17*, 547-558.

Timmins,J. and Moe,E. (2016). A Decade of Biochemical and Structural Studies of the DNA Repair Machinery of *Deinococcus radiodurans*: Major Findings, Functional and Mechanistic Insight and Challenges 1. Comput. Struct. Biotechnol J *14*, 168-176.

Tornow,S. and Mewes,H.W. (2003). Functional modules by relating protein interaction networks and gene expression. Nucleic Acids Res. *31*, 6283-6289.

Valdivia,R.H. (2008). *Chlamydia* effector proteins and new insights into chlamydial cellular microbiology. Curr. Opin. Microbiol. *11*, 53-59.

von Mering,C., Jensen,L.J., Snel,B., Hooper,S.D., Krupp,M., Foglierini,M., Jouffre,N., Huynen,M.A., and Bork,P. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. *33*, D433-D437.

Vujicic-Zagar,A., Dulermo,R., Le Gorrec,M., Vannier,F., Servant,P., Sommer,S., de Groot,A., and Serre,L. (2009). Crystal structure of the IrrE protein, a central regulator of DNA damage repair in *Deinococcaceae*. J Mol Biol *386*, 704-716.

Wagner,K., Moolenaar,G.F., and Goosen,N. (2011). Role of the insertion domain and the zinc-finger motif of *Escherichia coli* UvrA in damage recognition and ATP hydrolysis
1. DNA Repair (Amst) *10*, 483-496.

Wang,J., Mueller,K.L., and Grossman,L. (1994). A mutational study of the C-terminal zinc-finger motif of the *Escherichia coli* UvrA protein
. J Biol Chem. *269*, 10771-10775.

Ward,M.E. (1983). Chlamydial classification, development and structure. Br. Med. Bull. *39*, 109-115.

White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D., Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L., Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M., Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C., Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W., Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S., Smith,H.O., Venter,J.C., and Fraser,C.M. (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science *286*, 1571-1577.

Weisburg,W.G., Giovannoni,S.J., and Woese,C.R. (1989). The Deinococcus-Thermus phylum and the effect of rRNA composition on phylogenetic tree construction. Syst Appl. Microbiol. *11*, 128-134.

Wiederstein,M. and Sippl,M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. *35*, W407-W410.

Wong,K.M., Qiu,X., Cheng,D., Azad,A.K., Habbous,S., Palepu,P., Mirshams,M., Patel,D., Chen,Z., Roberts,H., Knox,J., Marquez,S., Wong,R., Darling,G., Waldron,J., Goldstein,D., Leighl,N., Shepherd,F.A., Tsao,M., Der,S., Reisman,D., and Liu,G. (2014). Two BRM promoter insertion polymorphisms increase the risk of early-stage upper aerodigestive tract cancers. Cancer Med. *3*, 426-433.

Wyrick,P.B. (2010). *Chlamydia trachomatis* persistence in vitro: an overview. J Infect. Dis. *201 Suppl 2*, S88-S95.

Xu,G., Lu,H., Wang,L., Chen,H., Xu,Z., Hu,Y., Tian,B., and Hua,Y. (2010). DdrB stimulates single-stranded DNA annealing and facilitates RecA-independent DNA repair in *Deinococcus radiodurans*. DNA Repair (Amst) *9*, 805-812.

Xu,D. and Zhang,Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J *101*, 2525-2534.

Yu,H.H., Kibler,D., and Tan,M. (2006). In silico prediction and functional validation of sigma28-regulated genes in *Chlamydia* and *Escherichia coli*. J Bacteriol *188*, 8206-8212.

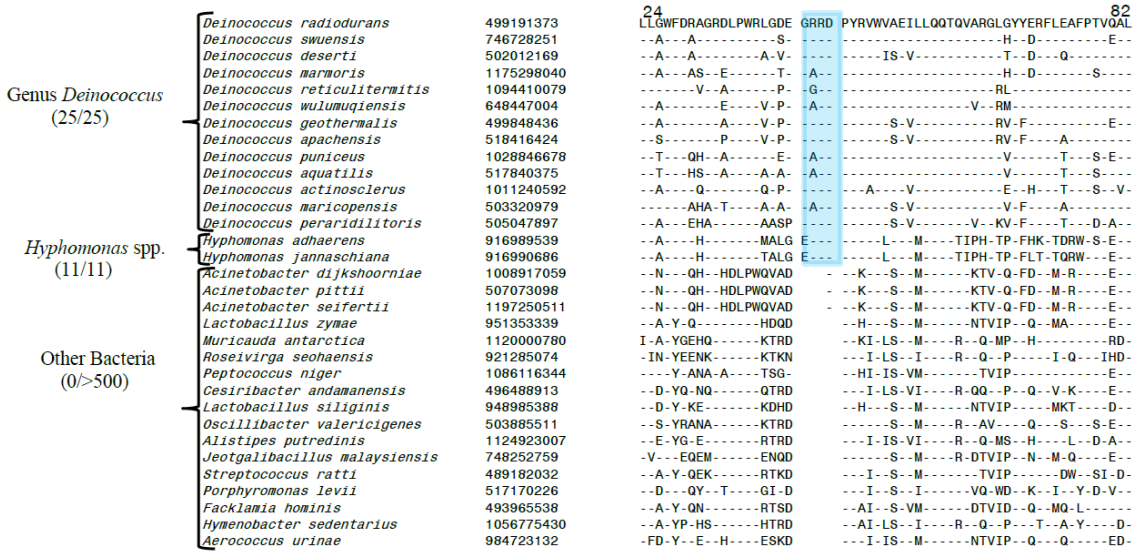
Yu,H.H. and Tan,M. (2003). Sigma28 RNA polymerase regulates hctB, a late developmental gene in *Chlamydia*. Mol. Microbiol. *50*, 577-584.

Zahradka,K., Slade,D., Bailone,A., Sommer,S., Averbeck,D., Petranovic,M., Lindner,A.B., and Radman,M. (2006). Reassembly of shattered chromosomes in *Deinococcus radiodurans*. Nature *443*, 569-573.

