RESEARCH ARTICLE

INTERNATIONAL MICROBIOLOGY (2007) 10:201-208
DOI: 10.2436/20.1501.01.28 ISSN: 1139-6709 **www.im.microbios.org**

INTERNATIONAL MICROBIOLOGY

Identification of signature proteins that are distinctive of the *Deinococcus-Thermus* phylum

Emma Griffiths, Radhey S. Gupta*

Department of Biochemistry, McMaster University, Hamilton, Ontario, Canada

Received 23 June 2007 · Accepted 30 July 2007

Summary. The members of the *Deinococcus-Thermus* phylum, which include many species that are resistant to extreme radiation, as well as several thermophiles, have been recognized solely on the basis of their branching patterns in 16S rRNA and other phylogenetic trees. No biochemical or physiological characteristic is currently known that is unique to this group of species. To identify genes/proteins that are exclusive of this group of species, systematic protein basic local alignment tool (Blastp) searches were carried out on each open reading frame (ORF) in the genome of *Deinococcus radiodurans*. These studies identified 65 proteins that were only found in all three sequenced *Deinococcus-Thermus* genomes (viz. *D. radiodurans*, *D. geothermalis* and *Thermus thermophilus*), but not in any other bacteria. In addition, these studies also identified 206 proteins that are exclusively found in the two *Deinococcocci* species, and 399 proteins that are unique to *D. radiodurans*. The identified proteins, which represent a genetic repertoire distinctive to the *Deinococcus-Thermus* group, or to *Deinococci* species, provide novel molecular markers for their identification and characterization. The cellular functions of most of these proteins are not known and their studies should prove useful in identifying novel biochemical and physiological characteristics that are exclusive of these groups of bacteria and also those responsible for the extreme radiation resistance of *Deinococci*. [**Int Microbiol** 2007; 10(3):201-208]

Key words: *Deinococcus* spp. · *Thermus thermophilus* · *Deinococci*-specific proteins · radiation-resistant bacteria · extremophilic bacteria · ORFans proteins · lateral gene transfer

Introduction

The *Deinococcus-Thermus* group of species has enormous biochemical, physiological and phenotypic diversity [12]. While most species from this phylum stain gram-negative, and all of them contain an outer membrane, which is a distinctive feature of gram-negative bacteria [16], those belonging to the genus *Deinococcus* have a thick layer of peptidoglycan that causes them to stain gram-positive [12,28]. Of the two main groups that comprise this phylum, the *Deinococci* species are known for their very high resistance to ionizing radiations; they can withstand up to 150 kGy (1.5 Mrads), which is more than 100-fold increased resistance as compared to *Escherichia coli* [3,6,12,26,27]. In addition, these organisms are also more resistant to ultraviolet radiation, oxidizing agents, and dessicating conditions [6,12,33]. In contrast to the *Deinococci,* the *Thermales* are highly thermophilic organisms; they inhabit niches up to 80ºC, but they do not have unusual radiation nor stress resistance capabilities [12]. The *Deinococcus-Thermus* phylum comprised two

***Corresponding author:** R.S. Gupta
Department of Biochemistry
McMaster University
Hamilton, Ontario, Canada L8N 3Z5
Tel. +1-9055259140 ext. 22639. Fax +1-9055229033
E-mail: gupta@mcmaster.ca

This article contains supporting information [SI] ONLINE, at: www.im.microbios.org

orders or families, *Deinococcaceae* and *Thermaceae* [1,12, 13]. However, recently a third family, *Trueperaceae*, has also been proposed [1]. The *Deinococcus-Thermus* group of species are thought to have a common ancestor due to their close branching in the16S rRNA, as well as other phylogenetic trees [4,9,15,24,26,32,37]. Nevertheless, except for a number of conserved inserts or deletions (i.e., indels) in protein sequences that are exclusive to these bacteria [14], no other distinctive biochemical or molecular characteristic is known that is specific for this diverse group.

The complete genomes of three species from this group, viz. *Deinococcus radiodurans* R1 [26], *Deinococcus geothermalis* DSM 11300 (sequenced by the DOE Joint Genome Institute) and *Thermus thermophilus* strains HB27 [20] and HB8, are now available. The 3.28-Mb *D. radiodurans* genome is segmented over two chromosomes and two extrachromosomal elements (MP1 and CP1 plasmids) [26], while the 2.12–2.13-Mb *T. thermophilus* genomes consist of one chromosome and one or two extrachromosomal plasmids (pTT27 in both strains, HB8 also has a second plasmid pTT8) [20]. Earlier studies have shown that *T. thermophilus* has far fewer predicted protein-coding genes (2210 genes) than *D. radiodurans* (3191 genes) and, of these, only 1081 belong to a shared gene core [33]. Notably, many of the expanded paralogous gene families of *D. radiodurans* were absent in *Thermus*, and vice versa [33]. However, these previous studies have not identified proteins that are shared only by the *Deinococcus-Thermus* species, or by the two *Deinococci*.

