MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Mar. 2001, p. 44–79 Vol. 65, No. 1
1092-2172/01/$04.00+0 DOI: 10.1128/MMBR.65.1.44–79.2001

Genome of the Extremely Radiation-Resistant Bacterium *Deinococcus radiodurans* Viewed from the Perspective of Comparative Genomics

KIRA S. MAKAROVA,[1,2] L. ARAVIND,[2] YURI I. WOLF,[2] ROMAN L. TATUSOV,[2] KENNETH W. MINTON,[1] EUGENE V. KOONIN,[2] AND MICHAEL J. DALY[1]*

Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-4799,[1] and National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20814[2]

INTRODUCTION 44
Extreme Radiation Resistance 44
Isolation 44
Cell Structure 45
DNA Damage Resistance 46
Logistics of Extreme DNA Damage Resistance 46
DNA Repair Pathways 47
SEQUENCE ANALYSIS 47
Metabolic Pathways 48
Energy production and conversion 48
Carbohydrate metabolism 48
Amino acid and nucleotide metabolism 50
Metabolism of lipids and cell wall components 50
Metabolism of coenzymes 51
Translation System 51
Replication, Repair, and Recombination 51
Stress Response and Signal Transduction Systems 55
Distinctive Features of Predicted Operon Organization and Transcription Regulation 59
Expansion of Specific Protein Families 63
Proteins with Unusual Domain Architectures 66
Horizontal Gene Transfer 67
Mobile Genetic Elements 70
Inteins 70
Insertional sequences 70
Small noncoding repeats 71
Prophages 72
Evolutionary Relationships to Other Bacteria and Phylogeny 73
CONCLUSIONS 74
AVAILABILITY OF COMPLETE RESULTS 75
ACKNOWLEDGMENTS 75
ADDENDUM IN PROOF 75
REFERENCES 75

INTRODUCTION

Extreme Radiation Resistance

The evolution of organisms that are able to grow continuously at 6 kilorads (60 Gy)/h (119) or survive acute irradiation doses of 1,500 kilorads (50–52) is remarkable, given the apparent absence of highly radioactive habitats on Earth over geologic times. Notwithstanding a few natural fission reactors like those that gave rise to the Oklo uranium deposits (Gabon) 2 billion years ago (151), the radiation levels in the Earth's surface environments, including its waters containing dissolved radionuclides, have provided only about 0.05 to 20 rads/year over the last 4 billion years (193). DNA damage is readily inflicted on organisms by a variety of other common physicochemical agents (e.g., UV light or oxidizing agents) or nonstatic environments (e.g., cycles of desiccation and hydration or cycles of high and low temperatures) and it seems more likely that radiation resistance evolved in response to chronic exposure to nonradioactive forms of DNA damage.

Isolation

Bacteria belonging to the family *Deinococcaceae* are some of the most radiation-resistant organisms discovered, and they are vegetative, easily cultured, and nonpathogenic (23, 137, 138). Despite their ubiquitous distribution and apparent ancient derivation, only seven species of *Deinococcaceae* have been described (69, 138, 145). *Deinococcus radiodurans* strain

* Corresponding author. Mailing address: Department of Pathology, Room B3153, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814-4799. Phone: (301) 295-3750. Fax: (301) 295-1640. E-mail: mdaly@usuhs.mil.

R1 was the first of the deinobacteria to be discovered and was isolated in Oregon in 1956 (7) from canned meat that had spoiled following exposure to X rays. Culture yielded a red-pigmented, nonsporulating, gram-positive coccus that was extremely resistant to ionizing radiation, UV light, hydrogen peroxide, and numerous other agents that damage DNA (119, 137, 142, 215), as well as being highly resistant to desiccation (135). It is an aerobic, large (1- to 2-μm) tetrad-forming soil bacterium that is best known for its supreme resistance to ionizing radiation. It not only can survive acute exposures to gamma radiation that exceed 1,500 krads without dying or undergoing induced mutation (53), but it also displays luxuriant growth in the presence of high-level chronic irradiation (6 kilorads/h) (119, 212) without there being any effect on its growth rate or ability to express cloned foreign genes (31). For comparison, *Escherichia coli* will not grow and is killed in the presence of 6 kilorads/h (119) and an acute dose of only 100 to 200 kilorads needed to sterilize a culture. Similarly, vegetative cells of *Bacillus* spp. cannot grow at 6 kilorads/h and *Bacillus* spores show a 5-order-of-magnitude decrease in viability following acute exposure to 200 to 1,000 kilorads (207).

Shortly after the isolation of *D. radiodurans* R1 in 1956, a second strain of *D. radiodurans* (SARK) was discovered as an air contaminant in a hospital in Ontario (R. G. E. Murray and C. F. Robinow, Seventh International Congress for Microbiology, 1958). Since then, six closely related radioresistant species have been identified: *Deinococcus radiopugnans* from haddock tissue (54), *Deinococcus radiophilus* from Bombay duck (122), *Deinococcus proteolyticus* from the feces of *Lama glama* (108), the rod-shaped *Deinococcus grandis* from elephant feces (158), and the two thermophilic species *Deinococcus geothermalis* and *Deinococcus murrayi* from hot springs in Portugal and Italy, respectively (69). These species together form a distinct eubacterial phylogenetic lineage, believed to be most closely related to the *Thermus* genus. Based on 16S rDNA sequence analysis, it has been proposed that *Deinococcus* and *Thermus* form a eubacterial phylum (168). To date, the natural distribution of the deinococci has not been explored systematically. Isolations have occurred worldwide but are diverse and patchy in distribution. In addition to those noted above, sites of isolation include damp soil near a lake in England (133), weathered granite from the Antarctic Dry Valleys (44), irradiated medical instruments, and air purification systems (10, 41, 114, 145). As suggested above, it is possible that their extreme proficiency at DNA repair is related to the selective advantage in environments where they are prone to damage during long periods of desiccation (135). More recently, it has been proposed that adaptation could also occur in permafrost or other semifrozen conditions where cryptobiotic microbes with extremely long generation times could be selected with metabolic processes able to repair the unavoidable accumulation of background radiation-induced DNA damage (171).

Of the deinococcal species, *D. radiodurans* (138) and *D. geothermalis* (48) are the only ones for which a system of genetic transformation and manipulation has been developed. Now adding to this genetic technology is the recent complete sequencing and annotation of the *D. radiodurans* genome (218). The *D. radiodurans* strain R1 genome consists of two chromosomes (DR_Main [2.65 Mbp] and DR412 [412 kbp]), one megaplasmid (DR177 [177 kbp]), and one plasmid (46 kbp) (218), carrying 3,195 predicted genes. This combination of factors has positioned *D. radiodurans* as a promising candidate for the study of mechanisms of DNA damage and repair, as well as its exploitation for practical purposes such as cleanup and stabilization of radioactive waste sites. For example, *D. radiodurans* is being engineered to express metal-detoxifying and organic compound-degrading functions in environments heavily contaminated by radiation; 7×10^7 m^3 of ground and 3×10^9 liters of groundwater were contaminated by radioactive waste generated in the United States during the Cold War (31, 48, 119).

Cell Structure

The cell envelope of *D. radiodurans* is unusual in terms of its structure and composition (3). Although the cell envelope of *D. radiodurans* is reminiscent of the cell walls of gram-negative organisms (32, 61, 208, 221), *Deinococcus* often stains gram positive; this may result from the inability of its thick peptidoglycan layer to decolorize. Its cell envelope consists of the plasma and outer membranes, which are separated by a 14- to 20-nm peptidoglycan layer and an uncharacterized "compartmentalized layer." At least six layers have been identified by electron microscopy, with the innermost layer being the plasma membrane. The next layer is a peptidoglycan-containing cell wall and appears to be perforated (the holey layer), but it has no known physiological significance. The third layer appears to be divided into numerous fine compartments (the compartmentalized layer). The fourth layer is the outer membrane, and the fifth layer is a distinct electrolucent zone. The sixth layer consists of regularly packed hexagonal protein subunits (the S-layer, or hexagonally packed intermediate layer), typical of other bacterial S-layers (26, 115, 206). A few strains of *Deinococcus* also exhibit a dense carbohydrate coat (25, 26, 118, 187, 205, 208, 221). Only the cytoplasmic membrane and the peptidoglycan layer are involved in septum formation during cell division. The other layers are regarded as a sheath, since they surround groups of cells and form on the surface of daughter cells as they separate (187, 208, 221).

The chemical structure of the peptidoglycan layer of *D. radiodurans* SARK has been investigated using mass spectrometry (165), and the structure obtained is consistent with the A3β classification given to *D. radiodurans* (32, 176, 186). *Thermus thermophilus* HB8 (166) also has an A3β murein chemotype, and its peptidoglycan is built from the same monomeric subunit, underscoring the phylogenetic relationship between these genera.

The plasma and outer membranes appear to have the same lipid composition (206), yet there is no evidence for conventional lipopolysaccharides. The fatty acid composition of *D. radiodurans* is distinctive (69); attempts to identify hydroxy fatty acids, lipid A, and heptoses have been unsuccessful (145). A mixture of 15-, 16-, 17-, and 18-carbon saturated and monounsaturated acids are present, while polyunsaturated, cyclopropyl, and branched-chain fatty acids are not detectable. *D. radiodurans* has the distinguishing characteristic of lacking conventional phospholipids found in other bacteria (204). Of the *D. radiodurans* membrane lipid, 43% is composed of phosphoglycolipids containing a series of alkylamines as structural components, hitherto unknown as lipid constituents (8, 9). These

lipids appear to be derived from the same precursor, a novel phosphatidylglycerolalkylamine, and form when the precursor is glycosylated with galactose or glucosamine. Although glucosamine-containing lipids have been found in other species, notably members of the genus *Thermus* (160), these phosphoglycolipids are, at present, considered unique to *D. radiodurans*.

DNA Damage Resistance

The most extensively studied of the deinococci is *D. radiodurans*. Unlike other deinobacterial species, it is amenable to genetic manipulation due to its natural transformability by both high-molecular-weight chromosomal DNA and plasmid DNA (131, 143, 189). The natural transformability of *D. radiodurans* has facilitated the development of a variety of techniques for genetic manipulation of this organism (31, 49–52, 81–83, 119, 120, 131, 189–191), rendering it a highly susceptible target for molecular investigation. Transformability, however, is not integral to DNA damage resistance, since the other deinobacterial species are no less radioresistant than *D. radiodurans* (142) but are not transformable by any forms of DNA (*D. geothermalis*, however, is an exception since it has been transformed with plasmid recently [48]). In the exponential growth phase, *D. radiodurans* does not die in response to ionizing irradiation up to 0.5 megarad and shows 10% survival at 0.8 megarad (142), while exponentially growing *E. coli*, for comparison, shows a very small shoulder of complete resistance and 10% survival at 15 kilorads (188), a 50-fold difference in resistance (188). In the stationary growth phase, *D. radiodurans* does not die until exposed to 1.5 megarads, over 100-fold greater resistance than stationary-phase *E. coli* (53, 137). In exponential phase, *D. radiodurans* is 33-fold more resistant to UV than is *E. coli* (197). Compared to other organisms, the *D. radiodurans* DNA sustains the expected amount of damage in vivo at high irradiation doses, on the order of 150 to 200 double-stranded DNA breaks (DSBs) at 1.5 megarads per haploid chromosome under aerobic irradiation conditions, all of which are mended within hours following irradiation (53, 107, 123), nor is its DNA less susceptible than that of *E. coli* to UV in vivo (183). Furthermore, survivors of extreme ionizing radiation, UV, or bulky chemical-adduct exposures do not show any mutagenesis greater than that occurring after a single round of normal replication (197, 198). On the other hand, *D. radiodurans* is mutable by *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine and other agents that can cause mispairing of bases during replication (197, 198). Of the many forms of damage imposed on DNA by ionizing radiation, DSBs are considered the most lethal due to the inherent difficulty in their repair, since no single-strand template for accurate repair remains in the double helix (117). Other organisms, such as *E. coli*, can repair at most a few DSBs per chromosome without dying (112).

Logistics of Extreme DNA Damage Resistance

D. radiodurans contains 8 to 10 haploid genome copies during exponential growth and 4 genome copies during stationary phase (87, 89). In comparison, *E. coli* contains four or five haploid chromosomes during vigorous exponential growth, and this multiplicity in *E. coli* has been shown to be necessary for repair of DSBs (112). However, multiplicity in itself is insufficient for radioresistance. *Micrococcus luteus* and *Micrococcus sodonensis* also contain multiple genome equivalents but are radiosensitive (142). *Azotobacter vinelandii*, which contains up to 80 chromosomes per cell (164, 172), is quite sensitive to UV damage (125), to which *D. radiodurans* is highly resistant. Using various growth media, Harsojo et al. (89) were able to vary the genomic complement of *D. radiodurans* between 5 and 10 during the exponential phase and demonstrated that there was no correlation between chromosome number and radioresistance. The authors concluded that if chromosome multiplicity is important in repair, five or fewer chromosomes are sufficient. On high-level irradiation (1.75 megarads), *D. radiodurans* can reconstitute its genome from 1,000 to 2,000 DSB fragments compared to the maximum capability of *E. coli* of restoring its genome from 10 to 15 DSB fragments. Since most recombination models postulate that all DSB fragments search all others for homology during repair, this would call for an astronomical number of combinations to ensure genome restoration in *D. radiodurans*. Therefore, it may be that *D. radiodurans* can use redundant information in ways that other organisms do not. An alternative repair model has been postulated for *D. radiodurans* in which its chromosomes are always aggregated and aligned, thus dramatically simplifying the search for repair templates (51, 139) following DNA damage.