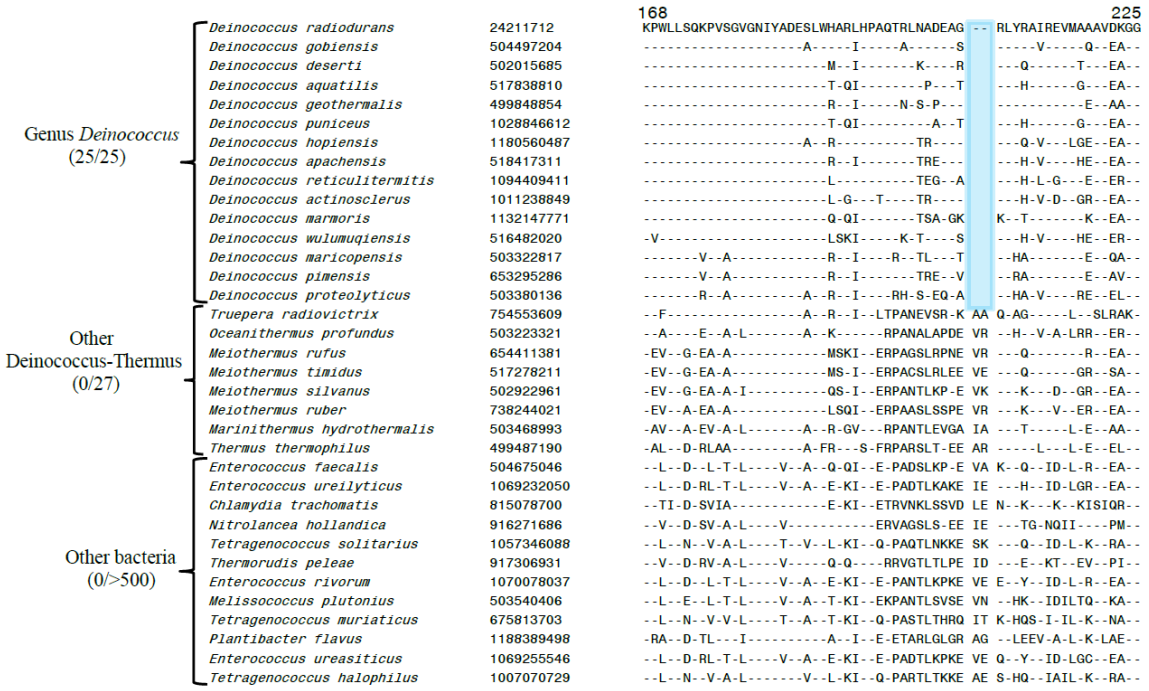
Zhi,X.Y., Zhao,W., Li,W.J., and Zhao,G.P. (2012). Prokaryotic systematics in the genomics era. Antonie Van Leeuwenhoek *101*, 21-34.

Zhong,G., Fan,P., Ji,H., Dong,F., and Huang,Y. (2001). Identification of a chlamydial protease-like activity factor responsible for the degradation of host transcription factors. J Exp. Med. *193*, 935-942.

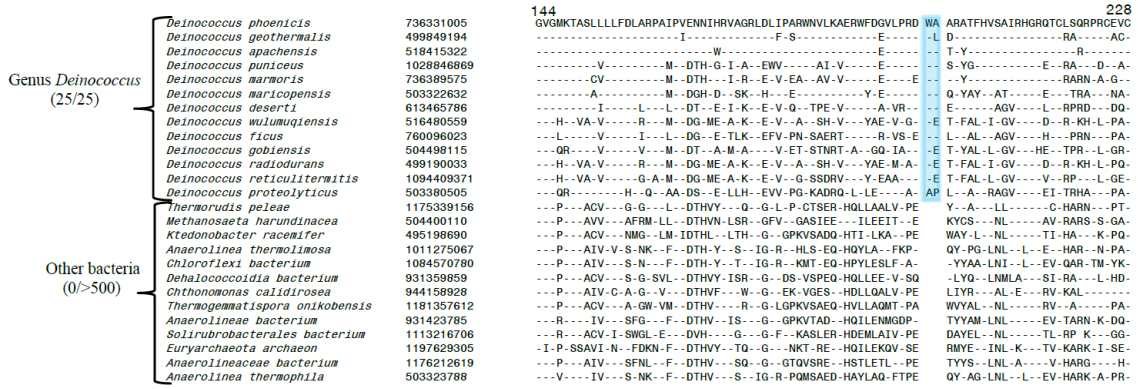
Appendix:



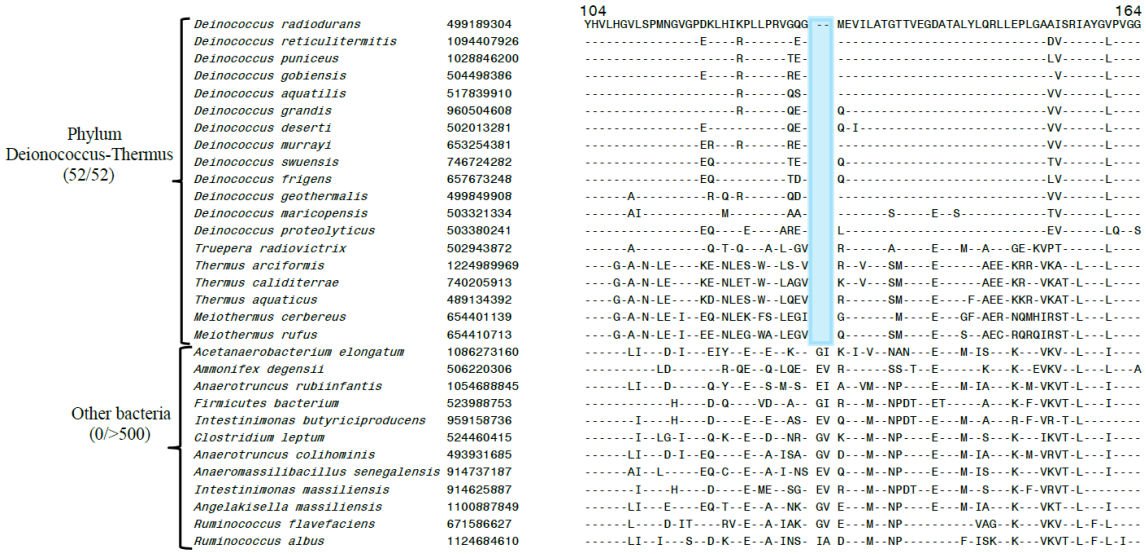
Supplementary Figure 1: Partial sequence alignment of conserved region of 8-oxoguanine DNA glycosylase (MutY) protein showing a 4 amino acid insertion that is a distinctive characteristic from homologs of *Deinococcus* group of bacteria. *Hyphomonas* spp. also contain this insertion in this position.