In this work, we have performed systematic protein basic local alignment tool (Blastp) searches on all open reading frames (ORFs) in the *D. radiodurans* genome to identify proteins that are uniquely shared by species from this phylum but not found in other organisms. Similar investigations on other groups of bacteria have shown that such proteins provide particularly useful tools for taxonomic, biochemical and molecular biological studies [11,17–19,22]. Although proteins which are specific for either *D. radiodurans* or *T. thermophilus* were reported in earlier genomic publications [20,26], the proteins that are shared uniquely by the available species from these two divergent groups (i.e., *Deinococcus* and *Thermus*) were not examined. Subsequent availability of the genome of *D. geothermalis* has also facilitated identification of proteins that are specific for the *Deinococcus* genus.

Materials and methods

Blastp searches were carried out on each ORF in the *D. radiodurans* genome against all available sequences in the GenBank non-redundant database to identify proteins for which all significant blast hits obtained were from the *Deinococcus-Thermus*-group of species [2]. The searches were performed using the default parameters of the Blastp program, which included the low complexity filter. The expected (E) values for all Blastp results were individually examined to identify proteins that were specific for the *Deinococcus-Thermus* group. Proteins of interest were those in which either all significant hits were from this group or which involved a large increase in E values from the last hit for the *Deinococcus-Thermus* species to the first hit from any other bacteria and the latter E values were $>1\ e^{-04}$, indicating weak similarity that can occur by chance. However, for smaller proteins, higher E values were often considered significant, as the magnitude of the E value depends upon the length of the query sequence [2]. In this study, we have also retained some proteins in which one or two isolated hits from other groups of bacteria had acceptable E values, as they provide possible cases of lateral gene transfers (LGT). For all proteins of interest, their loci numbers in *D. radiodurans* genome, accession numbers, protein lengths, and any information concerning their cellular function were tabulated and presented.

Results

Proteins specific for *Deinococcus-Thermus*. Our comparative genomic analysis has identified 65 proteins (or ORFs) that are uniquely present in all three sequenced *Deinococcus-Thermus* genomes, but they are absent from all other species (Table 1). Except for four of these proteins, for which a putative functional assignment has been made, all the remaining proteins are of unknown functions (Table 1). Of the proteins with an assigned function, DR1021 is an S-layer like protein. The *Deinococci* and *Thermales* are surrounded by a crystalline proteinaceous shell known as the S-layer, which arrays itself in a regular hexagonal geometry [25,30]. In addition to their role as protective coats, the S-layers have adapted new functions, e.g., as molecular sieves, attachment sites for extracellular enzymes, and virulence factors [34]. DR1021 probably carries out a novel function in the S-layer, which is specific for the *Deinococcus-Thermus* species. Another *Deinococcus-Thermus*-specific protein, DR1194, is a putative Fe-S binding reductase. The two other proteins with putative assigned functions are a V-type ATP synthase E subunit (DR0697) and a diacylglycerol kinase (DR2093).

Among the proteins that are specific for *Deinococcus-Thermus*, 13 of them are present in six gene clusters (viz. DR0041-DR0042, DR0213-DR0214, DR1230-DR1231, DR1872-DR1873, DR2006-DR2007 and DR2318-DR2319-DR2320) on chromosome 1. Two additional proteins, DR0040 and DR0212, which are located immediately adjacent to two of these gene clusters, are specific for the *Deinococci*. It is possible that the proteins in these gene clusters are part of a functional unit (e.g., operon) and thus may be involved in related functions. Of the *Deinococcus-Thermus*-specific proteins, many of them (e.g., DR0042, DR0213, DR0573, DR1013, DR1230, DR1231, DR1750, DR2319, DRA0288, DRA0369 and DRB0145) show high degree of sequence conservation. Sequence alignment for two of these proteins,