Repair of DNA damage in *D. radiodurans* follows an ordered series of events (137, 142). Physical repair of lesions requires conditions compatible with growth (212). For colony formation assays, this is simply achieved by plating on nutrient agar. For liquid cultures, this requires fresh nutrient medium and adjustment of cellular density to a level suitable for exponential growth. This has been demonstrated in liquid cultures for excision of pyrimidine dimers (30), repair of DSBs (56, 53), and recombinational repair of plasmids and chromosomes (51). While growth-promoting conditions are essential for removal of lesions from cellular DNA, the cells themselves do not immediately divide. Indeed, there is a dramatic inhibition of growth for extended durations following acute exposure to nonlethal (or partially lethal) DNA damage. This growth lag is associated with limited degradation of chromosomal DNA intrinsic to the DNA repair processes. Degradation proceeds at a rate independent of dose (the initial extent of damage), but its duration is positively correlated with dose (137; also see reference 142 and citations therein). Thus, the greater the dose, the longer the growth lag, which may exceed the duration of DNA degradation. Following a nonlethal exposure of stationary-phase *D. radiodurans* to 1.5 megarads under anoxic conditions, dilute liquid cultures of *D. radiodurans* show no growth for about 10 h and then resume rapid exponential growth (53). The dose-dependent delay of the onset of cellular replication suggests the existence of a checkpoint that monitors the extent of repair and accordingly controls the initiation of replicative DNA synthesis. During the period of stasis, it can be expected that the cell undergoes several phases of repair. The first can be termed cellular cleansing, and it involves several modalities, including the export of damaged DNA components. Initially, the products formed are DNA fragments about 2,000 bp long and consists of a mixture of damaged and undamaged nucleotides and nucleosides (22, 213). These products are found in the cytoplasm and also in the surrounding growth medium, suggesting that *D. radiodurans*

exports the DNA degradation products once they are formed (reference 22 and citations therein). The removal of damaged nucleotides outside the cell might protect the organism from elevated levels of mutagenesis by preventing the reincorporation of damaged bases during DNA synthesis (22). Remaining intracellular mutagenic precursors could be sanitized via pyrophosphohydrolases of the Nudix superfamily (for "nucleoside diphosphate linked to some other moiety x"), the founding member of which is the repair enzyme MutT (28). MutT has an 8-oxo-dGTPase activity, which produces 8-oxo-dGMP plus inorganic pyrophosphate. Since 8-oxo-dGTP is highly mutagenic, the enzyme "sanitizes" the nucleoside triphosphate pool. *D. radiodurans* is markedly rich in Nudix proteins, some of which may act to sanitize other mutagenic DNA precursors (218). Finally, activated oxygen species with long half-lives may be eliminated by superoxide dismutases and catalases such as SodA and KatA (129). During this initial phase of cellular cleansing, amino acids, nucleotides, nucleosides, sugars, and phosphate may be imported into the cell while precursors for DNA synthesis are made by way of ribonucleoside diphosphate reductase (104). Subsequent phases of repair are genomic restoration and coordination of repair activities.

DNA Repair Pathways

D. radiodurans has repair pathways that include excision repair, mismatch repair, and recombinational repair. Generally, no marked error-prone SOS response is observed in *D. radiodurans* (142). However, there have been a few reports consistent with SOS response, where preexposure to low doses of ionizing radiation, UV, or hydrogen peroxide causes a low level of subsequent increased resistance to DNA damage (twofold or less) (199, 215). Since the SOS response is not always mutagenic, the absence of DNA damage-induced mutagenesis observed in *D. radiodurans* cannot be taken as evidence against the existence of the SOS response in this bacterium. Photoreactivation is not present (142), and it has been reported that the adaptive response to alkylation damage is also absent (170). It is known that following DNA damage, there are changes in the cellular abundance of proteins, with enhanced synthesis of four to nine proteins, as judged by sodium dodecyl sulfate-polyacrylamide protein gels (86, 200). Included in this group of proteins are probably RecA (36), elongation factor Tu (200), and KatA (129). While there are many predicted DNA repair genes and pathways in the *D. radiodurans* genome (218), only a few of its DNA repair enzymatic activities and/or genes have been evaluated for their biochemical activities. The UvrA protein and its gene have been detected (1, 149), and it has been identified as a component of nucleotide excision repair. UV endonuclease-beta has been purified and found to be a 36-kDa manganese-requiring protein, which is thus far only known to recognize UV-induced pyrimidine cyclobutane dimers, incising them as an endonuclease rather than as a glycosylase (63–65). Other repair-related activities detected in extracts of *D. radiodurans* include uracil DNA glycosylase (132), a thymine glycol glycosylase, and a deoxyribophosphodiesterase (144). DNA polymerase I activity is present and is necessary for resistance to both UV and ionizing radiation (81). Both UvrA and DNA polymerase I deficiencies can be fully complemented by the expression of *E. coli* UvrA and DNA polymerase I proteins in *D. radiodurans* mutants, respectively (1, 81). However, this is not the case for *D. radiodurans recA*, which appears to play a more important role in the extreme radiation resistance phenotype.

The *D. radiodurans* RecA protein has been detected and its gene has been sequenced; it shows greater than 50% identity to the *E. coli* RecA protein (81). Mutants with mutations in this gene are highly sensitive to UV and ionizing radiation. Unlike UvrA and DNA polymerase I proteins, expression of *E. coli* RecA in *D. radiodurans* does not complement the RecA deficiency and appears to have no effect on *D. radiodurans* (82, 36). Expression of *D. radiodurans* RecA in *E. coli* has been reported to be lethal (36); however, recently it has been successfully expressed in *E. coli* with less toxicity (M. M. Cox and K. W. Minton, unpublished data), and it has been reported to complement *E. coli* RecA deficiency (150).

D. radiodurans RecA has recently been purified and characterized (M. M. Cox, unpublished data). In vitro, it has been shown to catalyze the spectrum of activities classically attributed to RecA proteins: (i) it forms striated filaments on single-stranded DNA and double-stranded DNA; (ii) it promotes an efficient DNA strand exchange reaction; and (iii) it has a DNA-dependent nucleoside triphosphatase activity. However, *D. radiodurans* RecA is distinct from other well-characterized RecAs (e.g., from the gram-negative *E. coli*) in its nucleoside triphosphatase and DNA strand exchange activities. Unlike *E. coli* RecA, *D. radiodurans* RecA does not hydrolyze ATP at pH 7.5, although it exhibits some ATPase activity at lower pHs. In contrast, it is very effective at hydrolyzing dATP over a broad pH range.

The existence of a very efficient *recA*-independent single-stranded DNA annealing repair pathway has been reported for *D. radiodurans* (50). This pathway is active during and immediately after DNA damage and before the onset of *recA*-dependent repair. It can repair about one-third of the 150 to 200 DSBs per chromosome following exposure to 1.75 megarads (50). It has also been reported that unlike other organisms, *D. radiodurans* RecA is not present in the undamaged deinococcal cell but is synthesized only following DNA damage and following repair. *D. radiodurans* RecA is apparently expressed in *D. radiodurans* only following extreme DNA damage (36), and it is noteworthy that the *recA*-defective *D. radiodurans* strain rec30 is more radiation resistant than *E. coli* (138). It is possible that the greater resistance of rec30 arises from the presence of multiple copies of its genome in combination with the single-stranded DNA-annealing repair pathway, which is fully functional in this mutant (50). Together, this evidence supports the idea that *D. radiodurans* RecA is not necessary for the repair of nonextreme DNA damage (~10 DSB/chromosome, ~100 kilorads) and that Dr RecA may be activated only when DNA is highly damaged (>100 kilorads) (M. J. Daly, unpublished data).

SEQUENCE ANALYSIS

To further our understanding of the functions of individual genes and cellular systems in *D. radiodurans* as well as their relationship with other organisms, we undertook a detailed computational analysis of the *D. radiodurans* genome. In addition to the standard genome annotation procedure of The

Institute for Genomic Research (218), we used several approaches for deeper protein characterization. In particular, we systematically applied sensitive profile-based methods that included PSI-BLAST, which constructs a position-dependent weight matrix from multiple alignments generated from the BLAST hits above a certain expectation value (e-value) and allows iterative database searches using the information derived from such a matrix (5, 6), IMPALA (175), which searches the matrix against profile databases, and SMART (179, 178), which uses a Hidden Markov Model algorithm (59) to search a sequence against a multiple-alignment database. In addition to the database of profiles included in the SMART system, two other profile collections were used: (i) 5,640 profiles derived from the structurally characterized domains contained in the SCOP database (100, 219), and (ii) 150 profiles for widespread domains primarily involved in different forms of signaling that were employed in previous genome comparisons (40, 163, 175).

Paralogous families of proteins encoded in the *D. radiodurans* genome were initially identified by comparing the complete set of *D. radiodurans* proteins to itself (after filtering for low-complexity regions with the SEG program [220]) using the PSI-BLAST program run for three iterations and clustering proteins by single linkage (clustering threshold e-value, 0.001) using the GROUPER program (214). One sequence from each cluster was used to generate a position-specific matrix by running an iterative PSI-BLAST search first against a *D. radiodurans* protein and then against the nonredundant protein database. These profiles were used to search for additional family members in the *D. radiodurans* proteome. Families that were recognized by the same profile were joined into superfamilies.

The phylogenetic affinities of *D. radiodurans* were explored using the COGNITOR program. This program assigns query proteins to conserved protein families that consist of apparent orthologs, termed clusters of orthologous groups (COGs) (201, 202). The functional assignments embedded in the COG database were also used to reconstruct metabolic pathways and other functional systems in *D. radiodurans* together with the KEGG (105) and WIT (157) databases.

Analysis of the phyletic distribution of homologs of *Deinococcus* proteins detected in database searches was performed using the TAX_COLLECTOR program of the SEALS package (214). This was followed by phylogenetic tree construction for specific cases. Multiple alignments for phylogenetic reconstruction were generated using the ClustalW program (93) and, when necessary, further adjusted on the basis of PSI-BLAST search outputs. Phylogenetic trees were constructed using the neighbor-joining methods with bootstrap replications as implemented in the NEIGHBOR program of the PHYLIP package (67).

Intergenic repeats were identified using the BLASTN program (6). As a result of this analysis, 2,007 *D. radiodurans* proteins were assigned to 1,272 COGs, which placed them into specific phylogenetic and functional contexts. In conjunction with profile analysis, this allowed us to define the domain architectures of multidomain proteins, to identify protein families that are unusually expanded in *D. radiodurans*, and to assign function and/or structure to a number of proteins previously described as hypothetical.

Below, we present an overview of the principal functional systems of *D. radiodurans* as determined by these analyses and describe unusual aspects of the genome that may be relevant to understanding the extreme resistance of this organism to radiation, desiccation, and other stress factors.

Metabolic Pathways

Analysis of the genome of *D. radiodurans* shows that it has a typical set of proteins for housekeeping and regulatory functions. As demonstrated by the COG analysis, the metabolic capabilities of *D. radiodurans* are similar to those of *E. coli* (152) but less diverse (Table 1); *D. radiodurans* is an obligatory heterotroph (212). Table 1 lists and compares the standard metabolic pathways of *Deinococcus* to the corresponding pathways in *E. coli, Synechocystis, Bacillus subtilis*, and *Mycobacterium tuberculosis*.

Energy production and conversion. Probably the most interesting feature of the systems for energy production in *D. radiodurans* is that, unlike most other free-living bacteria, it uses the vacuolar type of proton ATP synthase instead of the F_1F_0 type. Vacuolar (V)-type H^+-ATPase is typical of eukaryotes and archaea; all archaea have a conserved operon that consists of eight genes encoding the ATPase subunits. This operon is partially conserved (with some of the subunits missing) in a minority of characterized bacteria, where it replaces the F_1F_0 ATPase, e.g., in *Deinococcus, Thermus*, spirochetes, chlamydiae, and *Enterococcus*. The scattered distribution of the V-ATPase operon among bacteria, in contrast to its conservation in archaea, suggests that this operon has been disseminated in the bacterial world by horizontal transfer. The genes for the standard five complexes of electron transport and oxidative phosphorylation are present in *D. radiodurans*, with a few exceptions, but some genes of the cytochrome *bd* quinol oxidase complex are missing. Given that this complex is active predominantly under low-oxygen conditions in other bacteria, its apparent loss in *Deinococcus* is consistent with *D. radiodurans* being strictly aerobic. Interestingly, *D. radiodurans* encodes a multisubunit Na^+/H^+ antiporter (DR0880 to DR0886) that is characteristic of thermophiles and a few other bacteria (*B. subtilis* and *Rickettsia prowazekii*), but is absent in *E. coli, Synechosystis*, and *Mycobacterium*. It has been shown that this system is necessary for cells to grow under alkaline conditions (95).