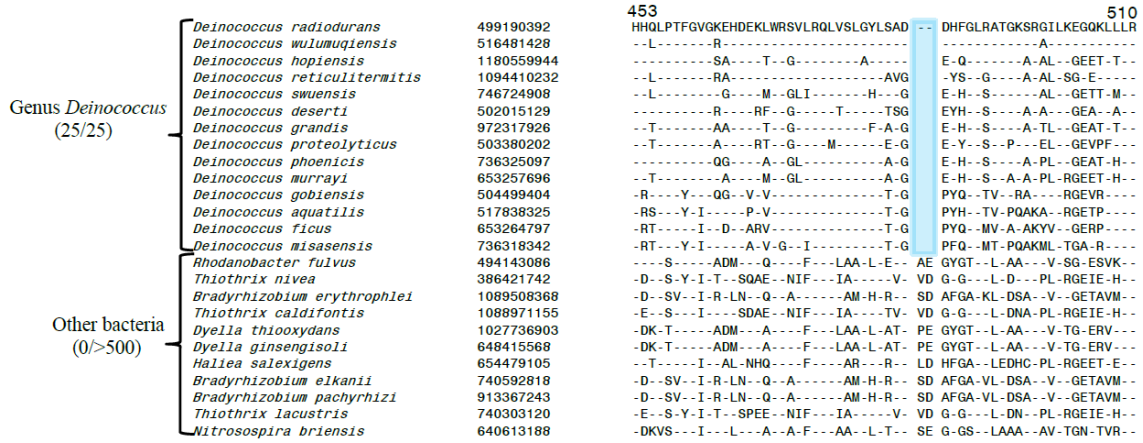
Supplementary Figure 2: Partial sequence alignment of conserved region of formamidopyrimidine and 8-oxoguanine DNA glycosylase (MutM) protein showing a 2 amino acid deletion that is the distinctive molecular characteristics of *Deinococcus* spp.



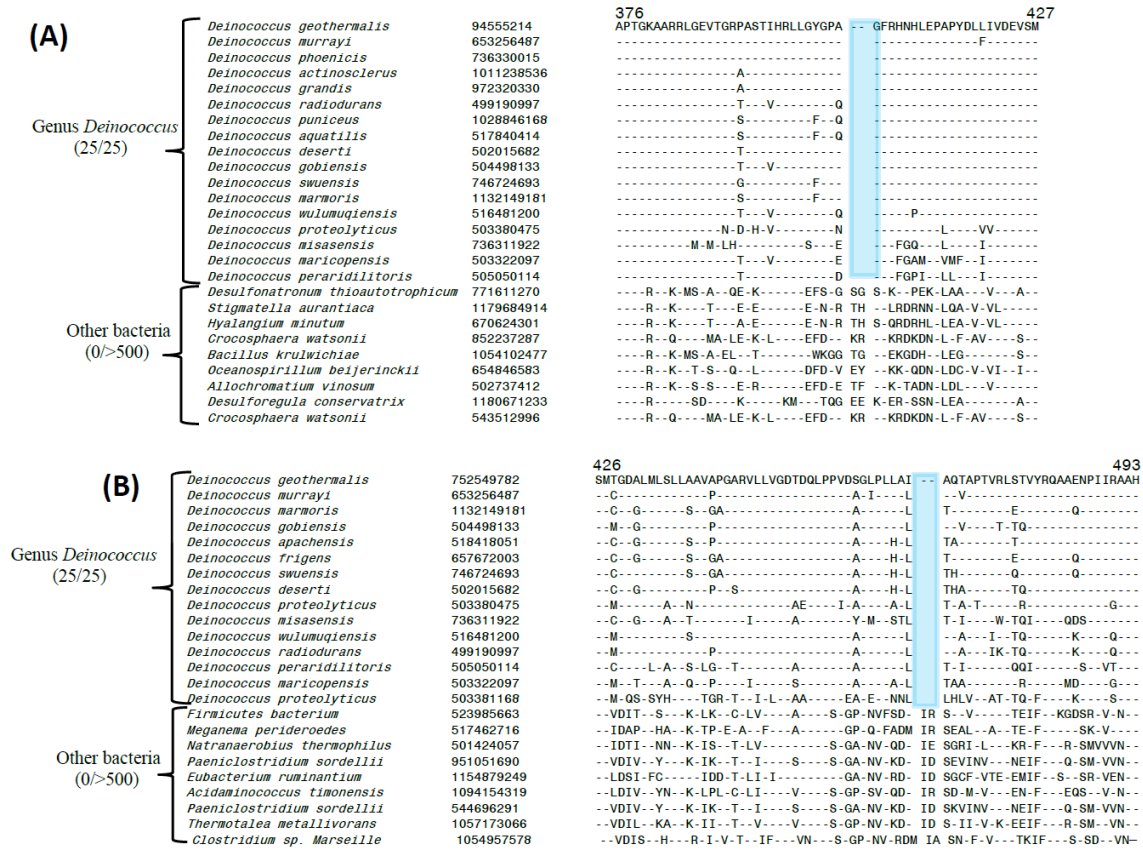
Supplementary Figure 3: Partial sequence alignment of conserved region of Endonuclease III (Nth) protein showing a 2 amino acid insertion that is specific for all *Deinococcus* group of bacteria.



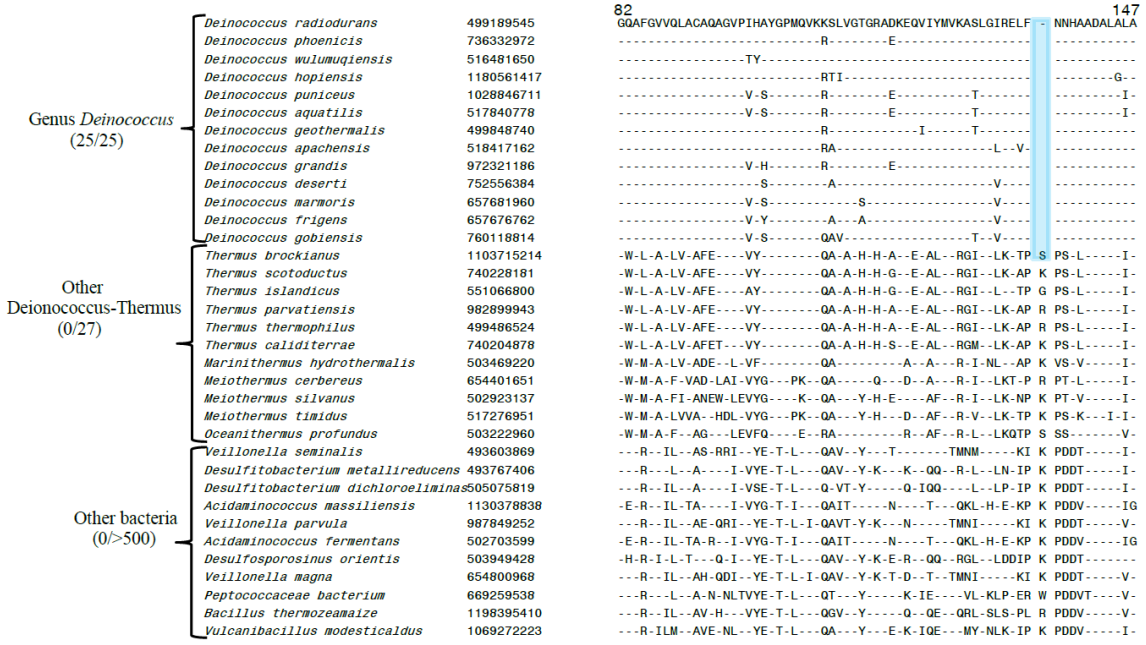
Supplementary Figure 4: Partial sequence alignment of conserved region of DNA helicase recombination protein RecR showing a 2 amino acid deletion that is the distinctive molecular characteristics of *Deinococcales* spp..