Table 1. Proteins specific for *Deinococcus-Thermus*

Gene ID/Accession no.	Length	Function
Chromosome 1		
DR0041/NP 293767	192aa	Unknown
DR0042/NP 293768	257aa	Unknown
DR0060/NP 293786	142aa	Unknown
DR0090/NP 293816	232aa	Unknown
DR0116/NP 293842	129aa	Unknown
DR0131/NP 293857	550aa	Unknown
DR0213/NP 293937	305aa	Unknown
DR0214/NP 293938	259aa	Unknown
DR0459/NP 294182	236aa	Unknown
DR0548/NP 294271	314aa	Unknown
DR0570/NP 294293	93aa	Unknown
DR0573/NP 294296	971aa	Unknown
DR0697/NP 294420	185aa	V-ATP synth, E sub[a]
DR0727/NP 294450	151aa	Unknown
DR0771/NP 294495	238aa	Unknown
DR0900/NP 294624	149aa	Unknown
DR0972/NP 294696	231aa	Unknown
DR0988/NP 294712	160aa	Unknown
DR0994/NP 294718	112aa	Unknown
DR1000/NP 294724	195aa	Unknown
DR1013/NP 294737	257aa	Unknown
DR1021/NP 294745	167aa	S-layer-like protein
DR1088/NP 294812	288aa	Unknown
DR1139/NP 294863	131aa	Unknown
DR1166/NP 294890	194aa	Unknown
DR1194/NP 294918	166aa	Fe-S reductase[b]
DR1201/NP 294925	1021aa	Unknown
DR1221/NP 294945	245aa	Unknown
DR1230/NP 294954	218aa	Unknown
DR1231/NP 294955	802aa	Unknown
DR1388/NP 295111	217aa	Unknown
DR1460/NP 295183	675aa	Unknown
DR1474/NP 295197	228aa	Unknown
DR1479/NP 295202	74aa	Unknown
DR1750/NP 295473	124aa	Unknown
DR1752/NP 295475	548aa	Unknown
DR1768/NP 295491	134aa	Unknown
DR1852/NP 295575	369aa	Unknown
DR1870/NP 295593	284aa	Unknown
DR1872/NP 295595	73aa	Unknown
DR1873/NP 295596	637aa	Unknown
DR1924/NP 295647	168aa	Unknown
DR1981/NP 295704	184aa	Unknown
DR2006/NP 295729	157aa	Unknown
DR2007/NP 295730	79aa	Unknown
DR2093/NP 295816	127aa	Dk[c]
DR2136/NP 295859	206aa	Unknown
DR1564/NP 295287	278aa	Unknown
DR2156/NP 295879	128aa	Unknown
DR2286/NP 296007	502aa	Unknown
DR2295/NP 296016	137aa	Unknown
DR2297/NP 296018	98aa	Unknown
DR2318/NP 296039	210aa	Unknown
DR2319/NP 296040	303aa	Unknown
DR2320/NP 296041	366aa	Unknown
DR2349/NP 296070	270aa	Unknown
DR2472/NP 296192	169aa	Unknown
DR2491/NP 296211	104aa	Unknown
DR2500/NP 296220	157aa	Unknown
DR2527/NP 296247	188aa	Unknown
Chromosome 2		
DRA0046/NP 285370	384aa	Unknown
DRA0282/NP 285605	504aa	Unknown
DRA0288/NP 285611	354aa	Unknown
DRA0369/NP 285692	402aa	Unknown
MP1 Plasmid		
DRB0145/NP 051674	397aa	Unknown

All significant BLAST hits for these proteins are from *Deinococcus-Thermus* species.
[a]V-type ATP synthase, E subunit.
[b]Fe-S binding reductase, putative.
[c]Diacylglycerol kinase.

DR042 and DR1750, are presented in Fig. 1. These alignments contain several stretches of identical residues that can be used for the design of PCR primers or other molecular probes for these bacteria.

Proteins specific for *Deinococci*. These studies have also identified 206 proteins or ORFs whose homologs are found only in *D. radiodurans* and *D. geothermalis* (Table 2). For five additional proteins (marked with *), a homolog showing significant similarity is also present in an isolated bacterial species from other groups (viz. DR0082 and DR2593, *Symbiobacterium thermophilum*; DR0691, *Mesorhizobium* sp.; DR1098, *Arthrobacter* sp.; DR1644, *Bacillus subtilis*), which can be attributed to lateral gene transfers. Similar to the *Deinococcus-Thermus*-specific proteins, all except five of the 211 proteins listed in Table 2 are of unknown functions. The five proteins with predicted function are involved in a variety of cellular activities. These include a putative V-type ATP synthase K subunit (DR0696), a putative M-related protein (DR0920), which emanate from the cell surface and provide an anchoring motif for cell surface proteins [5], a putative peptidase (DR2058) and a protein (DR2347), which has been annotated as Roadblock/LC7 protein [35]. Another *Deinococci* specific protein DRA0346, located on chromosome 2, has been annotated as the DNA damage repair protein PprA (or PprI) and it may play a significant role in the radiation resistance of the *Deinococci* species [21,29]. Mutation studies with the PprI have shown that this protein acts as a general switch controlling downstream DNA repair pathways [21]. *In vitro* studies provide