Carbohydrate metabolism. The *D. radiodurans* genome appears to encode functional pathways for glycolysis, gluconeogenesis, the pentose phosphate shunt, and the tricarboxylic acid (TCA) cycle. A few genes are missing, but these may not be essential since they are also absent in some bacteria that are functional in these pathways (Table 1). The *D. radiodurans* Entner-Doudoroff pathway may be disrupted since a key enzyme, 2-keto-3-deoxy-6-phosphogluconate aldolase (an ortholog of *E. coli* Eda), is missing. However, this enzyme is also absent in archaea, where the Entner-Doudoroff pathway appears to be functional, and therefore the enzyme could be displaced by a nonorthologous aldolase in *Deinococcus*. The glyoxalate bypass that has only been described for *E. coli* and *M. tuberculosis* is present and complete in *Deinococcus*. It remains unclear, however, why some intermediates of the TCA cycle cannot support the growth of *D. radiodurans* (212). As

TABLE 1. Basic metabolic pathways in *D. radiodurans*

Pathway	Genes in the pathway[a]	Genes missing[b]	Comments
Glycolysis	*glk, pgi, pfkA, fba/dhnA, tpi, gapA, pgk, pgm/yibO, eno, pykA*		Complete pathway
Gluconeogenesis	*ppsA, eno, pgm, pgk, gapA, tpi, fba/dhnA, fbp, pgi*	*fbp* (CE--)	Likely functional pathway
Pentose phosphate shunt and pentose biosynthesis	*zwf, gnd, tktA, talA, yhfD, rpiA, deoC*		Complete pathway
Entner-Doudoroff pathway	*zwf, edd, eda, gnd*	*eda* (CEB-)	Likely functional pathway
TCA cycle	*gltA, acnA, icd, sucA, sucB, sucC, sucD, frdA, frdB, fumA, fumC, mdh*	*fumA* (-E--)	Likely functional pathway
Glyoxalate bypass	*glcB, AceA*		Rare pathway present in *E. coli, M. tuberculosis*, and few other bacteria
Purine biosynthesis	*prsA, purF, purD, purN/purT, purL, purM, purK, purE, purC, purB, purH2, purH1, purA, guaB, guaA*		Complete pathway
Purine salvage	*purU, deoD, xapA, apt, xpt, hpt*	*xapA* (CEBR)	*D. radiodurans* has two *apt* genes of archaeal type
Pyrimidine biosynthesis	*curA, carB, pyrB, pyrC/ygeZ, pyrD, pyrE, pyrF, pyrH, ndk, pyrG*		Complete pathway
Pyrimidine salvage	*cdd, upp, udk, deoD, deoA, nrdF, nrdE, pfs/amn, tdk*		Complete pathway with two *nrdE* (one is of archaeal type with intein)
Thimidylate biosynthesis	*dcd, dut, thyA, tmk, ndk*	*dcd* (CE-R), *dut* (CEBR)	Pathway could be functional if unknown analogs of *dcd* and *dut* are present
Histidine biosynthesis	*prsA, hisG, hisI2, hisI, hisA, hisH, hisF, hisB2, hisC, hisB1, hisD*		Complete pathway
Branched-chain amino acid biosynthesis	*ilvA, ilvB, ilvN, ilvC, ilvD, leuA, leuC, leuD, leuB, ilvE*		Complete pathway
Glutamate and glutamine biosynthesis	*gltB, gdhA, glnA*		Complete pathway; *D. radiodurans* has two *glnA* genes, One is for the rare class III glutamine synthase; in R1 strain this gene has a frameshift
Aspartate and asparagine biosynthesis	*aspC, asnB, asnA, ansA*	*asnB* (-EBR), *asnA* (-E--)	Pathway could be functional if unknown analogs of *asnB* and *asnA* are present
Aromatic amino acid biosynthesis	*aroG/kdsA, aroB, aroD, aroE, aroK, aroA, aroC, pheA1, pheA2, tyrA2, tyrB, trpD1, trpE, trpD2, trpC2, trpC1, trpA, trpB*	*tyrB* (-E--)	*D. radiodurans* has both *aroG* and *kdsA*; *D. radiodurans* and *B. subtilis* have rare bifunctional protein: chorismate mutase (*tyrA*1) and 2-dehydro-3-deoxyphosphoheptonate aldolase (*aroG*); *D. radiodurans* has two *trpE* genes, one of which is fused to *trpG*; same fusion is also found in *Asospirillum* and *Rhizobium*; reverse fusion is in *Streptomyces*
Serine and glycine metabolism	*serA, serC, serB, glyA, gcvP, gsvT, gsvH, lpd*	*serC* (-EB-), *serB* (-E-R)	Pathway could be functional if unknown analogs of *serB* and *serC* are present
Threonine biosynthesis	*thrA, asd, thrB, thrC*		Complete pathway
Methionine biosynthesis	*metL1/thrA1, asd, metL2/thrA2, metA, metB, metC, metE/metH*	*metA* (-EB-)	Incomplete and unlikely to be a functional pathway
Cysteine biosynthesis	*cysD/cysH, cysC, cysN, cysI, cysJ, cysK/cysM, cysE*	*cysD/cysN, cysM, cysE* (CEBR), *cysJ* (CEB-)	Unlikely to be a functional pathway
Arginine biosynthesis	*argJ, argB, argC, argD, argE, argF, argG, argH, argI*		Likely functional pathway; circular type as in gram-positive bacteria; some genes were acquired from archaea (see Table 11)
Proline metabolism	*argB, argE, proB, proA, proC, putA*		Complete pathway
Lysine biosynthesis	*dapA, dapB, dapD, dapC, dapE, dapF, lysA*	*dapA, dapB, dapF* (CEBR), *dapD* (-EB-)	Unlikely to be a functional pathway; *dapC* may be substituted by other aminotransferase; the closest gene to *dapE* is more likely to be an ortholog of *B. subtilis rocB* and therefore is probably involved in degradation of amino acids rather than in lysine biosynthesis
Fatty acid biosynthesis	*accB, accC, accA, accD, acpP, fabB/fabF, fabH, fabD, fabG, fabI, fadA, BS_mmgB, caiD*		Complete pathway; *D. radiodurans* encodes four *accA*, four *accD*, four BS_*mmgB*, and five *caiD*
NAD biosynthesis	*nadB, nadA, nadC, nadD, nadE, pncB*	*nadB, nadA, nadC, nadD* (CEBR)	Unlikely to be a functional pathway
Riboflavin and FAD biosynthesis	*ribA, ribD, ribB, ribE, ribC, ribG*		Complete pathway

Continued on following page

TABLE 1—*Continued*

Pathway	Genes in the pathway[a]	Genes missing[b]	Comments
Siroheme biosynthesis	*hemA, hemL, hemB, hemC, hemD, cysG2, cysG1*		*D. radiodurans* has two other genes related to this pathway; *hemF* and *hemY*
Cobalamin biosynthesis	*cysG2, cbiL, cbiH, cbiF, cbiJ, cbiE, cbiT, cbiC, cbiA, cobN, cobA, cbiP, cobD, cbiB, cobT, cobS, cobU*	*cbiL, cbiH, cbiJ, cbiE, cbiT, cbiC, cobN* (C--R)	Possible partly functional pathway
Biotin biosynthesis	*bioW, bioF, bioA, bioD, bioB, birA, bioH*	*bioW* (--B-), *bioA, bioD, bioB* (CEBR)	Pathway could be functional if unknown analogs of *bioD* and *bioW* are present; *bioA* aminotransferase can be substituted by paralogous enzyme, and any biotin synthase-related enzyme may replace *bioB*
Pyridoxal phosphate biosynthesis	*yaeM, ldh, serC, pdxA, pdxJ, BS_yaad, pdxH, pdxK*	*pdxA, pdxJ*, (CE--)	*D. radiodurans* has an ortholog of BS_*yaad* which is found so far only in archaea and eukaryotes
Thiamine biosynthesis	*thiC, thiD, thiK, thiE, thiL*	*thiK* (-EB-), *thiL* (CEBR)	Pathway could be functional if unknown analogs of *thiK* and *thiL* are present
Ubiquinone and menaquinone biosynthesis	*menF, menD, menC, menE, menB, menA, menG, ubiA, ubiX, ubiB, ubiH, ubiE, ubiG*	*menF, menD, menC, menE, menB, menA*, (CEBR), *menG* (CE-R)	Unlikely functional pathway of menaquinone biosynthesis; there are some paralogs of *menC*, but they are unlikely to be related to this pathway; synthesis of ubiquinone is likely to be present; only *ubiG* is missed, but it exists only in *E. coli, Rickettsia* and yeast
NAHD-ubiquinone oxidoreducatase	All 14 subunits in one operon		Complete pathway
H^+-ATPase	8 subunits in one operon		Complete pathway; vacuolar-type H^+-ATPase like in archaea, *Thermus*, spirochetes, and *Chlamydia*
Cytochrome *c* and *b*-dependent electron transport	*cccA/cccB, qcrB, ctaA, ctaE, ctaF, ctaD, ctaB, ctaC, ccdA, sdhC, ccmG, ccmF, ccmE, ccmD, ccmC, ccmB, ccmA, ccmH, cydB, cydA*	*ctaF, ccmA, ccmD* (-E--), *cydA, cydB* (CEBR)	Probably functional pathway; component of heme exporter (such proteins are definitely present and some of them can perform this function)

[a] The gene names and pathway classification follow the biochemical data and nomenclature described for *E. coli* and *S. enterica* serovar *Typhimurium* (152).
[b] The presence or absence in bacteria with large genomes is indicated in parentheses after the names of genes that are missing in *D. radiodurans*. Abbreviations are as follows: C, *Synechocystis* sp.; E, *E. coli*; B, *B. subtilis*; R, *M. tuberculosis*.

expected of a heterotroph, *Deinococcus* encodes several enzymes for complex carbohydrate metabolism; for some of these, e.g., glycogen-debranching enzymes (DR0405 and DR0191), phylogenetic analysis suggests that horizontal transfer from eukaryotes has occurred (data not shown). Other enzymes for sugar conversion, as well as most of the known sugar transport systems, are encoded in *D. radiodurans*, and this is consistent with the observation that a variety of different sugars can be used by this bacterium as carbon and energy sources (212).

Amino acid and nucleotide metabolism. *D. radiodurans* is unable to use ammonia as a nitrogen source despite the presence of apparently functional genes for glutamate ammonia ligase and carbamoyl-phosphate synthase, which are key enzymes for ammonia utilization. While there is currently no explanation for this, it has been shown that *D. radiodurans* can use amino acids effectively as a nitrogen source and that sulfur-containing amino acids appear to be the most readily utilized form of nitrogen. Notably, *D. radiodurans* lacks the standard pathways for cysteine and methionine biosynthesis yet is able to produce these amino acids using unidentified biosynthetic pathways when provided with other amino acids (212). The absence of all key enzymes for lysine biosynthesis is another puzzling feature of *Deinococcus* metabolism since it does not require lysine for growth (212). All of the other standard amino acid pathways appear to be functional. Although a few genes seem to be missing from these pathways, they are also absent in some of the other free-living bacteria, where they probably have been displaced by paralogous or nonhomologous enzymes. Some of the genes for enzymes of arginine metabolism are likely to have been acquired by the common ancestor of the *Thermus-Deinococcus* group from archaea (see Tables 10 and 11).

Most of the known genes for nucleotide metabolism are present in *D. radiodurans*. The most conspicuous gap is the absence of purine nucleoside phosphorylase, a key enzyme of purine salvage, which has been found in all free-living organisms investigated. Another noteworthy absence is that of two related enzymes of pyrimidine salvage, cytidine deaminase and dUTPase (important in preventing DNA damage), which are present in most bacteria. As may be the case for absent amino acid biosynthetic genes, there might also be unidentified enzymes that compensate for these pyrimidine salvage activities.

Metabolism of lipids and cell wall components. *D. radiodurans* lacks only one gene from the standard bacterial set of genes coding for enzymes of lipid metabolism, namely, phosphatidylglycerophosphate synthase, which is involved in the biosynthesis of acidic phospholipids. With the exception of the archaeon *Methanococcus jannaschii*, phosphatidylglycerophosphate synthase has been detected in all organisms with completely sequenced genomes. Its absence in *Deinococcus*, therefore, is unexpected. *Deinococcus* encodes multiple copies of several fatty acid biosynthesis genes, of which some could have been transferred horizontally into *Deinococcus* from distant taxa (Table 1). Consistent with the unusual structure of the peptidoglycan layer in *Deinococcus* (see above), we identified

all essential genes for ornithine metabolism but did not detect several key enzymes for diaminopimelic acid biosynthesis.

Metabolism of coenzymes. Our experimental data show that *Deinococcus* is capable of de novo biosynthesis of all principal coenzyme components except for nicotinic acid (212). Consistent with this result, we find that genes for several key enzymes of NAD biosynthesis are missing in the genome, which is unusual since this pathway is present in most free-living organisms. Several other conventional pathways for coenzyme biosynthesis are also not complete (Table 1), but, given the ability of *Deinococcus* to grow in the absence of these coenzymes, it probably encodes functional analogs of these.

Translation System

The translation apparatus is arguably the most highly conserved and uniform of cellular systems, and *D. radiodurans* is no exception. It contains a typical bacterial complement of translation machinery components. This general uniformity notwithstanding, there are several unique features in the translation apparatus of *Deinococcus* that have been revealed both experimentally and by genome analysis. In particular, *Deinococcus* has a unique repertoire of genes and reactions for the formation of glutaminyl-tRNA and asparaginyl-tRNA. Generally, there are two pathways for the activation of glutamine and asparagine: (i) direct charging of $tRNA^{Gln}$ and $tRNA^{Asn}$ by glutaminyl- and asparaginyl-tRNA synthetase (Gln-RS and Asn-RS), respectively, and (ii) transamidation of Glu-$tRNA^{Gln}$ and Asp-$tRNA^{Asn}$ by the respective amidotransferases (AdT), Glu-AdT and Asp-AdT (101). Usually, the two pathways and the corresponding genes are not present in the same organism. The transamidation pathway for glutamine is predominant in bacteria and archaea, whereas glutaminyl-tRNA synthetase is typical of eukaryotes and gamma proteobacteria (101). In the case of asparagine, archaea primarily use the transamidation pathway, eukaryotes use the direct pathway, and bacteria have a patchy distribution of both systems. Glu-AdT has been studied in detail; it consists of three subunits encoded by the *gatABC* genes (45). The nature of Asp-AdT is less clear; it has been suggested that it shares A and C subunits with Glu-AdT whereas the B subunit (the likely determinant of tRNA binding) is unique. *D. radiodurans* encodes Asn-RS, Gln-RS, and the GatABC proteins (45). A recent genome survey has shown that the two systems also coexist in several members of the proteobacteria (85), but *Deinococcus* is the only nonproteobacterial species with this combination of asparagine and glutamine activation systems. Furthermore, in addition to the intact GatB, *Deinococcus* encodes a C-terminal domain of this protein that is fused to Gln-RS (Fig. 1). The GatABC complex of *D. radiodurans* is capable of catalyzing the formation of both Gln-$tRNA^{Gln}$ and Asn-$tRNA^{Asn}$, but in vivo apparently only Asn-$tRNA^{Asn}$ is formed, since the discriminating Glu-RS of *Deinococcus* does not produce the mischarged Glu-$tRNA^{Gln}$ (45). In contrast, *Deinococcus* encodes two copies of Asp-RS, a typical bacterial discriminating copy and nondiscriminating copy that probably was acquired from the archaea by horizontal gene transfer (45) (see below). The nondiscriminating Asp-RS produces Asp-$tRNA^{Asn}$, which serves as the substrate for the GatABC enzyme. It has been suggested that the main role of the Asn-$tRNA^{Asn}$ formation in *Deinococcus* is the synthesis of asparagine, rather than its incorporation into proteins, since *Deinococcus* does not encode orthologs of known asparagine synthetases (45). Given that GatB is thought to be the tRNA-binding component of Glu-AdT and Asp-AdT, the C-terminal GatB-related domain in *Deinococcus* Gln-RS could enhance the specificity of this enzyme for $tRNA^{Gln}$. This domain is missing in other Gln-RSs, but the respective organisms do not encode GatB, which in *Deinococcus* could compete with Gln-RS for binding $tRNA^{Gln}$.