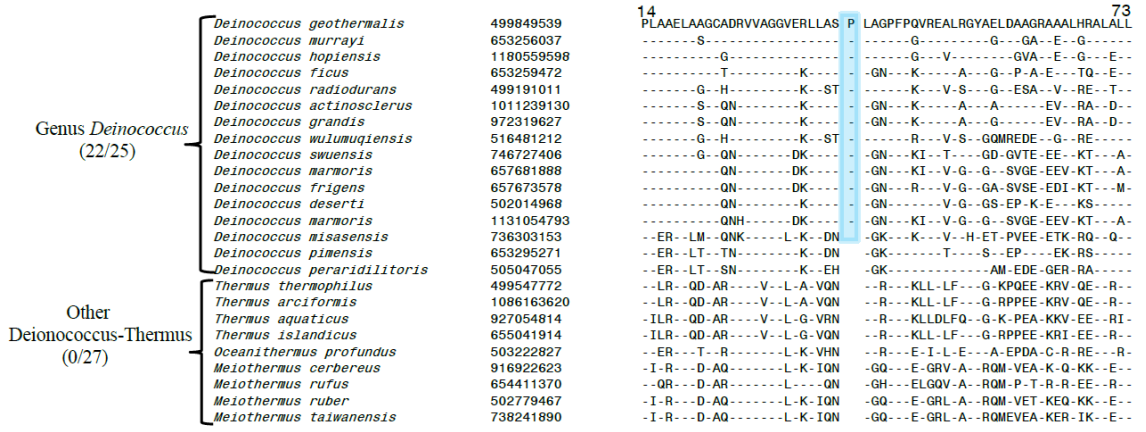
Supplementary Figure 5: Partial sequence alignment of conserved region of DNA helicase RecQ protein showing a 2 amino acid deletion that is uniquely shared by all *Deinococcus* group of bacteria.



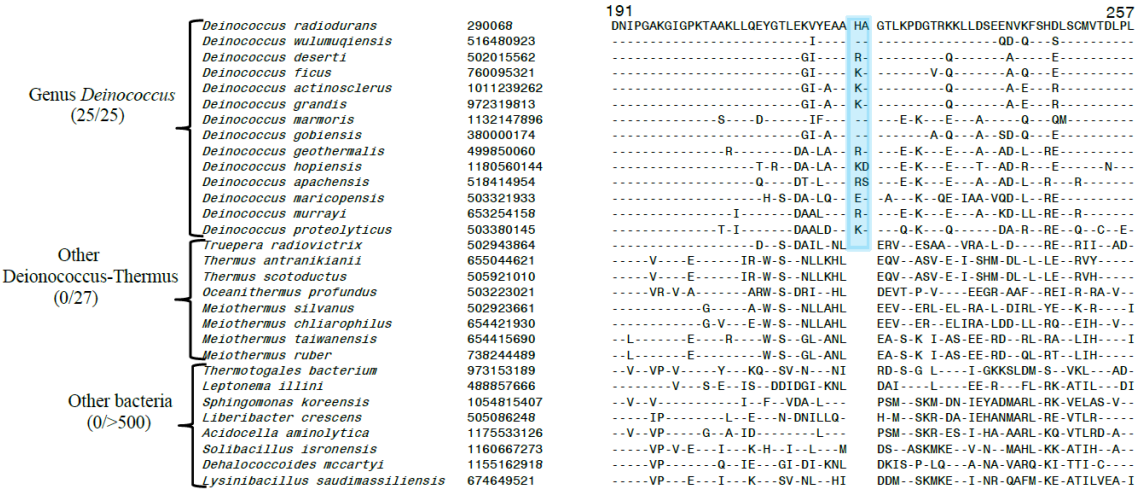
Supplementary Figure 6: Partial sequence alignment of conserved region of helicase RecD protein showing two 2 amino acid deletions that are specific for all *Deinococcus* group of bacteria



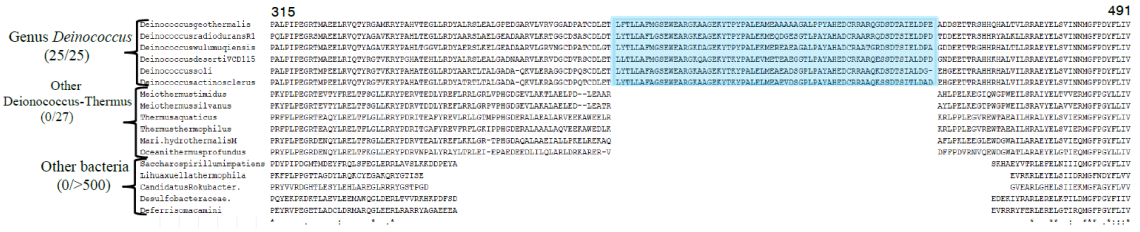
Supplementary Figure 7: Partial sequence alignment of conserved region of crossover junction endodeoxyribonuclease RuvC protein showing a 2 amino acid deletion that is the distinctive molecular characteristics of *Deinococcus* spp.