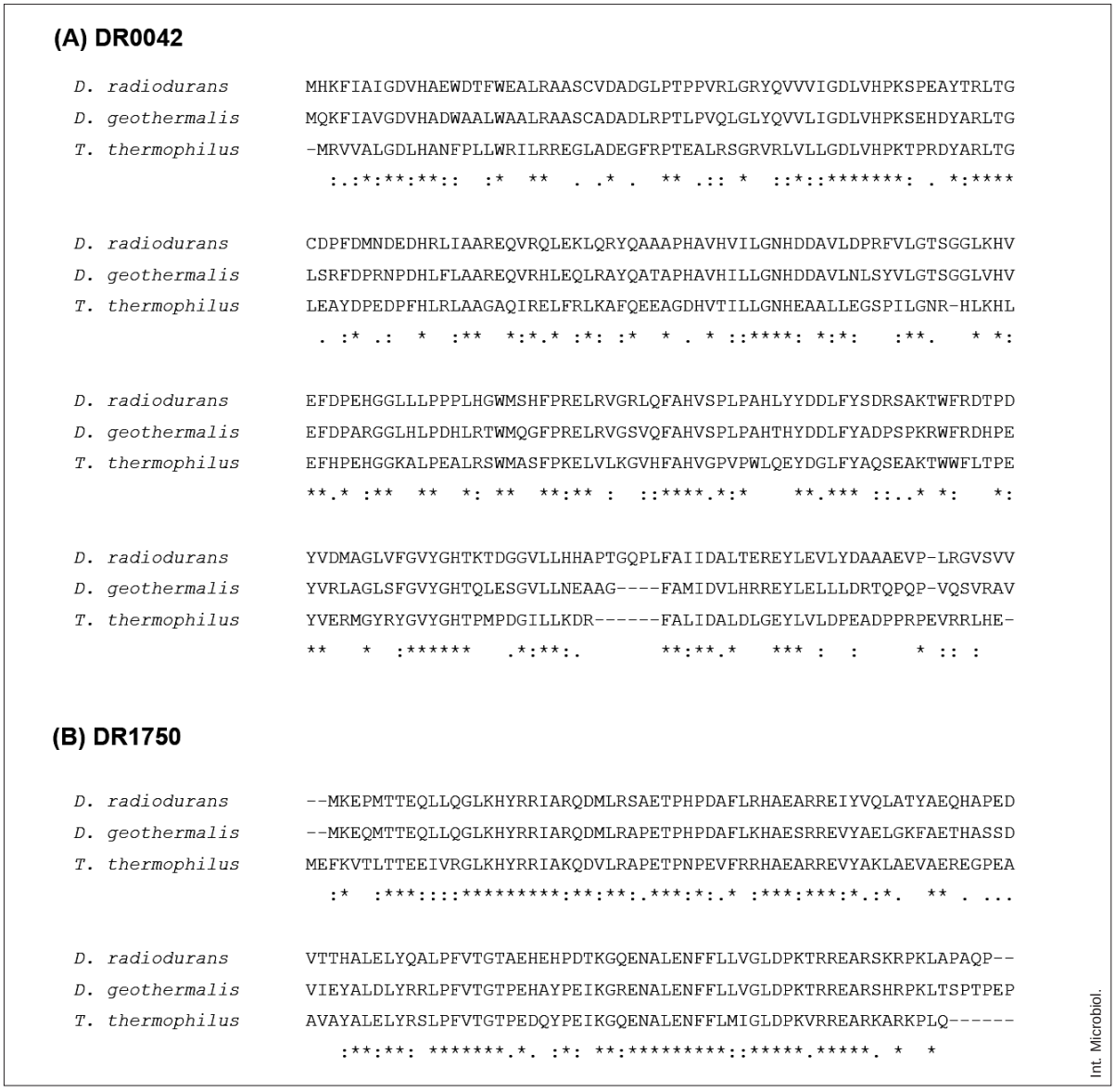

Fig. 1. Sequence alignments of two Deinococcus-Thermus-specific proteins (**A**) DR042 (accession no. NP 293768) and (**B**) DR1750 (accession no. NP 295473) for all available homologs. The * indicate residues that are completely conserved, whereas those marked **.** or **:** denote residues where conservative substitutions have occurred.

evidence that PprA binds preferentially to double-stranded DNA carrying strand breaks, and it stimulates the DNA end-joining reaction catalyzed by DNA ligases [29].

Among the proteins that are specific for *Deinococci*, 45 proteins were found in 20 clusters each containing between 2–4 genes. These clusters were composed of the following loci: DR0067-DR0068, DR0359-DR0360, DR0637-DR0638, DR0672-DR0673, DR0736-DR0737, DR0849-DR0850, DR0863-DR0864, DR1371-DR1372, DR1404-DR1405-DR1406-DR1407, DR1421-DR1422, DR14254-DR1455, DR1483-DR1484), DR1780-DR1781, DR1830-DR1831, DR2001-DR2002-DR2003, DR2157-DR2158, DR2229-DR2230, DR2347-DR2348-DR2349-DR2350, DR2558-DR2559 and DRA0140-DRA0141.

Proteins specific for *Deinococcus radiodurans*. These studies have also revealed 399 ORFs that are specific for *D. radiodurans* and not found in any other species including *D. geothermalis*. All except four of these proteins are of unknown function, as indicated in Table 3 [SI] ONLINE. Of the four genes with predicted functions, DR0092 is related to the MutT/nudix family of proteins. The MutT

Table 2. Proteins specific for *Deinococci* (*D. radiodurans* and *D. geothermalis*)