The repertoire of aminoacyl-tRNA synthetases (aminoacyl-RSs) in *Deinococcus* also shows several other peculiarities. In addition to the corresponding functional enzymes, *Deinococcus* encodes truncated and apparently inactive forms of Glu-RS and Ala-RS, as well as apparently active paralogs of Trp-RS and His-RS. Possible horizontal transfer of these additional enzymes as well as other aminoacyl-RSs from archaea and thermophilic bacteria could be readily examined once more of these organisms are sequenced.

Replication, Repair, and Recombination

D. radiodurans contains all the typical bacterial genes that comprise the basal DNA replication machinery (Table 2). The number of paralogs and the domain organization of the DNA polymerase III α-subunit is variable in the major bacterial divisions in terms of the presence of an active or inactivated PHP domain, which is predicted to possess phosphatase activity, and the proofreading 3′-5′ exonuclease domain. *D. radiodurans* encodes a single α-subunit that is most similar to proteobacterial polymerases and does not contain the 3′-5′ exonuclease, which is encoded by a separate gene orthologous to *E. coli dnaQ*. Unlike the proteobacterial orthologs, however, the *Deinococcus* polymerase contains an apparently active PHP domain. This appears to represent the ancestral bacterial state of the replicative DNA polymerase, which is also seen in bacteria like *Synechocystis* and *Aquifex*. In addition to typical proteins involved in replication, *Deinococcus* encodes DNA polymerase X, which is similar to the eukaryotic DNA polymerase beta (references 27 and 217 and references therein), and is relatively uncommon in prokaryotes. *Deinococcus* polymerase X contains an N-terminal nucleotidyltransferase domain and a C-terminal PHP hydrolase domain, the same domain architecture that is seen in homologs from *B. subtilis* and *Methanobacterium thermoautotrophicum*; this conservation of domain organization suggests horizontal transfer of the polymerase X gene (13). Notably, along with a few other bacteria, such as *Synechocystis* and *Aquifex, Deinococcus* encodes three small nucleotidyltransferases (DR1806, DR0679, and DR0248), which are expanded in archaea (13). These "minimal" nucleotidyltransferases are typically accompanied by a small protein that is fused to the nucleotidyltransferase in the DR0248 protein; the function of this protein, however, has not been characterized directly but is likely to be coupled to that of the nucleotidyltransferases.

The repertoire of DNA-associated proteins in *Deinococcus* is similar to that in other bacteria, but some unique features were noticed. Like other bacteria, *Deinococcus* encodes an ortholog of the chromosomal DNA-binding protein HU, which is believed to play a central role in DNA packaging and also as a cofactor in recombination (reference 184 and references

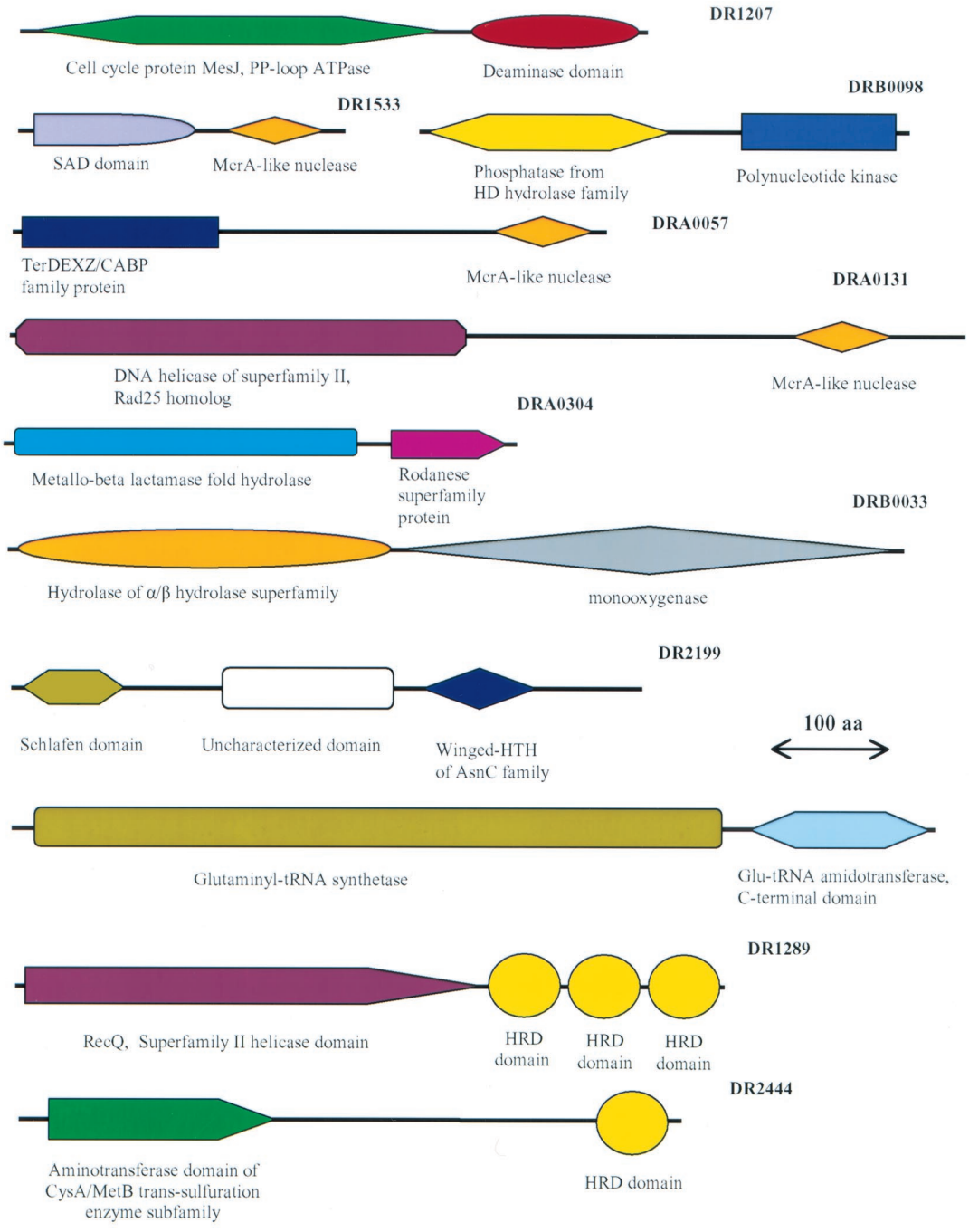

FIG. 1. Examples of unique domain architectures of *Deinococcus* proteins.

therein). Interestingly, the sequenced genome of the *Deinococcus* R1 strain contains three adjacent open reading frames (ORFs) encoding fragments of the single-stranded DNA-binding protein (SSB) but lacks a complete gene for SSB; so far, all sequenced bacterial genomes encoded an intact SSB. Because of the 10-fold coverage during the TIGR sequencing project (218), two sequencing errors in this short gene would seem unlikely. Two explanations arise: (i) *Deinococcus* could encode an as yet unrecognized SSB analog (or an extremely diverged homolog), making the SSB gene expendable; or (ii) a tripartite SSB gene could be expressed by a translational readthrough mechanism or even a unique RNA-editing mechanism.

Bacterial DNA repair includes several partially redundant pathways and generally shows considerable flexibility (20, 60, 70). We investigated the predicted repair system components of *D. radiodurans* in detail, to detect any possible correlation with its exceptional radioresistant and desiccation-resistant phenotype. Generally, it appears that *Deinococcus* possesses a typical bacterial system for DNA repair and that, commensurate with the genome size, its repair pathways even appear to

TABLE 2. Genes coding for replication, repair, and recombination functions in *D. radiodurans*[a]

Gene name[b]	Gene_ID	Protein description and comments	Pathway[c]	Phylogenetic pattern[d]
yhdJ	DRC0020	Adenine-specific DNA methylase	mMM?	-m-k--vd-e--huj-------
ogt/ybaZ	DR0248	*O*-6-methylguanine DNA methyltransferase	DR	amtkyqvd-ebrhuj---lin-
mutT	DR0261	8-oxo-dGTPase; *D. radiodurans* encodes another 22 paralogs; only some predicted to function in repair	DR	--t----d-ebrhuj---lin-
alkA	DR2074, DR2584	3-methyladenine DNA glycosylase II; DR2584 is of eukaryotic type	DR, BER	-------d--br-----o--nxa -tky--dcebr----------
mutY	DR2285	8-oxoguanine DNA glycosylase and AP-lyase, A-G mismatch DNA glycosylase	BER, MMY	--t----d-ebrhuj---lin-
nth	DR2438, DR0289, DR0928	Endonuclease III and thymine glycol DNA glycosylase; DR0928 and DR2438 are of archaeal type, and DR0289 is close to yeast protein	BER	amtkyqvdcebrhuj--olinx
mutM/fpg	DR0493	Formamidopyrimidine and 8-oxoguanine DNA glycosylase	BER	-------dcebrh--gp-----
nfi (*yjaF*)	DR2162	Endonuclease V	BER	a--k-qvd-eb----------
polA	DR1707	DNA polymerase I	BER	--t--qvdcebrhujgpolinx
ung	DR0689, DR1663	Uracil DNA glycosylase; DR0689 is a likely horizontal transfer from a eukaryote or a eukaryotic virus	BER	----y--d-ebrhujgpo-inx
mug	DR0715	G/T mismatch-specific thymine DNA glycosylase, distantly related to DR1751; present as a domain of many multidomain proteins in many eukaryotes	BER	-------d-e------------
	DR1751, DR0022	Uracil DNA glycosylase	BER	a--k-qvdc-br-----ol--x
xthA	DR0354	Exodeoxyribonuclease III	BER	a-t-y--dcebrhuj--ol--x
sms	DR1105	Predicted ATP-dependent protease	NER, BER	-----qvdcebrhuj---linx
mfd	DR1532	Transcription repair coupling factor; helicase	NER	------vdcebrhuj--olinx
uvrA	DR1771, DRA0188	ATPase, DNA binding	NER	--t--qvdcebrhujgpolinx
uvrB	DR2275	Helicase	NER	--t--qvdcebrhujgpolinx
uvrC	DR1354	Nuclease	NER	--t--qvdcebrhujgpolinx
uvrD	DR1775, DR1572	Helicase II; initiates unwinding from a nick; DR1572 has a frameshift	NER, mMM, SOS	--t-yqvdcebrhujgpolinx
mutL	DR1696	Predicted ATPase	mMM, VSP	----yqvdceb-h----olinx- -tk-qvdc-b--uj--o----
mutS	DR1976, DR1039	ATPase; DR1039 has a frameshift	mMM, VSP	----yqvdceb-h----olinx
xseA/nec7	DR0186	Exonuclease VII, large subunit	MM	------vd-ebrhuj----inx
sbcC	DR1922	Exonuclease subunit, predicted ATPase	RER	amtkyqvdceb------ol---
sbcD	DR1921	Exonuclease	RER	amtkyqvdcebr-----ol---
recA	DR2340	Recombinase; single-stranded DNA-dependent ATPase, activator of *lexA* autoproteolysis	RER, SOS	amtkyqvdcebrhujgpolinx
recD	DR1902	Helicase/exonuclease; contains three additional N-terminal helix-hairpin-helix DNA-binding modules; closely related to *RecD* from *B. subtilis* and *Chlamydia*	RER	-m--y--d-ebrh----o-in-
recF	DR1089	Predicted ATPase; required for daughter strand gap repair	RER	-------dcebrh-----li-x
recG	DR1916	Holliday junction-specific DNA helicase; branch migration inducer	RER	-----qvdcebrhuj--ol--x
recJ	DR1126	Nuclease	RER	amtk-qvdceb-huj--olinx
recN	DR1477	Predicted ATPase	RER	-----q-dcebrhuj---l--x
recO	DR0819	Required for daughter strand gap repair	RER	-------dcebrh-----lin-
recQ	DR2444, DR1289	Helicase; suppressor of illegitimate recombination	RER	----y--dceb-h-----l---
recR	DR0198	Required for daughter-strand gap repair	RER	-----q-dcebrhuj---linx
ruvA	DR1274	Holliday junction-binding subunit of the RuvABC resolvasome	RER	--t--qvdcebrhujgpolinx
ruvB	DR0596	Helicase subunit of the RuvABC resolvasome	RER	------vdceb-hujgpolinx
ruvC	DR0440	Endonuclease subunit of the RuvABC resolvasome	RER	------vdce-rhuj---linx
dnaE	DR0507	Polymerase subunit of the DNA polymerase III holoenzyme	MP	-----qvdcebrhujgpolinx
dnaQ	DR0856	3′-5′ exonuclease subunit of the DNA polymerase III holoenzyme	MP	-----qvdcebrhujgpolinx
dnlJ	DR2069	DNA ligase	MP	-----qvdcebrhujgpolinx
ssb	DR0099	Single-strand-binding protein; *D. radiodurans* R1 has three incomplete ORFs corresponding to different fragments of the SSB	MP	-----qvdcebrhujgpolinx