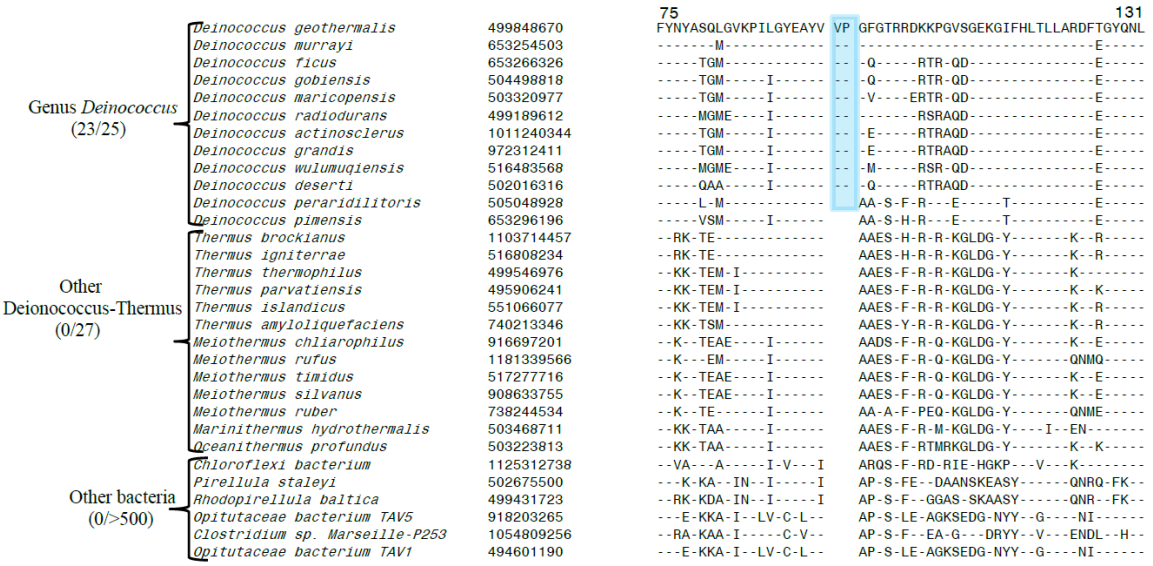
Supplementary Figure 8: Partial sequence alignment of conserved region of DNA helicase RecG protein showing a 1 amino acid insertion that is specific for all *Deinococcus* group of bacteria. Three *Deinococcus spp.* do not contain this insertion.



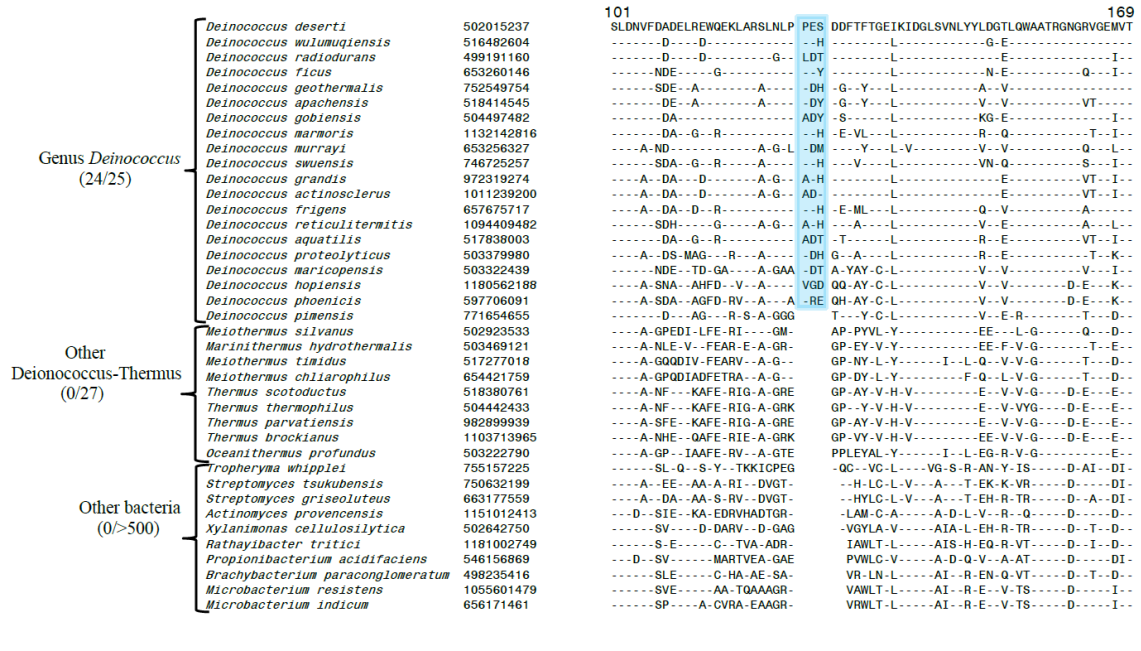
Supplementary Figure 9: Partial sequence alignment of conserved region of DNA polymerase I (PolA) protein showing a 2 amino acid insertion that is specific for all *Deinococcus* group of bacteria.



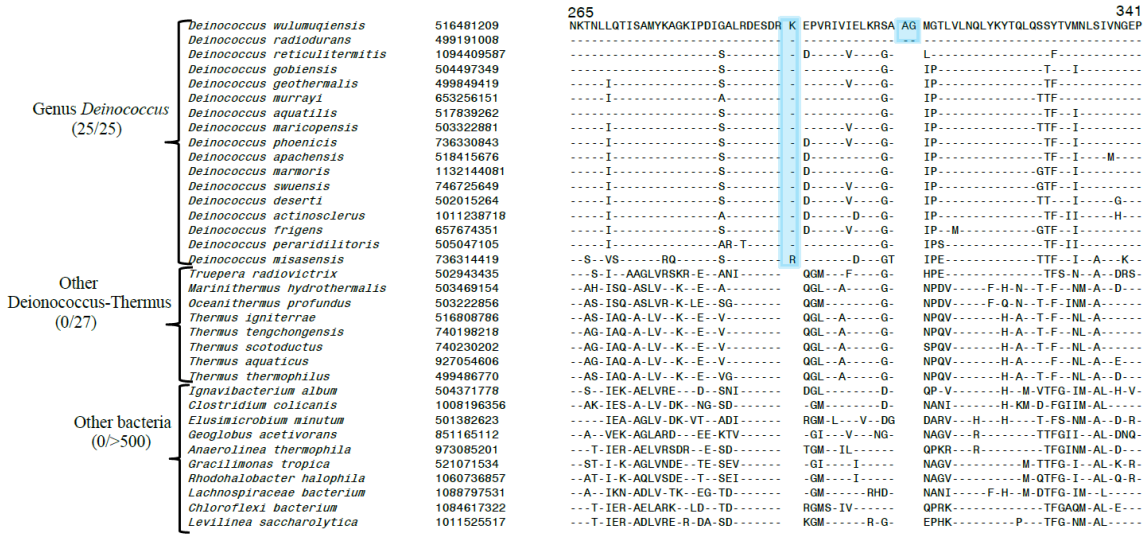
Supplementary Figure 10: Partial sequence alignment of conserved region of DNA polymerase III alpha (DnaE) protein showing a 65 amino acid insertion that is uniquely shared by all *Deinococcus* group of bacteria.