Gene ID/ Accession	Length	Gene ID/ Accession	Length	Gene ID/ Accession	Length	Gene ID/ Accession	Length
Chromosome 1		DR0800/NP 294524	60aa	DR1421/NP 295144	93aa	DR2020/NP 295743	131aa
DR0003/NP 293729	291aa	DR0849/NP 294573	177aa	DR1422/NP 295145	190aa	DR2058/NP 295781	328aa [3]
DR0015/NP 293741	118aa	DR0850/NP 294574	105aa	DR1425/NP 295148	206aa	DR2077/NP 295800	169aa
DR0018/NP 293744	205aa	DR0857/NP 294581	285aa	DR1429/NP 295152	67aa	DR2090/NP 295813	553aa
DR0021/NP 293747	88aa	DR0863/NP 294587	295aa	DR1432/NP 295155	89aa	DR2132/NP 295855	150aa
DR0037/NP 293763	201aa	DR0864/NP 294588	293aa	DR1446/NP 295169	142aa	DR2157/NP 295880	67aa
DR0040/NP 293766	158aa	DR0869/NP 294593	228aa	DR1454/NP 295177	238aa	DR2158/NP 295881	76aa
DR0047/NP 293773	65aa	DR0887/NP 294611	128aa	DR1455/NP 295178	98aa	DR2172/NP 295895	139aa
DR0061/NP 293787	478aa	DR0889/NP 294613	116aa	DR1465/NP 295188	188aa	DR2183/NP 295906	210aa
DR0067/NP 293793	309aa	DR0895/NP 294619	128aa	DR1469/NP 295192	477aa	DR2193/NP 295915	90aa
DR0068/NP 293794	124aa	DR0903/NP 294627	518aa	DR1483/NP 295206	316aa	DR2207/NP 295929	95aa
DR0070/NP 293796	199aa	DR0909/NP 294633	160aa	DR1484/NP 295207	163aa	DR2229/NP 295951	175aa
DR0072/NP 293798	179aa	DR0913/NP 294637	117aa	DR1490/NP 295213	218aa	DR2230/NP 295952	75aa
DR0082/NP 293808*	154aa	DR0920/NP 294644	87aa [2]	DR1527/NP 295250	133aa	DR2237/NP 295959	122aa
DR0124/NP 293850	94aa	DR0931/NP 294655	231aa	DR1539/NP 295262	110aa	DR2240/NP 295962	142aa
DR0193/NP 293917	116aa	DR0938/NP 294662	395aa	DR1557/NP 295280	343aa	DR2271/NP 295993	695aa
DR0212/NP 293936	101aa	DR0956/NP 294680	260aa	DR1575/NP 296298	215aa	DR2281/NP 296002	164aa
DR0218/NP 293942	146aa	DR0993/NP 294717	116aa	DR1584/NP 295307	152aa	DR2292/NP 296013	85aa
DR0239/NP 293963	134aa	DR1048/NP 294772	162aa	DR1590/NP 295313	183aa	DR2314/NP 296035	117aa
DR0269/NP 293992	336aa	DR1098/NP 294822*	346aa	DR1603/NP 295326	360aa	DR2334/NP 296055	386aa
DR0296/NP 294019	98aa	DR1104/NP 294828	95aa	DR1607/NP 295330	144aa	DR2342/NP 296063	220aa
DR0308/NP 294031	147aa	DR1116/NP 294840	341aa	DR1615/NP 295338	124aa	DR2344/NP 296065	177aa
DR0326/NP 294049	198aa	DR1125/NP 294849	115aa	DR1643/NP 295366	81aa	DR2347/NP 296068	332aa [4]
DR0338/NP 294061	128aa	DR1138/NP 294864	254aa	DR1644/NP 295367*	210aa	DR2348/NP 296069	559aa
DR0352/NP 294075	157aa	DR1140/NP 294864	124aa	DR1663/NP295386	153aa	DR2350/NP 296071	418aa
DR0355/NP 294078	102aa	DR1172/NP 294896	298aa	DR1665/NP 295388	152aa	DR2414/NP 296135	157aa
DR0359/NP 294082	159aa	DR1180/NP 294904	169aa	DR1671/NP 295394	172aa	DR2458/NP296178	89aa
DR0360/NP 294083	143aa	DR1195/NP 294919	296aa	DR1693/NP 295416	250aa	DR2504/NP 296224*	274aa
DR0367/NP 294090	266aa	DR1202/NP 294926	249aa	DR1697/NP 295420	86aa	DR2541/NP 296261	177aa
DR0381/NP 294104	134aa	DR1210/NP 294934	225aa	DR1699/NP 295422	262aa	DR2554/NP 296274	222aa
DR0437/NP 294160	72aa	DR1212/NP 294936	168aa	DR1744/NP 295467	170aa	DR2558/NP 296278	156aa
DR0449/NP 294172	243aa	DR1216/NP 294940	158aa	DR1770/NP 295493	247aa	DR2559/NP 296279	75aa
DR0452/NP 294175	132aa	DR1218/NP 294942	114aa	DR1780/NP 295503	354aa	DR2563/NP 296283	70aa
DR0458/NP 294181	839aa	DR1226/NP 294950	363aa	DR1781/NP 295504	473aa	DR2569/NP 296289	270aa
DR0545/NP 294268	95aa	DR1234/NP 294958	293aa	DR1786/NP 295509	102aa	DR2572/NP 296292	496aa
DR0554/NP 294277	244aa	DR1242/NP 294966	180aa	DR1788/NP 295511	127aa	DR2593/NP 296312*	185aa
DR0569/NP 294292	342aa	DR1245/NP 294966	165aa	DR1816/NP 295539	258aa	**Chromosome 2**	
DR0571/NP 294294	397aa	DR1249/NP 294973	192aa	DR1820/NP 295543	307aa	DRA0056/NP 285379	162aa
DR0574/NP 294297	370aa	DR1254/NP 294978	114aa	DR1830/NP 295553	139aa	DRA0140/NP 285464	272aa
DR0600/NP 294323	79aa	DR1256/NP 294980	87aa	DR1831/NP 295554	291aa	DRA0141/NP 285465	144aa
DR0637/NP 294360	77aa	DR1269/NP 294993	124aa	DR1833/NP 295556	144aa	DRA0215/NP 285538	131aa
DR0638/NP 294361	142aa	DR1272/NP 294996	102aa	DR1840/NP 295563	79aa	DRA0219/NP 285542	268aa
DR0644/NP 294367	206aa	DR1293/NP 295017	276aa	DR1847/NP 295570	90aa	DRA0346/NP 285669	300aa [5]
DR0672/NP 294395	169aa	DR1301/NP 295025	231aa	DR1867/NP 295590	120aa	**CP1Plasmid**	
DR0673/NP 294396	195aa	DR1315/NP 295039	210aa	DR1881/NP295604*	113aa	DRC0030/NP 051700	432aa
DR0691/NP294414*	254aa	DR1320/NP 295044	176aa	DR1882/NP 295605	142aa	**Megaplasmid**	
DR0696/NP 294419	101aa [1]	DR1331/NP 295055	81aa	DR1887/NP 295610	252aa	DRB0035/NP 051576	318aa
DR0714/NP 294437	266aa	DR1364/NP 295087	155aa	DR1896/NP 295619	94aa	DRB0110/NP 051642	87aa
DR0724/NP 294447	617aa	DR1371/NP 295094	313aa	DR1923/NP 295646	276aa	DRB0116/NP 051647	237aa
DR0736/NP 294459	136aa	DR1372/NP 295095	164aa	DR1962/NP 295685	170aa		
DR0737/NP 294460	213aa	DR1404/NP 295127	73aa	DR1987/NP 295710	173aa		
DR0746/NP 294469	79aa	DR1405/NP 295128	107aa	DR2001/NP 295724	103aa		
DR0760/NP 294484	341aa	DR1406/NP 295129	157aa	DR2002/NP 295725	129aa		
DR0769/NP 294493	267aa	DR1407/NP 295130	365aa	DR2003/NP 295726	210aa		
DR0780/NP 294504	92aa	DR1414/NP 295137	527aa	DR2006/NP 295729	157aa		
DR0795/NP 294519	211aa	DR1416/NP 295139	210aa	DR2016/NP 295739	79aa		