Continued on following page

TABLE 2—*Continued*

Gene name[b]	Gene_ID	Protein description and comments	Pathway[c]	Phylogenetic pattern[d]
lexA	DRA0344, DRA0074	Transcriptional regulator, repressor of the SOS regulon, autoprotease	SOS	-----vdcebrh---------
ycjD	DR0221, DR2566	Uncharacterized proteins related to *vsr*	VSP?	--t---vd-e-rh---------
BS_*dinB*	13 homologs (see Fig. 5)	Uncharacterized family of presumably metal-dependent enzymes	?	-------dc-br----------
ham1/*yggV*	DR0179	Xanthosine triphosphate pyrophosphatase, prevents 6-*N*-hydroxylaminopurine mutagenesis	DR	amtkyqvdcebrh----olin-
uve1/BS_*ywjD*	DR1819	UV endonuclease; activity was characterized in *Neurospora*	NER	-------d-b----------
*yej*H/*rad25*	DRA0131	DNA or RNA helicase of superfamily II; also predicted nuclease; contains an additional *mcrA* nuclease domain	NER	a--ky--d-e-r------l---
	DR0690	Topoisomerase IB; currently the only bacterial representative of topoisomerase IB	?	----y--d--------------
	DR1721	3′→5′ nuclease; related to baculoviral DNA polymerase exonuclease domain	?	-------d--------------
	DR1262	Ro RNA binding protein; ribonucleoproteins complexed with several small RNA molecules; involved in UV resistance in *Deinococcus*	?	-------d--------------
	DR1757	Predicted nuclease and zinc finger domain-containing protein; an ortholog is present in *Pseudomonas aeruginosa*	?	-------d--------------
mrr	DR1877, DR0508, DR0587	MRR-like nuclease; restrictase of the *recB* archaeal Holliday junction resolvase superfamily	?	------vdc----u--------
tag[e]		3-Methyladenine DNA glycosylase I	BER	---------e-rh---------
vsr[e]		Strand-specific, site-specific, GT mismatch endonuclease; fixes deamination resulting from *dcm*	VSP	---------e------------
rusA (*ybcP*)[e]		Endonuclease/Holliday junction resolvase	RER	-------v-eb-----------
xseB[e]		Exonuclease VII, small subunit	MM	-------v-ebrh--------x
recB[e]		Helicase/exonuclease	RER	---------ebrhuj--olinx
recC[e]		Helicase/exonuclease	RER	---------e-rh----o-in-
ada[e]		*O*-6 alkylguanine, *O*-4 alkylthymine alkyltransferase; removes alkyl groups of many types; transcription activator	DR	amtkyqv--ebrhuj---lin-
alkB[e]		Unknown	DR, BER(?)	---------e------------
dut[e]		DUTPase	DR	----yq---ebrhuj---linx
dcd[e]		dCTP deaminase	DR	amtk-q--ce-rhuj----inx
nfo[e]		Endonuclease IV	BER	-mtkyqv--ebr---gp--in-
phrB[e]		Photolyase	DR	--t-y---ce------------
mutH[e]		Endonuclease	mMM	---------e------------
dam[e]		GATC-specific *N*-6 adenine methlytransferase; imparts strand specificity to mismatch repair	mMM	-m-k----ce--huj---l---
polB[e]		DNA polymerase II	SOS	amtky----e------------
sbcB[e]		Exodeoxyribonuclease I	mMM, RER	---------e--h---------
dcm[e]		Site-specific C-5 cytosine methyltransferase; VSP is targeted toward hot spots created by *dcm*	mMM	-mtk---dceb-huj-------
dinP[e]		Specific function unknown (predicted nucleotidyltransferase)	MM, RER	See umuC
recE[e]		Exonuclease VIII	RER	---------e------------
recT[e]		Annealing protein	RER	---------eb-----------
dinG[e]		Predicted helicase; SOS inducer	SOS	-mtkyq---ebrh---------
umuC[e]		Error–prone DNA polymerase; in conjunction with *umuD* and *recA*, catalyzes translesion DNA synthesis	SOS	----y---cebr---gp-----
umuD[e]		In conjunction with *umuC* and *recA*, facilitates translesion DNA synthesis; autoprotease	SOS	See LexA
radC[e]		Predicted acyltransferase; predicted DNA-binding protein	BER	-----qv-ceb-h---------

[a] Based largely on reference 20, with modifications

[b] The gene names are from *E. coli*, whenever an *E. coli* ortholog exists, or from *B. subtilis* (with the prefix BS_). *ham1* and *uve1* genes are from *Saccharomyces cerevisiae* and *Neurospora crassa*, respectively; where no ortholog was detectable in either *E. coli* or *B. subtilis*, no gene is indicated.

[c] Abbreviation of DNA repair pathways: DR, direct damage reversal; BER, base excision repair; NER, nucleotide excision repair; mMM, methylation-dependent mismatch repair; MMY, *mutY*-dependent mismatch repair; VSP, very-short-patch mismatch repair; RER, recombinational repair, SOS, SOS repair; MP, multiple pathways; ?, unknown possible repair pathways or uncertain assignments.

[d] Abbreviations in phylogenetic patterns: a, *Archaeoglobus fulgidus*; m, *Methanococcus jannaschii*; t, *Methanobacterium thermoautotrophicum*; k, *Pyrococcus horikoshii*; y, *Saccharomyces cerevisiae*; q, *Aquifex aeolicus*; v, *Thermotoga maritima*; c, *Synechocystis*; e, *E. coli*; b, *Bacillus subtilis*; r, *Mycobacterium tuberculosus*; h, *Haemophilus influenzae*; u, *Helicobacter pylori*; j, *Helicobacter pylori* J99; g, *Mycoplasma genitalium*; p, *Mycoplasma pneumoniae*; o, *Borrelia burgdorferi*; l, *Treponema pallidum*; i, *Chlamydia trachomatis*; n, *Chlamydia pneumoniae*, x, *Rickettsia prowazekii*.

[e] *E. coli* repair genes with no orthologs in *D. radiodurans*.

be less complex and diverse than those of bacteria with larger genomes, such as *E. coli* and *B. subtilis*. At the same time, there are several interesting and unusual aspects of the predicted layout of the repair systems in *Deinococcus* that may be linked to its phenotype (Table 2).

The nucleotide excision repair system that consists of the UvrABC excinuclease and the UvrD and Mfd (transcription-repair coupling factor) helicases is fully represented in *D. radiodurans*. Also present are the main components of the base excision repair system including several nucleotide glycosylases and endonucleases, namely, MutM (formamidopyrimidine and 8-oxoguanine DNA glycosylase); MutY (8-oxoguanine DNA glycosylase and apurinic DNA endonuclease-lyase); two paralogous uracil DNA glycosylases (Ung homologs); an additional, recently identified enzyme that has the same activity but is unrelated to Ung (DR1751) (174); endonucleases III (Nth) and V (YjaF); and exonuclease III (XthA). *Deinococcus* lacks two key enzymes involved in the repair of UV-damaged DNA in other organisms, namely, endonuclease IV (AP-endonuclease) and photo-lyase. Instead, it encodes a typical bacterial UV endonuclease III (thymine glycol-DNA glycosylase) and, more unexpectedly, a TIM-barrel fold nuclease characteristic of eukaryotes and most closely related to the UV endonuclease of *Neurospora* (20, 223). Eukaryotic-type topoisomerase IB is a truly unexpected protein to be identified in the *Deinococcus* genome and also could play a role in UV resistance (see "Horizontal gene transfer" below).

The repertoire of recombinational repair genes in *Deinococcus* includes orthologs of most of the *E. coli* genes involved in this process (Table 2), but the RecBCD recombinase is missing. While this complex is not universal in bacteria, it is a major component of recombination systems in most free-living species. In *Deinococcus*, where recombination is thought to be an important contributor to damage-resistance, the absence of this ATP-dependent exonuclease is unexpected. *Deinococcus* does encode an apparent ortholog of one of the helicase-related subunits of this complex, RecD, but not the other subunits. The RecD protein in *Deinococcus* is unusual in that it contains an N-terminal region of about 200 amino acid residues that consist of three tandem predicted HhH DNA-binding domains; this unusual domain organization of the RecD protein is shared with *B. subtilis* and *Chlamydia*. Such dissociation of RecD from the RecB and RecC subunits is not unique to *Deinococcus*; "solo" RecD-related proteins are also present in *M. jannaschii* and in yeast. The function(s) of RecD, once outside the recombinase complex, is unknown.

Another component of the recombinational repair system in *Deinococcus* that has an unusual domain architecture is the RecQ helicase. It contains three tandem copies of the C-terminal helicase-RNase D (HRD) domain, instead of the single copy present in all other bacteria except *Neisseria* that similarly possesses three copies (141) (also see below). RecQ sequences from *Neisseria* and *Deinococcus* are more similar to each other than to any other homologs, which, together with the distinctive triplication of the HRD domain, indicates that the *recQ* gene has been exchanged between bacteria from these two distant lineages. In addition, *Deinococcus* encodes a protein (DR2444) that contains an HRD domain and a domain homologous to cystathionine gamma-lyase; this is the first example of an HRD domain that is not associated with either a helicase or a nuclease (although it is possible that the domain organization of this protein is an artifact caused by a frameshift). This propagation of the HRD domain in *Deinococcus* could contribute to the repair phenotype given the interactions of RecQ with RecA in recombination (88).

The methylation-dependent mismatch repair system of *D. radiodurans* includes the MutS and MutL ATPases and endonuclease VII (XseA). Orthologs of the site-specific methylases Dcm and Dam, which are associated with mismatch repair, are not readily detectable. It appears likely, however, that other distantly related DNA methylases predicted in *D. radiodurans* could perform similar functions.

Like other bacteria with large genomes, *D. radiodurans* encodes the LexA repressor-autoprotease (DRA0344), which in *E. coli* and *B. subtilis* controls the expression of the SOS regulon. In addition, unlike any of the other bacterial genomes studied, *D. radiodurans* encodes a second, diverged copy of LexA (DRA0074), which retains the same arrangement of the helix-turn-helix (HTH) DNA-binding domain and the autoprotease domain. Attempts to identify LexA-binding sites and the composition of the putative SOS regulon in *D. radiodurans* have been unsuccessful (M. S. Gelfand, personal communication). This suggests that *D. radiodurans* does not possess a functional SOS response system, which is in agreement with the results of previous experimental studies (142). Furthermore, *Deinococcus* does not encode proteins of the DinP/UmuC family, nonprocessive DNA polymerases that play a critical role in translesion DNA synthesis and associated error-prone repair such as SOS repair in *E. coli* (117).

In addition to orthologs of well-characterized repair proteins discussed in this section, *Deinococcus* encodes several unusual proteins and expanded protein families that are less confidently associated with repair but might contribute to the unusual effectiveness of the repair and recombination systems in this bacterium; these proteins are discussed below in the section on the unique features of the *Deinococcus* proteome.

Stress Response and Signal Transduction Systems

D. radiodurans encodes a broad spectrum of proteins that have been associated with various forms of stress response in other bacteria as well as several proteins that appear to be unique and could contribute to more specific forms of the stress response (Table 3). Orthologs of almost all known genes involved in different stress responses in other bacteria (109) are present in *Deinococcus*. The few stress response proteins that are missing are either specific to the adaptation of a particular organism to its environment or, when of more general significance, likely to be replaced by nonorthologous proteins with similar functions. For example, instead of using the OtsA and OtsB proteins for the synthesis of the osmoprotection disaccharide trehalose, *Deinococcus* probably uses an alternative pathway via trehalose synthase (DR0933), which has been recently characterized in *Thermus* (209). Trehalose plays a major role in the desiccation resistance of *E. coli* (216) and is also likely to be important in *Deinococcus. Deinococcus* has two additional genes for trehalose metabolism: maltooligosyl trehalose synthase (DR0463), which provides yet another route of trehalose formation, and trehalohydrolase (DR0464). These genes apparently form a mobile operon and probably