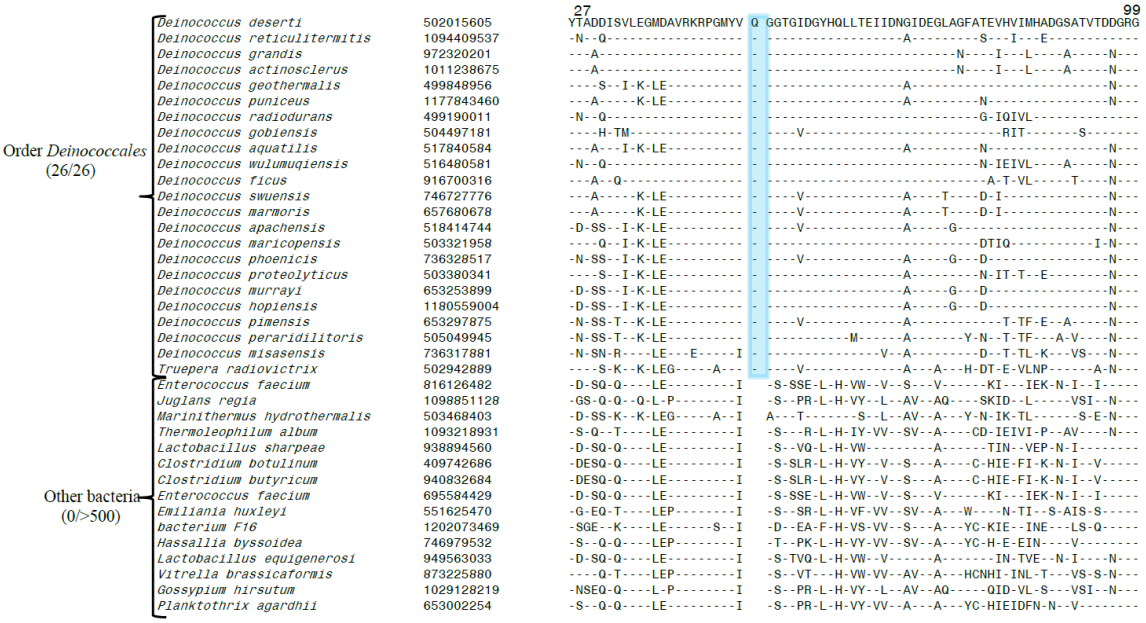
Supplementary Figure 11: Partial sequence alignment of conserved region of DNA polymerase III alpha (DnaE) showing a 2 amino acid insertion that is a distinctive characteristics from homologs of *Deinococcus* group of bacteria. Two *Deinococcus* spp. do not contain this CSI.



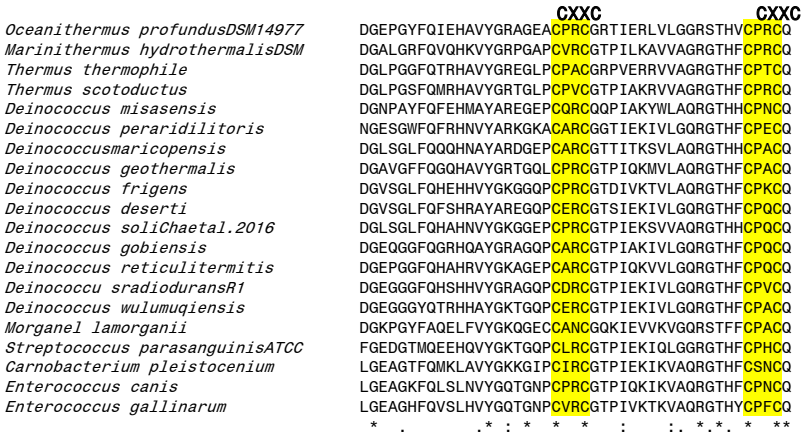
Supplementary Figure 12: Partial sequence alignment of conserved region of NAD (+)-dependent DNA ligase (LigA) protein showing a 3 amino acid insertion that is uniquely shared by all *Deinococcus* group of bacteria. *Deinococcus pimensis* does not contain this insertion.



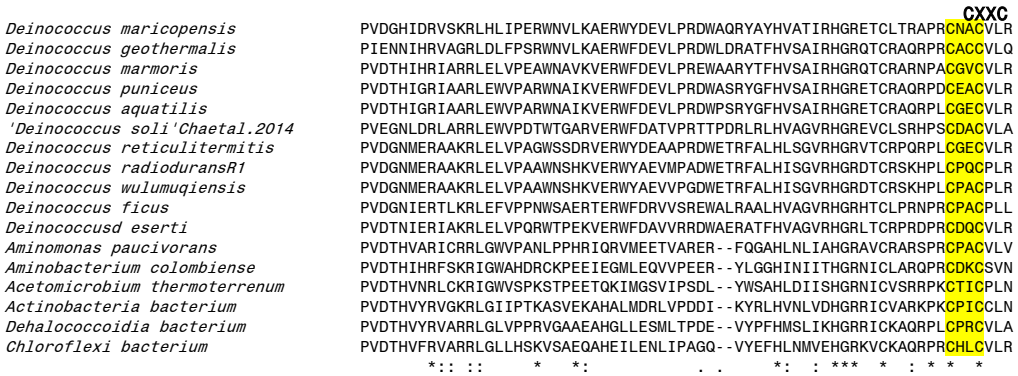
Supplementary Figure 13: Partial sequence alignment of conserved region of DNA gyrase A (GyrA) protein showing a 1 amino acid insertion that is uniquely shared by all *Deinococcus* group of bacteria. This sequence alignment also contains a 2 amino acid size insertion which is only specific for *Deinococcus radiodurans* and *Deinococcus wulumuquiensis*.