[1]V-type ATP synthetase, subunit K; [2]M-related protein; [3]Peptidase, archeal and bacterial C-terminal; [4]Roadblock/LC7 protein;

[5]DNA Damage repair protein PprI; * One significant hit also found in other bacteria.

proteins are key component of an antimutagenic system that prevents the incorporation of 8-oxo-dGTP into DNA [26]. DRA0254 is annotated as a *D. radiodurans* specific resolvase that participates in site-specific homologous recombination [38]. DRA0268 is a *D. radiodurans*-specific adenine deaminase-related protein. The last of the proteins with annotated function, DRC0001, is a *D. radiodurans* specific cytochrome P450-related protein. Of the various *D. radiodurans* specific proteins, the genes for the following were clustered in the genome:

Chromosome 1: (DR0103-0105), (DR0206-DR0208), (DR0285-0286), (DR0514-0529), (DR0531-0533), (DR0539-0540), (DR0591-0592), (DR0648-0650), (DR0663-0665), (DR0725/0726), (DR0817-0818), (DR0838-0839), (DR1043-1045), (DR1128-1129), (DR1134-1136), (DR1250-1251), (DR1264-1265), (DR1408-1409), (DR1463-1466), (DR1639-1640), (DR1657-1658), (DR1801-1802), (DR1840-1842), (DR1859-1860), (DR1900-1901), (DR1951-1953), (DR2037-2039), (DR2060-2061), (DR2082-2083), (DR2137-2138), (DR2170-2171), (DR2309-2310), (DR2401-2403), (DR2445-2446), (DR2455-2457), (DR2485-2486), (DR2532-2533).

Chromosome 2: (DRA0077-0096), (DRA0100-0102), (DRA0104-0107), (DRA0109-0111), (DRA0113-0116), (DRA0295-0296), (DRA0305-0307).

MP1 plasmid: (DRB0003-0004), (DRB0021-0022), (DRB0060-0061), (DRB0063-0064), (DRB0084-0085).
CP1 plasmid: (DRC0021-0024) and (DRC0026/0027).

Two of these gene clusters, DR0514-0529 and DRA0077-0096, are particularly large and contain 16 and 20 proteins, respectively.

Discussion

The work presented here is the first report on identifying specific sets of genes/proteins that are uniquely shared by the *Deinococcus-Thermus* group of species. Our analyses have identified 65 proteins that are uniquely found in the sequenced *Deinococcus-Thermus* species to the exclusion of all other species. Sequence information from *Deinococcus-Thermus* phylum is presently available for only a limited number of species. However, the sequenced genomes include species from both the main families (i.e., *Deinococcaceae* and *Thermaceae*) within this phylum. Therefore, it is probable that many of the genes/proteins identified here will prove to be distinctive characteristics of this phylum. We have also identified >200 genes/proteins that are exclusive to the two sequenced *Deinococcus* species. Although some of these proteins may prove to be unique to only these two species, many other proteins in this set may be found in other *Deinococci* species and could comprise distinctive characteristics of the *Deinococcaceae* family. The unique presence of these proteins in the *Deinococcus-Thermus* and *Deinococci* species, suggests that their genes first evolved in a common ancestor of these taxonomic groups.