TABLE 3. Stress response-related genes in *D. radiodurans*

Gene name	Gene_ID	Protein description and comments[a]	Type of stress	Phylogenetic pattern[a]
groL	DR0607	Hsp10, molecular chaperone	Heat, general	amtkyqvdcebrhujgpolinx
grpE	DR0128	Hsp20, molecular chaperone	Heat, general	--t--qvdcebrhujgpolinx
groS	DR0606	Hsp60, molecular chaperone	Heat, general	----yqvdcebrhujgpolinx
dnaK	DR0129	Hsp70, molecular chaperone	Heat, general	--t-yqvdcebrhujgpolinx
dnaJ	DR0126, DR1424	Hsp70 chaperone cofactor	Heat, general	--t-yqvdcebrhujgpolinx
ibpA/*ibpB*	DR1114, DR1691	Small heat shock protein	Heat, general	amtkyqvdcebr---------x
hslJ	DR2056, DR1940	Related to heat shock protein, HslJ; DR1940 contains three repeats of this domain	Heat	-------dce-----------
clpA/*clpB*	DR0588, DR1046, DR1117	ATPase subunit of Clp protease	General	--t-yqvdcebrhujgpolinx
clpX	DR1973, DR0202	ATPase subunit of Clp protease	General	----yqvdcebrhuj--olinx
clpP	DR1972	ATP-dependent protease with chaperone activity	General	----yqvdcebrhuj--olinx
lon	DR1974, DR2189, DR0349	ATP-dependent Lon serine protease	General	----yqvd-eb-hujgpolinx
sms	DR1105	ATP-dependent serine protease	General	-----qvdcebrhuj---linx
htrA	DR0327, DR0745, DR1599, DR1756, DR0984, DR0300	Do serine protease, with regulatory PDZ domain	General	--t-yqvdcebrhuj--olinx
prc	DR1308, DR1491, DR1551	Tail-specific periplasmic serine protease	General	-----qvdceb-huj--olinx
yaeL	DR1507	Membrane-associated Zn-dependent protease I	General	amtk-qvdcebrhuj--olinx
ftsH	DR0583, DR1020, DRA0290	ATP-dependent Zn protease	General	----yqvdcebrhujgpolinx
htpX	DR0190, DR0194	Predicted Zn-dependent proteases (possible chaperones)	General	amtkyq-dcebrhuj-------
sugE	DR1004, DR1005	Membrane chaperone	General	-------d-ebr----------
hit	DR1621	Diadenosine tetraphosphate (Ap4A) hydrolase, HIT family, cell cycle regulation	General	amtkyq-dcebrhujgpo-inx
yebL	DR2523	Zn-binding (lipo)protein of the ABC type Zn transport system (surface adhesin A)	General	amtk-qvdcebrh-----linx
hflX	DR0139, DR0646	GTPase, protease modulator	General	-m-k-qvdcebrhujgpolinx
BS_*yloA*	DR0559	Fibronectin-binding protein, function unknown	?	amtky-vdc-b--uj--o----
BS_*ytxJ*	DR1832	General stress protein, related to thioredoxin	General	-------d--b----------
thiJ	DR0491, DR1199	Protease I, related to general stress protein 18, ThiJ superfamily protein	General	am-kyq-dcebrhujgpol---
uspA	DR2363, DR2132	Universal stress protein, nucleotide-binding	General	amtk-q-dcebrh--------x
spoT	DR1838	Guanosine polyphosphate (ppGpp) pyrophosphohydrolase/synthetase: no RelA counterpart like in gram-positive bacteria	Starvation	-------dcebrhujgpo----
hupA	DRA0065	Histone-like DNA-binding protein	?	-----qvdcebrhujgpolinx
hmp	DRA0243	Haemoglobin-like flavoprotein	?	----yq-d-eb----------
mazF	DR0417, DR0662	ppGpp-regulated growth inhibitor	Starvation	-------d-ebr----------
mazE	DR0416	Regulatory protein, MazF antagonist	Starvation	-------dce-----------
ppx	DRA0185	Phosphatase of ppGpp	?	----yqvdce--huj------x
dps	DR2263, DRB0092	Starvation inducible DNA-binding protein	Starvation	-------dceb-huj-ol--x
mscL	DR2422	Large conductance mechanosensitive channel	Osmotic	-------dcebrh--------
yggB	DR1995, DR0211	Membrane protein	Osmotic	amtk-qvdcebrhuj-ol--x
kdpD	DRB0088	Osmosensitive K^+ channel histidine kinase sensor domain	Osmotic	-------dce-r----------
trkA	DR1666	Potassium uptake system, NAD-binding component	Osmotic	amtk-q-dcebrh-gpol---
trkH/*trkG*	DR1667, DR1668	Potassium uptake system component	Osmotic	amtk-qydceb-h-gpol---
proW	DRA0138, DRA0139	Proline/glycine betaine ABC-type transport, permease subunit	Osmotic	a------debr-hj--o----
proV	DRA0137	Proline/glycine betaine ABC-type transport, ATPase subunit	Osmotic	a------debr-hj--o----
yehZ	DRA0135	Proline/glycine betaine ABC-type transport, periplasmic binding subunit	Osmotic	a------debr-hj-------
pspA	DR1473	Phage shock protein A, controls membrane integrity	Phage	-----q-dceb-----------
BS_*yloU*/BS_*yqhY*	DR2068, DR0389	Alkaline shock protein, function unknown	Alkaline	------vd--b--------in-
csp	DR0907	Cold shock protein, OB fold nucleic acid-binding protein	Cold	-----qvd-ebrh--------x
cinA	DR2838	Competence damage protein, mitomycin-induced, function unknown	?	-----qvdcebr-hjgp-----

Continued on following page

TABLE 3—*Continued*

Gene name	Gene_ID	Protein description and comments[a]	Type of stress	Phylogenetic pattern[a]
katE	DR1998, DRA0259	Catalase; DRA0259 has C-terminal proteinase I-like domain	Oxidative	----y--d-eb-huj-------
catA (*S. pombe*)	DRA0146	Catalase; eukaryotic type, presumably acquired from nitrogen-fixing bacteria	Oxidative	-------d--------------
NA[c]	DRA0145	Peroxidase; Yet present only in plant Polyporaceae spp.	Oxidative	-------d--------------
sodA	DR1279	Superoxide dismutase, Mn or Fe dependent	Oxidative	-t--yq-dcebrhuj--o-inx
sodC	DR1546, DRA0202	Superoxide dismutase, Cu/Zn dependent	Oxidative	----yq-d-ebr----------
fur	DR0865	Ferric uptake regulation protein	?	a----qvdcebrhujgpo----
bcp	DR0846, DR1208	Antioxidant type thioredoxin fold protein	Oxidative	----yqvdcebrhuj-------
osmC	DR1538, DR1857	Protein involved in alkylperoxide and oxidative stress response, osmotically induced protein	Oxidative	-------d-eb----gp-----
yhfA	DR1177	Protein involved in alkylperoxide and oxidative stress response, osmotically induced protein	Oxidative	---k-qvd-e------------
msrA	DR1849	Peptide methionine sulfoxide reductase PMSR	Oxidative	-t-y--dcebrhujgp-l---
ahpC	DR2242, DR1209	Thiol-alkyl hydroperoxide reductases	Oxidative/ detoxication	amtkyqvdcebr-uj---linx
ahpF/*trxB*	DR1982, DR2623, DR0412, DRB0033	Thioredoxin reductase/alkyl hydroperoxide reductase	Oxidative/ detoxication	amtkyqvdcebrhujgpolinx
grxA	DR2085, DRA0072	Glutaredoxin	Oxidative/ detoxication	a-tky-vdcebrh-----l-x
BS_*cypA* or *terA* of *Alcaligenes*	DR2473, DR2538, DR1723, DRA0186, DRC0041, DRC0001	Cytochrome P450 (uses O_2)	Detoxication	----y--dc-b-----------
terB of *Alcaligenes*	DR2220	Function unknown; involved in tellurium resistance response in *Alcaligenes*	Detoxication	-------d--------------
terC of *Alcaligenes*	DR2226, DR1187, DRB0131	Function unknown; membrane protein	Detoxication	-------dcebrhuj------x
BS_*yceH*	DR1127	Toxic anion resistance protein; possibly tellurite resistance	Detoxication	--------d-b------------
BS_*scp2*	DR2225, DR2221, DR2224, DR2223, DRA203	Chemical damaging agent resistance; in *B. subtilis* it is involved in low-temperature and salt stress response	Toxins/general	-------dc-b------------
arsC	DRA0123, DR0136	Arsenate oxidoreductase (*arsC*-like rodanese protein)	Toxins	-------d-ebh-----------
NA	DR1372	Desiccation protectant, LEA14 family	Desiccation	am-k---d--------------
NA	DRB0118	Desiccation-related protein from *Craterostigma plantagineum*; found to date only in plants	Desiccation	-------d-------------
NA	DR0105, DR1172	LEA76 family desiccation resistance protein	Desiccation	-------d--------------
BS_*ybfO*	DRA0345, DR2257	Erythromycin esterase	Drugs	-------d--br----------
bacA	DR0454	BacA bacitracin resistance protein, undecaprenol kinase	Drugs	-----qvdcebr-----o----
strA of *Streptomyces*	DR0455	Streptomycin resistance protein, streptomycin phosphotransferase	Drugs	-------d--------------
BS_*ycbJ*	DR0066, DRA0194, DR0394, DR0669	Antibiotic (aminoglycoside) kinase family protein	Drugs	-------dc-br----------
nimABCD of *Bacteroides*	DR0842	5-Nitroimidazole antibiotic resistance protein; distantly related to pyridoxamine phosphate oxidase (PDXH)	Drugs	-------d--------------
BS_*bmrU*	DR2234, DR1363, DR1560	Function unknown; involved in multidrug resistance	Drugs	----y-vdcebr----------
thdR	DR1016	Thiophen and furan oxidation, predicted GTPase	Drugs	----yqvdceb-hujgpolinx
BS_*tmrB*	DR1419	Tunicamycin resistance protein, predicted ATPase	Drugs	-------d--b-----------
BS_*penP*	DRA0241, DR0433	β-Lactamase	Drugs	-------dc-br----------
ybgL	DRA0284	Function unknown, lactam utilization protein	Drugs	---k---d-eb-h----------
gloA	DR1695, DR2022, DR2104, DR2208, DR0109, DRA0224	Lactoylgluthation lyase, fosphomicin resistance protein	Drugs	amtk---dcebrh----------
BS_*yoaR*	DR1619, DR0009, DR0025	Induced by vancomycin in *Enterococcus faecalis*	Drugs	-------d--b-----------
BS_*yokD*	DR2034, DR0599	Aminoglycoside *N*3-acetyltransferase; present in many other bacteria	Drugs	-------d--b-----------

Continued on following page

TABLE 3—*Continued*

Gene name	Gene_ID	Protein description and comments[a]	Type of stress	Phylogenetic pattern[a]
BS_*yocD*	DR2000	Function unknown; homologs of microcine C7 resistance protein MccF	Drugs	-------dceb--------l--x
lytB	DR2164	Function unknown; penicillin tolerance protein	Drugs	-----qvdcebrhuj---lin-
BS_*yrpB*	DR2545	2-Nitropropane dioxygenase	Drugs	a---yqvd--br-uj-------
BS_*ywnH*	DR1182	Phosphinothricin aminoacetyltransferase	Drugs	-------dceb-----------
cstA[b]		Carbon starvation-induced protein, membrane	Starvation	-----q---ebrhuj-------
clpY[b]		ATPase subunit of clp proteopltic system	General	-----qv--eb-huj---o---x
btuE/BS_*bsaA*[b]		Glutathione peroxidase	Oxidative	----y---ceb-----------
katG[b]		Catalase (peroxidase I)	Oxidative	a---y---ce-r----------
htpG[b]		HSP90, molecular chaperone	Heat, general	----y---cebrhuj--ol-x
dksA[b]		DnaK suppressor protein	Heat, general	-----q---eb-h----olinx
hrcA[b]		Transcriptional regulator of heat shock genes	Heat, general	------v-c-br-ujgp--in-
yajQ[b]		Unknown	Acid	---------ebrh---------
otsA[b]		Trehalose-6-phosphate synthase	Osmotic	--t-y----e-r----------
otsB[b]		Trehalose-6-phosphatase	Osmotic	--t-y---c-r----------

[a] Abbreviations in the phylogenetic patterns and gene descriptions are as in Table 2.
[b] Genes that are absent in *D. radiodurans*.
[c] NA, not applicable.

have been acquired by *Deinococcus* through horizontal transfer, since their closest homologs are found in *Rhizobium*, where they appear to have the same operon organization (130).

Among the proteins associated with oxidative stress response, *Deinococcus* encodes three catalases (DR1998, DRA0259, and DRA0146), two of which are highly similar to one another and to catalases from other bacteria whereas the third is only distantly related to other catalases. The gene for this unusual predicted catalase (DRA0146) is closely linked to and probably forms an operon with a gene for a peroxidase (DRA0145). DRA0146 is most similar to its ortholog from *Rhizobium*, and these two proteins are, in turn, more closely related to eukaryotic catalases from plants than to bacterial catalases. This suggests that *Deinococcus* acquired the gene for this catalase from a nitrogen-fixing bacterium, which, in turn, had hijacked it from a plant. In contrast, DRA0145 is distinctly closer to certain peroxidases from fungi, such as *Galactomyces geotrichum*, than to bacterial forms from *Neisseria, E. coli*, and actinomycetes. Thus, the entire operon probably has been acquired horizontally. A broad spectrum of other genes that may be involved in the stress response include DRA0149 (agmatinase), DR1353 (an acid-inducible apolipoprotein aminoacetyltransferase), and DR2299, DR1605, and DR2245 (genes of the two-component response and cyclic diguanylate signaling system), which again are very similar to homologs from the family *Rhizobiaceae*, suggesting significant horizontal gene transfer between these distant bacteria.

In addition to the well-characterized components of stress response systems, *Deinococcus* encodes several proteins and entire protein families whose specific roles are unknown but are likely to be important for the multiple stress resistance phenotypes of the bacterium. An example of a poorly studied but potentially important system is the "addiction module" response (2), which is encoded by two genes, *mazE* and *mazF* (DR0416 and DR0417, respectively). MazF is a stable protein that is toxic to bacteria, whereas MazE protects cells from the toxic effect of MazF and is degraded by the ClpP serine protease. Expression of these two genes is regulated by ppGpp, which is produced by the RelA enzyme (or the bifunctional enzyme SpoT) in response to amino acid starvation. On the basis of these studies, Aizenman et al. (2) have proposed a model of programmed bacterial cell death dependent on the MazEF proteins. Currently, *Deinococcus* is the only bacterium other than *E. coli*, the model system in which the role of these proteins was elucidated, that has both genes and retains their operon organization. Another example of poorly characterized genes that are likely to be involved in stress response are two proteins (DR2056 and DR1940) that are homologous to the *E. coli* heat shock protein HslJ (42). One of these proteins, DR1940, contains three copies of the HslJ domain, a feature that has not yet been seen in this protein family. All the HslJ domains contain two conserved cysteines that could function as a redox pair, with the protein itself being a disulfide bond chaperone. The only prominent chaperone that is missing without an obvious replacement is HSP90, but this gene is also absent in archaea and bacterial thermophiles and therefore appears to be nonessential.

The signal transduction system of *D. radiodurans* has chimeric features of prokaryotic and eukaryotic systems. This form of chimerism in the signaling system is becoming increasingly evident in several bacterial lineages such as actinomycetes, myxobacteria, and spore-forming firmicutes that undergo cellular differentiation. The typically bacterial components of the signaling system include the two-component systems with the histidine kinase and receiver domains (159) and the cyclic diguanylate signaling system with the GGDEF, EAL, and HD_GYP domains, which appear to function as cyclases and phosphodiesterases (75). In addition, these signaling domains are typically combined with small molecule and protein-binding domains, such as PAS and GAF (17, 203), and the conformation-signaling HAMP domain (16). The two-component phosphorelay system is well developed in *Deinococcus*, which encodes 23 histidine kinase domains and 29 receiver domains that form several combinations with the GAF and PAS domains. This system is expected to play a major role in sensing redox, light, and other environmental stimuli. Consistent with

this, DRA0050, which is orthologous to the cyanobacterial and plant phytochromes, has been shown to be a photoreceptor involved in the regulation of pigment biosynthesis (55), which is likely to affect resistance to DNA-damaging agents (35). Genes encoding two proteins that consist of a sensory transduction histidine kinase and a receiver domain (DRB0028 and DRB0029) appear to be coregulated with an s^B operon (DRB0024 to DRB0027). This operon encodes the antisigma factor-regulatory system and is known to be involved in stress response in other bacteria (92, 109). As a whole, this array of six genes appears to comprise a stress response module unique for *Deinococcus*.