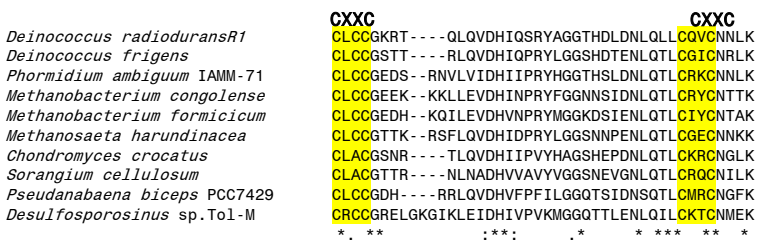
Supplementary Figure 14: Partial sequence alignment of conserved region of DNA gyrase subunit B (GyrB) protein showing a 1 amino acid insertion that is specific for all *Deinococcus* group of bacteria.



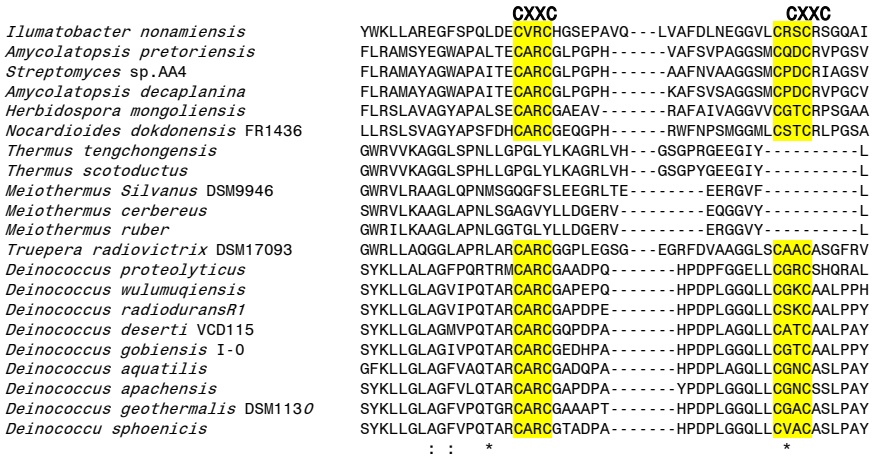
Supplementary Figure 15A: Comparative genomics analyses of CXXC motif in the formamidopyrimidine and 8-oxoguanine DNA glycosylase (MutM) protein. This protein contains two CXXC motifs and these motifs are conserved through the different bacterial species. These CXXC motifs are present in zinc finger



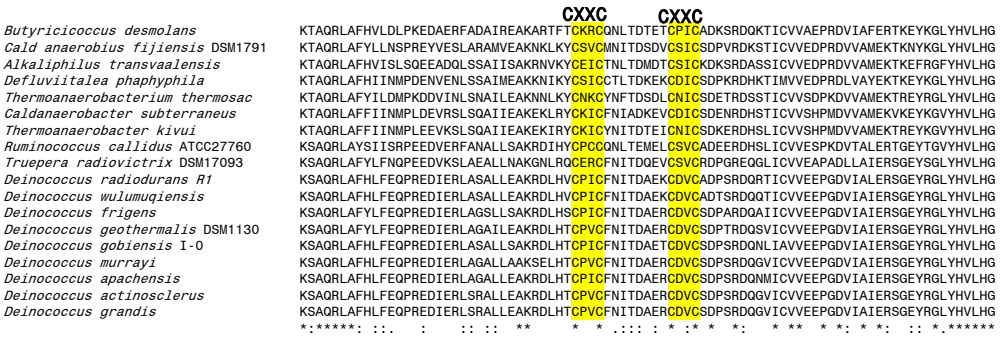
Supplementary Figure 15B: Comparative genomics analyses of CXXC motif in the Endonuclease III (Nth) protein. This protein contains a CXXC motif and this motif is conserved through the bacterial species.



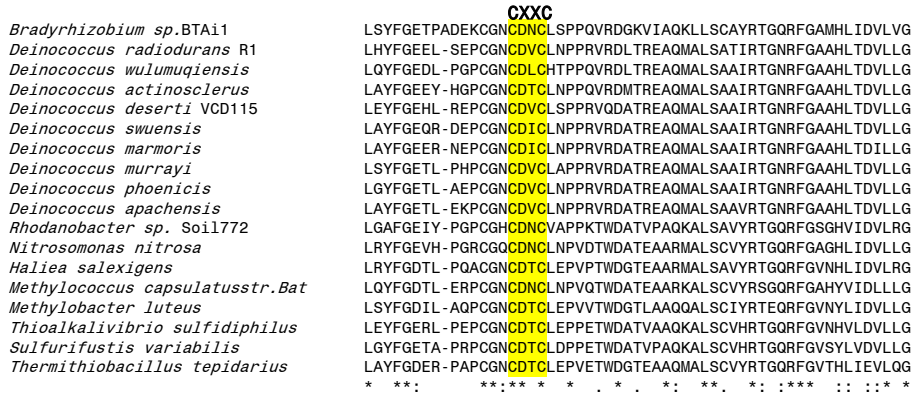
Supplementary Figure 15C: Comparative genomics analyses of CXXC motif in the DNA or RNA helicase of superfamily II (Rad25) protein. This protein contains two CXXC motifs and these motifs are conserved through the bacterial species.



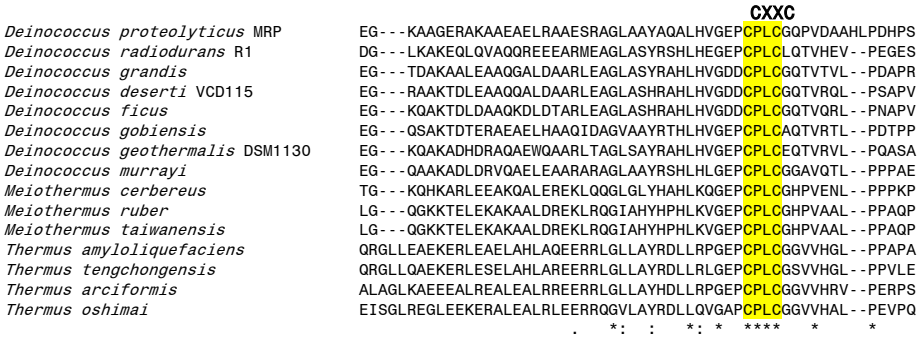
Supplementary Figure 15D: Comparative genomics analyses of CXXC motif in the DNA repair protein RecO. This protein contains two CXXC motifs and these motifs are conserved through most of the bacterial species. But this is not present in *thermus* spp.. These CXXC motifs are present in zinc finger.