The functions of most of the *Deinococcus-Thermus, Deinococci* and *D. radiodurans*-specific proteins identified in the present work are still unknown. However, significant proportions of their genes were located in distinct clusters in the *D. radiodurans* genome. Although it is not clear whether these gene clusters form functional units, this observation suggests that they could be involved in related functions [7,8]. The numbers of *D. radiodurans* specific ORFans (i.e., ORFs that have not matches with other ORFs) identified in the present work (about 400) is considerably smaller than that reported (720 proteins) previously for the same species [26]. We believe this difference is largely due to a paucity of sequence information available in the past, which hindered the identification of homologs in other species. Lateral gene transfer has widely been implicated in the evolution of members of the *Deinococcus-Themus* group, which are thought to have received genes from a number of other phyla from Archaea, Eukarya, and the cyanobacteria [10,23,26,31,33]. However, except for five of the *Deinococci*-specific proteins (see Table 2), for the remainder of the proteins described here, no homologs showing significant similarity were found in any other bacteria. These results suggest that the genes for these particular proteins have not been laterally transferred, either to or from *Deinococcus-Thermus* species to other bacteria.

The biochemical basis of extreme resistance of *Deinococcus* spp. to ionizing radiation is poorly understood [6,26, 36]. The sequencing of *D. radiodurans* genome has revealed that most of the DNA repair proteins found in *E. coli* are also present in *D. radiodurans* [36]. However, it is probable that *D. radiodurans* also has unique mechanisms for dealing with ionizing radiation-induced DNA damage. The genes/proteins that are exclusive to the *Deinococci* and/or *D. radiodurans* provide important candidates for such mechanisms. Analysis of the transcriptional response of *D. radiodurans* to ionizing radiation and desiccation revealed that novel proteins contribute to resistance under both stress conditions [36]. Of the 33 genes that are upregulated under these conditions, 20 ORFs are of unknown function [36]. Our analysis showed that three of these ORFs consisted of *Deinococci*-specific

proteins, DR0003 (DdrC), DR0070 (DdrB) and DRA0346 (PprA), while an additional 4 proteins DR0219 (DdrF), DR0227 (DdrG), DR0438 (DdrH) and DR1264 (DdrK) were *D. radiodurans*-specific. Thus, while some of the mechanisms to deal with DNA damage are shared by the *Deinococci* as a genus, others may be specific for *D. radiodurans*. In another proteomic study on *D. radiodurans* cells recovering from gamma-irradiation, 26 proteins were identified whose expression was significantly enhanced under these conditions [39]. These proteins are involved in a wide variety of house-keeping functions including transcription, translation, DNA replication, general chaperones, proteases, nucleotide and energy metabolisms, etc. The results of these studies agree with the finding of Karlin and Mrazek [23]: that an unusually large number of the genes from *D. radiodurans,* which have the characteristics of highly expressed genes, correspond to functions such as chaperones, proteases and detoxifying enzymes, and that this indicates the importance of such function in the unusual radiation resistance of this organism.

In addition to the proteins that are specific for the *Deinococcus-Thermus* or *Deinococci* species, we have recently described eight conserved indels in seven widely distributed proteins (viz. seryl-tRNA synthetase, threonyl-tRNA synthetase, RNA polymerase β′ subunit, signal recognition particle protein Ffh/SR54, major sigma factor 70, ribosomal protein L1 and UvrA) that are distinctive characteristics of various species from the *Deinococcus-Thermus* phylum [14]. These *Deinococcus-Thermus*-specific proteins and indels provide new and powerful means for the identification and classification of these bacteria as well for their genetic and biochemical studies. The understanding of the cellular functions of these proteins and indels should also lead to the identification of novel biochemical and physiological characteristics that are uniquely shared by the *Deinococci* and *Thermales*.

Acknowledgements. This work was supported by a research grant from the National Science and Engineering Research Council of Canada. We thank Divya Wilson for assistance in analyzing the results of Blast searches.