Deinococcus encodes 16 GGDEF domain-containing proteins, which suggests a major role for this uniquely bacterial module that is predicted to function as a cyclase in diguanylate signaling. The two predicted distinct phosphodiesterases of this system, the HD-GYP and EAL domains (six and four copies, respectively, in *Deinococcus*), complement each other in terms of their copy numbers, as has been observed for other bacterial genomes. These domains tend to combine with the stimulus-sensing PAS and GAF domains. One such interesting architecture is the combination of the GAF domain and the HD_GYP domain in two *Deinococcus* proteins (Fig. 2). The representation of this signaling system in *Deinococcus* is comparable to that in other bacteria with moderate-sized to large genomes.

While *Deinococcus* lacks flagella and is unlikely to be capable of chemotactic motility, it possesses certain remnants of the chemotactic signaling system that are likely to signal through alternative pathways. In particular, there are three methyl-accepting chemotactic receptor proteins (DRA0352, DRA0353, and DRA0354), each containing two HAMP domains, but there is no methyltransferase of the chemotactic signaling pathway. These three proteins are encoded by genes located in the vicinity of genes for a CheA-like histidine kinase and a CheY-like receiver domain, which suggests that the methyl-accepting receptor forms a single functional unit with this two-component system protein. Given the apparent absence of chemotaxis, the methyl-accepting receptors could form a scaffold for binding of the CheA kinase, which might signal the availability of amino acids in the environment.

The tetratricopeptide repeats (TPR) seem to play a special role in *Deinococcus* signaling. In three distinct proteins, these repeats are combined with typically bacterial signaling modules (Fig. 2). The TPR modules are likely to mediate protein-protein interactions within molecular complexes involving these proteins, as documented in eukaryotic systems (113). WD40 proteins, which often serve as interaction partners to TPR in eukaryotes (210), are also expanded in *Deinococcus* and could cooperate with the TPR-containing proteins. Of particular interest is another group of at least four β-propeller proteins that appear to be closer to the YWTD class of propellers than to WD40s (DR0960, DR1725, DR2062, and DR2484). In actinomycetes, these propeller domains are fused to protein kinases and are likely to perform specific protein-protein interaction functions in signaling (163).

The prominence of the "eukaryotic" component of the signal transduction systems in *Deinococcus* is underscored by the fact that it encodes 11 Pkn2-type kinases and 1 kinase of the RIO1 family (DR2209), which is typical of archaea and eukaryotes (121) and was detected in bacteria for the first time. This number is greater than in most other prokaryotes (121), suggesting that protein-serine/threonine phosphorylation-dependent regulatory pathways play a major role in *Deinococcus*. Consistent with this, *Deinococcus* also encodes PP2C phosphatases and a FHA domain that typically function in conjunction with the serine/threonine kinases.

Several protein families that have been implicated in stress response and signal transduction in other organisms have undergone specific expansion in *Deinococcus*; these are discussed in some detail below.

Distinctive Features of Predicted Operon Organization and Transcription Regulation

Generally, the genome organization of *D. radiodurans* is similar to that of other bacteria (218). Many functionally related genes are organized into clusters that are likely to comprise operons, including such common ones as ribosomal protein genes, ATP synthase, NADH dehydrogenase, and various ATP-binding cassette (ABC)-type transport systems. Beyond these generic operons, however, several unusual gene clusters were detected, and some of these are likely to be related to the unique features of *Deinococcus* (Table 4).

The first group of such unique gene arrays includes paralogous genes that encode protein families overrepresented in *Deinococcus*, such as amino-acetyltransferases, Nudix hydrolases, and genes of the TerE and DinB/YfiT families (see below). Some of these clusters appear to have evolved by tandem duplication within the *Deinococcus* lineage, e.g., an acetyltransferase cluster (DR2254 and DR2255) and a Nudix cluster (DR0783 and DR0784). Other clusters of paralogs clearly resulted from a single horizontal transfer event, e.g., the group of tellurium resistance genes (DR2220 to DR2226) that are related to the corresponding gene cluster on the broad-host-range plasmid R478. Finally, some clusters that consist of related genes with apparent phylogenetic affinities to different bacterial lineages (e.g., an acetyltransferase cluster [DR0675 to DR0677]) seem to have originated within the *Deinococcus* lineage through gene translocation. The second group of unusual predicted operons includes rare gene clusters that probably were acquired by horizontal transfer. Some of these operons could contribute to damage resistance, e.g., DNA repair-related functions (deoxypurine kinase operon [DR0298 and DR0299], eukaryotic-type uracil-DNA-glycosylase and topoisomerase IB [DR0689 and DR0690]), DNA transformation-related functions (competence genes [DR1854 and DR1855], restriction-modification system [DRB0143 and DRB0144]), stress response (DR0389 and DR0390; DR1160 and DR1161), and pigment biosynthesis (DR0861 and DR0862).

Two operons (DR0853 to DR0854 and DR2180 to DR2181) each consist of a gene for a small GTPase of the Ras/Rab family and a gene coding for a small protein of an uncharacterized family that is widespread in bacteria and archaea (L. Aravind and E. V. Koonin, unpublished data). The orthologous GTPase in *Myxococcus* is important for gliding motility (90), suggesting a role for these proteins in signaling. Expansion of the uncharacterized protein family encoded by the genes adjacent to the GTPase is seen in *Streptomyces* and *Deinococcus* and appears to result from relatively recent du-

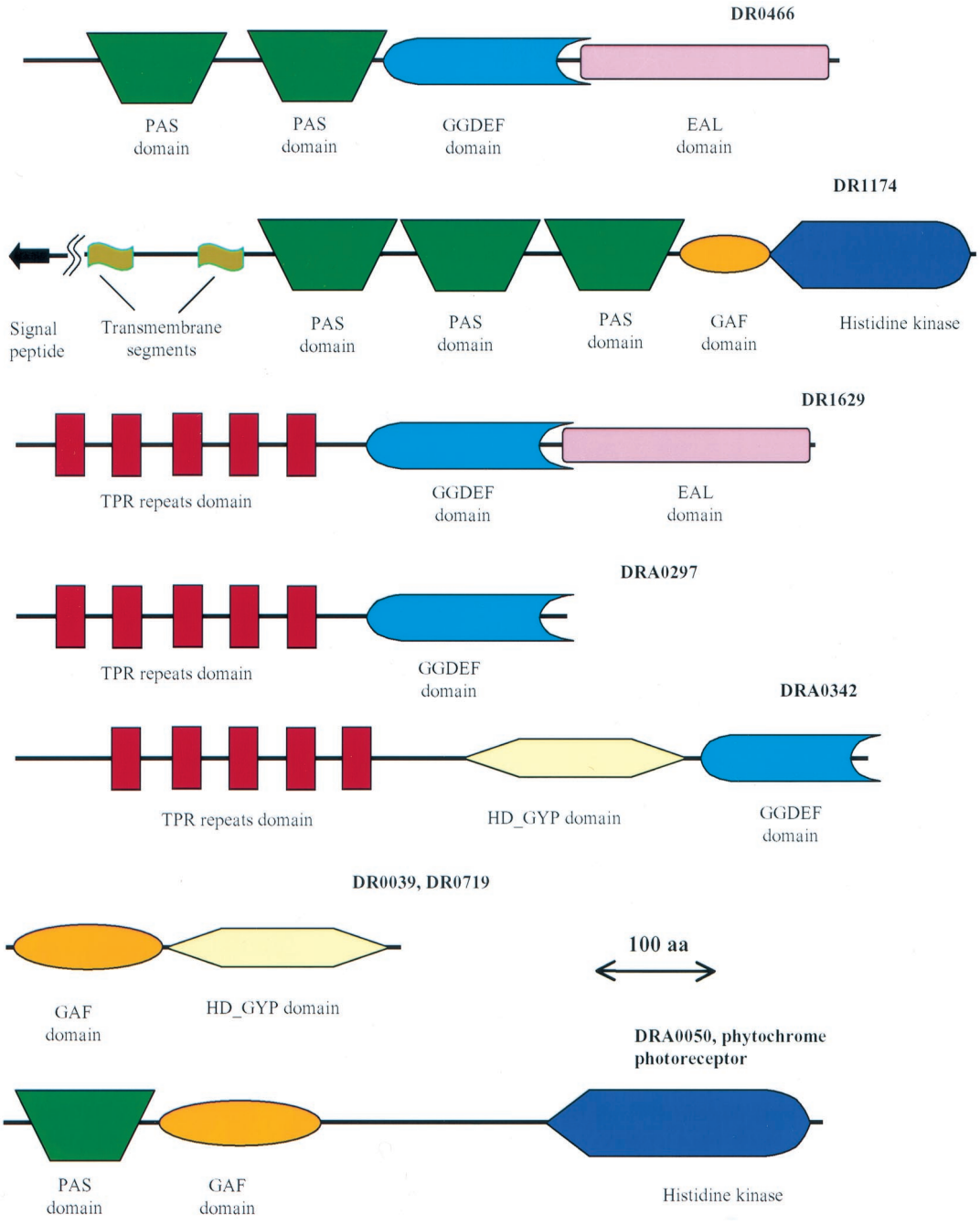

FIG. 2. Distinct domain architectures of selected proteins implicated in signal transduction in *Deinococcus*.

plications (DR0616, DR0995, and DR1612), with three of these genes forming a cluster in the chromosome (DR0993 to DR0995). Juxtaposition of these genes with genes for Ras/Rab-GTPases is frequently observed in other genomes, including *Myxococcus* and archaeal and bacterial thermophiles, suggesting that they form a mobile operon, with the encoded proteins being functionally coupled.

Another predicted operon (DR0332 to DR0335) that could have been horizontally transferred from cyanobacteria encodes components of a protein kinase-dependent regulatory pathway. These include two active Pkn2-type serine/threonine protein kinase with Zn ribbons, a PP2C-type phosphatase with an N-terminally disrupted Pkn2 kinase domain, and a protein that contains a phosphoserine-binding FHA domain combined with a Zn ribbon domain orthologous to proteins from cyanobacteria (FraH) and actinomycetes (121). The phosphorylation system encoded by this operon may play a role in cellular differentiation, with the Zn-ribbon-FHA protein functioning as the downstream effector that regulates transcription.

The general picture of transcription regulation in *Deinococcus* emerging from genome analysis is similar to that seen in other bacteria. Among *Deinococcus* gene products, we de-

TABLE 4. Some unusual predicted operons in *D. radiodurans*

Gene cluster[a]	Protein description	Best hit: species and GI number	Comment
DR0298	Deoxypurine kinase subunit, YAAF	*Bacillus subtilis* (586859)	Clear case of gene exchange with gram-positive bacteria; essential enzymes for biosynthesis of deoxyribonucleotides
DR0299	Deoxypurine kinase subunit, YAAG	*Bacillus subtilis* (586860)	
DR0398	Alkaline-shock-like protein	*Bacillus subtilis* (2337812)	Clear case of gene exchange with gram-positive bacteria; possibly stress related; conserved operon in *B. subtilis* and *T. maritima*; all three bacteria encode an additional copy of alkaline-shock-like protein
DR0390	Uncharacterized protein, YloV ortholog	*Bacillus subtilis* (2337813)	
DR0544	Highly conserved membrane transporter	*Pyrococcus horikoshii* (3256923)	Likely gene exchange with archaea; possibly related to the multidrug resistance system
DR0545	Small conserved membrane protein, possibly involved in multidrug resistance	*Methanobacterium thermoautotrophicum* (2621904)	
DR0674	Argininosuccinate synthase, ArgG	*Thermotoga maritima* (4982360)	Acetyltransferase cluster disrupts ArgG/ArgH operon present in some other bacteria
DR0675	Amino acetyltransferase	*Synechocystis* (1651699)	
DR0676	Amino acetyltransferase, related to phosphinothricin acetyltransferase	*Escherichia coli* (1742360)	
DR0677	Amino acetyltransferase	*Salmonella enterica* serovar Typhimurium (586786)	
DR0678	Argininosuccinate lyase (ArgH)	*Bacillus subtilis* (2293243)	DR0681 and DR0683 may have evolved by internal duplication; DR0679 and DR0680 together may comprise a mobile element; this pair of genes is present twice more in the *D. radiodurans* genome
DR0679	Small nucleotidyltransferase	*Synechocystis* (1653122)	
DR0680	Uncharacterized protein next to small nucleotidyltransferases	*Synechocystis* (1652090)	
DR0681	Amino acetyltransferase	*Aquifex aeolicus* (2983780)	
DR0682	Amino acetyltransferase	*Salmonella enterica* serovar Typhimurium (586786)	
DR0683	Amino acetyltransferase	*Aquifex aeolicus* (2983780)	
DR0689	Uracil-DNA glycosylase (Ung)	Human (137031)	Likely horizontal transfer from a eukaryote or a eukaryotic virus
DR0690	Topoisomerase IB	Orf virus (521138)	
DR0796	Amino acetyltransferase	*Bacillus subtilis* (1881232)	Acetyltransferase cluster; DR0796 and DR0797 are possible products of internal duplication
DR0797	Amino acetyltransferase	*Synechocystis* (1651699)	
DR0798	Amino acetyltransferase	*Bacillus subtilis* (1881232)	
DR0853	Rab/Ras family small GTPase	*Myxococcus xanthus* (94524)	Potential operon conserved also in *Thermus, Myxococcus* and archaea; involved in gliding motility in *Myxococcus*
DR0854	Protein associated with a GTPase	*Myxococcus xanthus* (94525)	
DR0861	Phytoene dehydrogenase, CRTI	*Flavobacterium* (1842244)	Carotenoid biosynthesis genes; Possibly involved in pigment biosynthesis in *Deinococcus*
DR0862	Phytoene synthase, CRTB	*Thermus thermophilus* (585011)	
DR0993	Uncharacterized protein associated with GTPase	*Methanococcus jannaschii* (1591982)	Cluster of genes that are expanded in *Deinococcus* and encode uncharacterized small proteins often associated with Ras/Rab family GTPase (see also DR0853–DR0854 and DR2180–DR2181)
DR0994	Uncharacterized protein associated with GTPase	Distantly related to the family of GTPase-associated proteins	