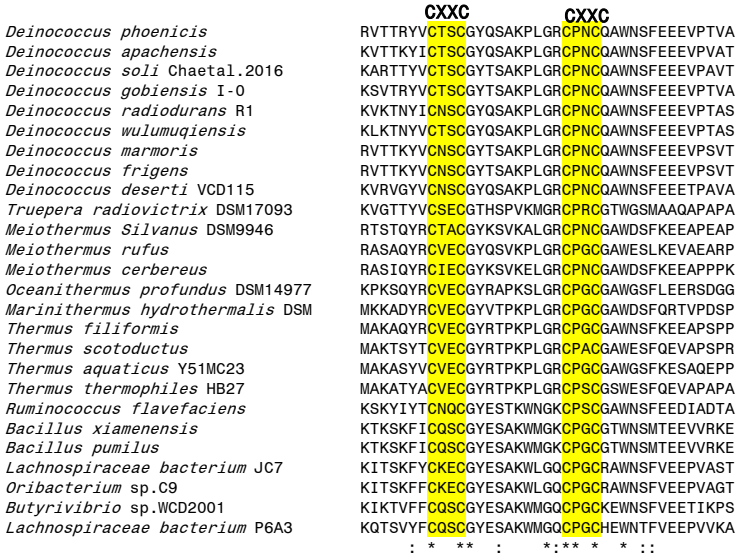
Supplementary Figure 15E: Comparative genomics and structural analyses of CXXC motif in the recombination protein RecR. This protein contains two CXXC motifs and these motifs are conserved through the different bacterial species. These CXXC motifs are present in zinc finger



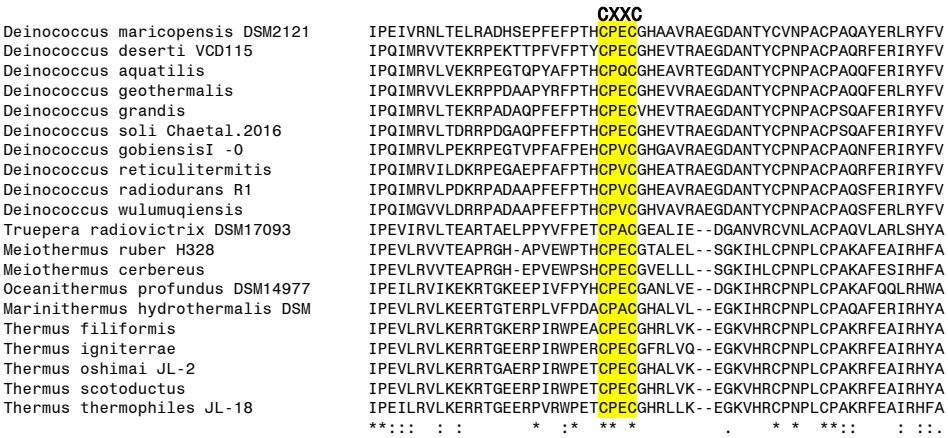
Supplementary Figure 15F: Comparative genomics and structural analyses of CXXC motif in the DNA helicase RecQ protein. This protein contains a CXXC motif and this motif is conserved through the different bacterial species. This CXXC motif is present in zinc finger domain.



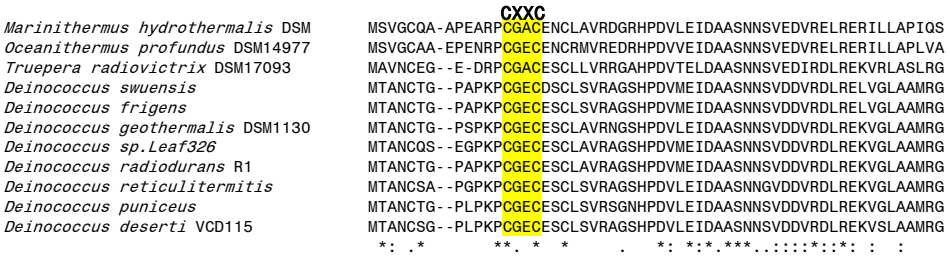
Supplementary Figure 15G: Comparative genomics and structural analyses of CXXC motif in the exonuclease SbcC protein. This protein contains a CXXC motif and this motif is conserved through the bacterial species.



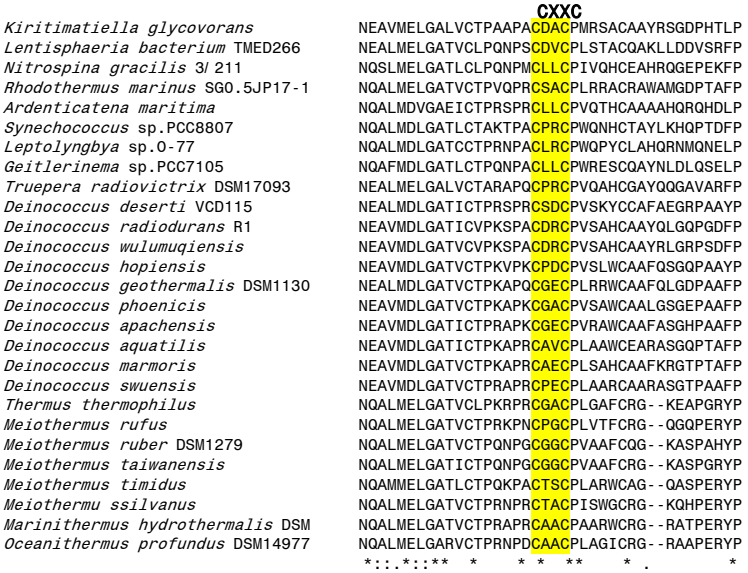
Supplementary Figure 15H: Comparative genomics and structural analyses of CXXC motif in the DNA repair protein RadA. This protein contains two CXXC motifs in zinc finger motif and these motifs are conserved through the bacterial species.



Supplementary Figure 15I: Comparative genomics and structural analyses of CXXC motif in the NAD- dependent DNA ligase (LigA) protein. This protein contains two CXXC motifs in zinc finger motif and these motifs are conserved through the bacterial species.



Supplementary Figure 15J: Comparative genomics and structural analyses of CXXC motif in the DNA polymerase III, tau/gamma subunit (DnaX) protein. This protein contains a CXXC motif and this motif is conserved through the different bacterial species.



Supplementary Figure 15K: Comparative genomics and structural analyses of CXXC motif in the 8-oxoguanine DNA glycosylase (MutY). This protein contains a CXXC motif and this motif is conserved through the different bacterial species.