References

1. Albuquerque L, Simões C, Nobre MF, Pino NM, Battista JR, Silva MT, Rainey FA, da Costa MS (2005) *Truepera radiovictrix* gen. nov., sp. nov., a new radiation resistant species and the proposal of *Trueperaceae* fam. nov. FEMS Microbiol Lett 247:161-169
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein databases search programs. Nucleic Acids Res 25:3389-3402
3. Battista JR, Earl AM, Park MJ (1999) Why is *Deinococcus radiodurans* so resistant to ionizing radiation? Trends Microbiol 7:362-365
4. Bustard K, Gupta RS (1997) The sequences of heat shock protein 40 (DnaJ) homologs provide evidence for a close evolutionary relationship between the *Deinococcus-Thermus* group and cyanobacteria. J Mol Evol 45:193-205
5. Courtney HS, Hasty DL, Dale JB (2006) Anti-phagocytic mechanisms of *Streptococcus pyogenes*: binding of fibrinogen to M-related protein. Mol Microbiol 59:936-947
6. Cox MM, Battista JR (2005) *Deinococcus radiodurans*–the consummate survivor. Nat Rev Microbiol 3:882-892
7. Danchin A (1999) From protein sequence to function. Curr Opin Struct Biol 9:363-367
8. Doerks T, von Mering C, Bork P (2004) Functional clues for hypothetical proteins based on genomic context analysis in prokaryotes. Nucleic Acids Res 32:6321-6326
9. Eisen JA (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
10. Elena SF, Whittam TS, Winkworth CL, Riley MA, Lenski RE (2005) Genomic divergence of *Escherichia coli* strains: evidence for horizontal transfer and variation in mutations rates. Int Microbiol 8:271-278
11. Gao B, Gupta RS (2007) Phylogenomic analysis of proteins that are distinctive of *Archaea* and its main subgroups and the origin of methanogenesis. BMC Genomics 8:86
12. 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
13. Garrity GM, Bell JA, Lilburn TG (2005) The revised road map to the *Manual*. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM (eds) Bergey's manual of systematic bacteriology, 2nd edn, vol 2. Springer, New York, pp 159-220
14. Griffiths E, Gupta RS (2004) Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the *Deinococcus-Thermus* phylum. J Bacteriol 186:3097-3107
15. Gupta RS, Bustard K, Falah M, Singh D (1997) Sequencing of heat shock protein 70 (DnaK) homologs from *Deinococcus proteolyticus* and *Thermomicrobium roseum* and their integration in a protein-based phylogeny of prokaryotes. J Bacteriol 179:345-357
16. Gupta RS (1998) Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol Mol Biol Rev 62:1435-1491
17. Gupta RS (2001) The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. Int Microbiol 4:187-202
18. Gupta RS (2006) Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria (*Campylobacterales*). BMC Genomics 7:167
19. Gupta RS, Griffiths E (2006) *Chlamydiae*-specific proteins and indels: novel tools for studies. Trends Microbiol 14:527-535
20. Henne A, Brüggemann H, Raasch C, et al. (2004) The genome sequence of the extreme thermophile *Thermus thermophilus*. Nat Biotechnol 22:547-553
21. Hua Y, Narumi I, Gao G, Tian B, Satoh K, Kitayama S, Shen B (2003) PprI: a general switch responsible for extreme radioresistance of *Deinococcus radiodurans*. Biochem Biophys Res Commun 306:354-360
22. Kainth P, Gupta RS (2005) Signature proteins that are distinctive of Alpha Proteobacteria. BMC Genomics 6:94
23. Karlin S, Mrazek J (2001) Predicted highly expressed and putative alien genes of *Deinococcus radiodurans* and implications for resistance to ionizing radiation damage. Proc Natl Acad Sci USA 98:5240-5245
24. Karlin S, Weinstock GM, Brendel V (1995) Bacterial classifications derived from recA protein sequence comparisons. J Bacteriol 177: 6881-6893
25. Koch AL (2003) Were Gram-positive rods the first bacteria? Trends Microbiol 11:166-170

26. Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, Koonin EV, Daly MJ (2001) Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. Microbiol Mol Biol Rev 65:44-79
27. Misra HS, Khairnar NP, Kota S, Shrivastava S, Joshi VP, Apte SK (2006) An exonuclease I-sensitive DNA repair pathway in *Deinococcus radiodurans*: a major determinant of radiation resistance. Mol Microbiol 59:1308-1316
28. Murray RGE (1992) The family *Deinococcaceae*. In: Balows A, Trüper HG, Dworkin M, Harder W, Schleifer KH (eds) The Prokaryotes, Springer, New York, pp 3732-3744
29. Narumi I, Satoh K, Cui S, Funayama T, Kitayama S, Watanabe H (2004) PprA: a novel protein from *Deinococcus radiodurans* that stimulates DNA ligation. Mol Microbiol 54: 278-285
30. Olabarria G, Carrascosa JL, de Pedro MA, Berenguer J (1996) A conserved motif in S-layer proteins is involved in peptidoglycan binding in *Thermus thermophilus*. J Bacteriol 178:4765-4772
31. Olendzenski L, Liu L, Zhaxybayeva O, Murphey R, Shin DG, Gogarten JP (2000) Horizontal transfer of archaeal genes into the Deinococcaceae: detection by molecular and computer-based approaches. J Mol Evol 51:587-599
32. Olsen GJ, Woese CR, Overbeek R (1994) The winds of (evolutionary) change: breathing new life into microbiology. J Bacteriol 176:1-6
33. Omelchenko MV, Wolf YI, Gaidamakova EK, et al. (2005) Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance. BMC Evol Biol 5:57
34. Sleytr UB, Beveridge TJ (1999) Bacterial S-layers. Trends Microbiol 7:253-260
35. Song J, Tyler RC, Lee MS, Tyler EM, Markley JL (2005) Solution structure of isoform 1 of Roadblock/LC7, a light chain in the dynein complex. J Mol Biol 354:1043-1051
36. Tanaka M, Earl AM, Howell HA, Park MJ, Eisen JA, Peterson SN, Battista JR (2004) Analysis of *Deinococcus radiodurans*'s transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance. Genetics 168:21-33
37. Weisburg WG, Giovannoni SJ, Woese CR (1989) The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. Syst Appl Microbiol 11:128-134
38. Zahradka D, Zahradka K, Petranovic M, Dermic D, Brcic-Kostic K (2002) The RuvABC resolvase is indispensable for recombinational repair in *sbcB15* mutants of *Escherichia coli*. J Bacteriol 184:4141-4147
39. Zhang C, Wei J, Zheng Z, Ying N, Sheng D, Hua Y (2005) Proteomic analysis of *Deinococcus radiodurans* recovering from γ-irradiation. Proteomics 5:138-143