Continued on following page

TABLE 4—*Continued*

Gene cluster[a]	Protein description	Best hit: species and GI number	Comment
DR0995	Uncharacterized protein associated with GTPase	*Aquifex aeolicus* (2984135)	
DR1175	N-terminal CheY family domain +C-terminal histidine kinase	*Mycobacterium tuberculosis* (2960188)	Signal transduction system; proteins with modified domain architectures compared to *M. tuberculosis* and *Synechocystis*
DR1174	Histidine kinase with 3 PAS +3 PAC + GAF domains	*Synechocystis* (1652132)	
DR1232	Pilin IV-like secreted protein	*Psedomonas putida* (544344)	Pilin IV cluster also including a GTPase with a possible regulatory role; probably responsible for DNA transformation
DR1233	Pilin IV-like secreted protein	*Legionella pneumophila* (3002996)	
DR1234	Pilin IV-like secreted protein	*Klebsiella oxytoca* (131598)	
DR1235	Dynamin-like GTPase	*Arabidopsis thaliana* (4587579)	
DR1596	Glucose-6-phosphate 1-dehydrogenase	*Synechocystis* (2494656)	Clear case of gene exchange with cyanobacteria
DR1597	OPCA, OxPPCycle gene, involved in assembly of glucose-6-phosphate 1-dehydrogenase	*Synechocystis* (2498703)	
DR1641	DinB/YfiT superfamily protein	Both distantly similar to *Bacillus subtilis* (2633163)	See main text
DR1642	DinB/YfiT superfamily protein		
DR1928	Glycerol kinase (GlpK)	*Borrelia burgdorferi* (2688136)	Clear case of gene exchange with spirochetes
DR1929	Glycerol uptake facilitator (GlpK)	*Borrelia burgdorferi* (2688137)	
DR2180	Uncharacterized protein associated with GTPase	*Aquifex aeolicus* (2984135)	Clear case of gene exchange with thermophiles (See above)
DR2181	RAB/RAS-like small bacterial GTPase, inactivated	*Aquifex aeolicus* (2984130)	
DR2220	Tellurium resistance protein (TerB)	Plasmid R478 (950680)	Tellurium resistance gene cluster; probable acquisition of a plasmid fragment; stress response-related genes; operon is probably disrupted by a transposon
DR2221	Tellurium resistance, member of *Dictyostelium*-type cAMP-binding protein family	*Alcaligenes* sp. (135597)	
DR2222	Transposase	No significant similarity	
DR2223	Tellurium resistance, member of *Dictyostelium*-type cAMP-binding protein family	*Alcaligenes* sp. (78205)	
DR2224	Tellurium resistance, member of *Dictyostelium*-type cAMP-binding protein family	Plasmid R478 (1181183)	
DR2225	Tellurium resistance, member of *Dictyostelium*-type cAMP-binding protein family	Plasmid R478 (950682)	
DR2226	Tellurium resistance, membrane protein (TerC)	*Mycobacterium tuberculosis* (2105065)	
DR2254	Amino-acetyltransferase	*Streptomyces coelicolor* (5531439)	Acetyltransferase cluster; these proteins are likely to be a product of internal duplication
DR2255	Amino-acetyltransferase	*Streptomyces coelicolor* (5531439)	
DR2311	Uncharacterized protein, YeiN ortholog	*Escherichia coli* (465602)	Among bacteria, YeiN orthologs are present only in *Deinococcus* and gamma proteobacteria; in eukaryotes, their counterparts are fused with the kinase gene; probable gene exchange with proteobacteria
DR2312	RBSK family ribokinase, YeiI ortholog fused to HTH domain	*Escherichia coli* (2507177)	

Continued on following page

TABLE 4—*Continued*

Gene cluster[a]	Protein description	Best hit: species and GI number	Comment
DRA0231	Oxidoreductase	*Escherichia coli* (2495497)	Highly conserved paralogous gene cluster is located next to this one in the chromosome (DRA0235–37), but in the opposite orientation
DRA0232	Flavoprotein dehydrogenase	*Escherichia coli* (2495498)	
DRA0233	Dehydrogenase, iron sulfur protein	*Escherichia coli* (2495499)	
DRA0331	Von Willebrand factor A domain, Mg^{2+} binding	*Synechocystis* sp. (2496792)	Serine/threonine protein kinase-based regulatory system
DRA0332	PKN2 family serine threonine kinase	*Anabaena* sp. (1709645)	
DRA0333	Zn-finger and FHA domain-containing protein, ortholog of cyanobacterial FraH	*Anabaena* sp. (556608)	
DRA0334	Inactive kinase +PP2C phosphoprotein phosphatase	*Mycobacterium tuberculosis* (1552573)	
DRB0143	AAA superfamily NTPase related to 5-methylcytosine-specific restriction enzyme subunit McrB	*Escherichia coli* (1790805)	Mcr operon is present only in *E. coli* and *D. radiodurans*; a clear case of horizontal gene exchange
DRB0144	Homolog of the McrC subunit of the McrBC restriction-modification system	*Escherichia coli* (1790804)	

[a] A gene cluster was considered a likely operon if the genes were localized on the same DNA strand and the distance between them was less than 100 bp.

tected 104 HTH domain-containing proteins that are predicted to function as transcriptional regulators. This number is close to those detected in other free-living bacteria with similar genome sizes (14); the repertoire of HTH-containing proteins identified in *Deinococcus* covers most of the diversity of prokaryotic transcriptional regulators. *Deinococcus* encodes seven members of the MerR/SoxR family of regulators (a greater number than in other characterized bacteria except *B. subtilis*), which could participate in the regulation of various stress response pathways (24, 155). Another family of predicted HTH regulators of unknown specificity that is expanded in *Deinococcus* consists of eight paralogs (e.g., DR1954); such an expansion is unprecedented in other bacteria and suggests a unique role in the regulation of a distinct set of genes.

Expansion of Specific Protein Families

Expansion of specific protein families has been observed for several complete genomes (43, 126, 194). Sometimes there is a clear relationship between the expansion of a particular protein family and the adaptation of the respective organism to its environment. Examples of such adaptive expansions include ferredoxins in autotrophic archaea (126), several families of enzymes involved in lipid degradation in *M. tuberculosis* (43), and *c*-type cytochromes in the metal-reducing bacteria *Shewanella* (148).

In the *D. radiodurans* genome, we detected several expansions, some of which appear to be related to stress response and damage control (Fig. 3). In particular, several different families of hydrolases are overrepresented compared to other sequenced genomes. These include MutT-like pyrophosphatases (Nudix), calcineurin-like phosphoesterases, lipase/epoxidase-like (α/β) hydrolases, subtilisin-like proteases, and sugar deacetylases. In addition to such specifically expanded families, several other families of hydrolases are present in *Deinococcus* in elevated numbers although they are also common in other bacteria and are not shown here. Some of these hydrolases are likely to be involved in the decomposition of damage products ("cell cleaning") under stress conditions. Independent expansions of certain families, such as α/β hydrolases in *Deinococcus* and *Mycobacterium* and subtilisin-like proteases in *Deinococcus* and *Bacillus*, are noteworthy and probably correlate with the adaptation of these organisms to the facultative or obligatory heterotrophic life-style (43, 116).

Expansion of the Nudix hydrolase protein superfamily is one of the most prominent features of the *Deinococcus* genome. The MutT protein, the prototype for this superfamily, has been identified as the central component of an antimutagenic system responsible for preventing incorporation of 8-oxo-dGTP into DNA (136). Subsequently, it has been shown that different MutT-like enzymes use a variety of substrates, and the Nudix pyrophosphohydrolases have been tentatively defined as a superfamily of "house-cleaning" enzymes that destroy potentially deleterious compounds (28). A detailed analysis of Nudix proteins in *Deinococcus* revealed five distinct multidomain proteins, in which the MutT domain is combined with other domains (Fig. 4). Orthologous proteins for three of them also exist in other bacteria. In particular, the family typified by *E. coli* YjaD contains a Zn ribbon module, which is probably involved in nucleic acid binding. Another *Deinococcus* protein contains an apparently inactivated (with the catalytic motif REXXEE missing) MutT domain combined with a TagD-like nucleotidyltransferase domain and is likely to perform a regulatory function. A second TagD-like nucleotidyltransferase from *Deinococcus* (DRA0273) is very similar, but the MutT domain has apparently eroded beyond recognition. Orthologs of a third Nudix protein, which contains an uncharacterized C-terminal domain, are present in *Streptomyces, Mycobacterium*, and *Synechocystis*. Again, in most of them, the Nudix pyrophosphohydrolase appears to be inactivated, suggesting a regulatory function.

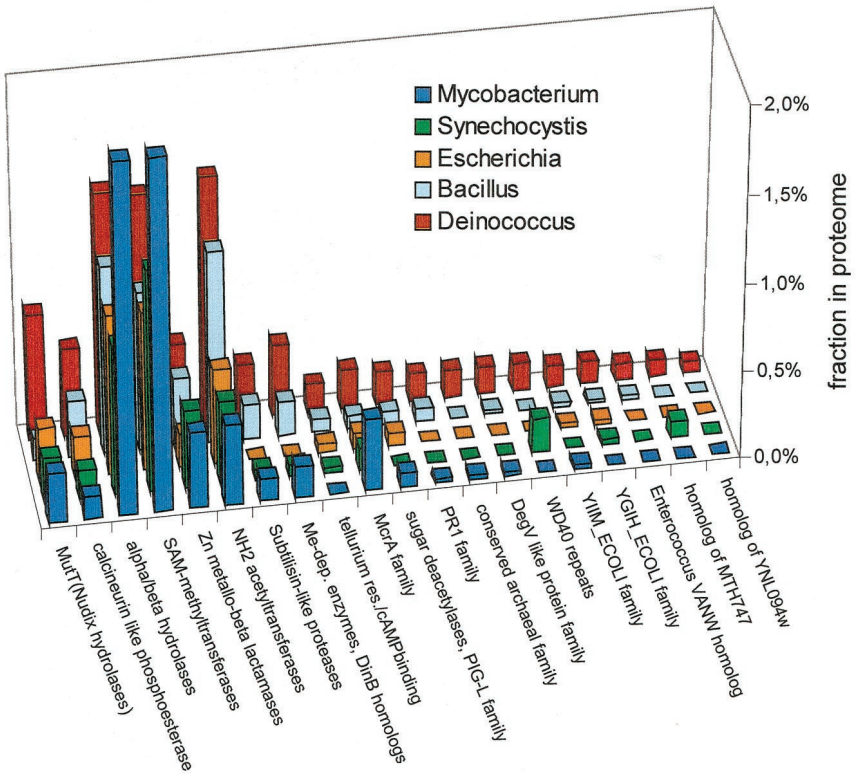

FIG. 3. Specific protein family expansion in *Deinococcus*.

Two closely related *Deinococcus* proteins contain a duplication of the MutT domain that has not yet been detected in any other organism. Three more Nudix proteins are specifically related to the proteins containing this duplication, and the genes for two of these are adjacent on the chromosome (DR0783 and DR0784). These seven related MutT domains appear to form a *Deinococcus*-specific family of Nudix hydrolases. Another Nudix protein consists of three domains, namely, *S*-adenosylmethionine (SAM)-dependent methylase, MutT, and cytosine deaminase (Fig. 4). This domain combination is unique to *Deinococcus* and suggests that the protein is involved in an as a yet uncharacterized repair pathway.

Altogether, *Deinococcus* encodes 23 Nudix superfamily proteins that contain 25 individual MutT domains. Some of these

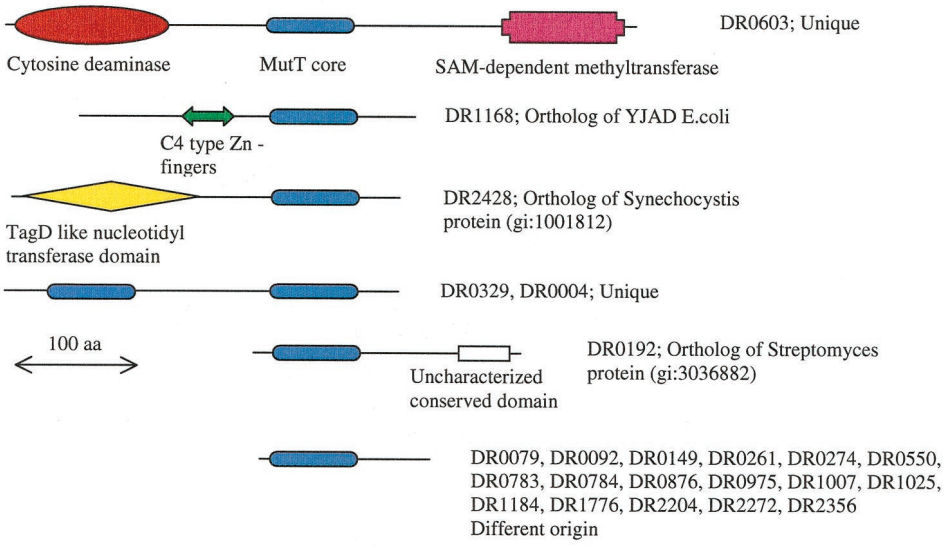

FIG. 4. Distinct domain architectures of proteins containing the MutT-like domain. aa, amino acids; SAM, *S*-adenosylmethionine.

proteins are likely to be repair enzymes with known activities, including the MutT ortholog (DR0261), while others will have novel functions, as suggested by the domain combinations discussed above. Other functions are likely to include utilization of damage products formed under various stress conditions. It is unlikely that a distant ancestor of the *Deinococcus* lineage encoded all these MutT-containing proteins. Rather, it appears that the heterogeneous collection of these proteins encoded by *D. radiodurans* was assembled via the mixed routes of serial duplication, particularly in the distinct deinococcal family of seven Nudix domains, and horizontal gene transfer.

Amino group acetyltransferases comprise another family that appears to have undergone independent expansion in *Deinococcus* and in *Bacillus*. Acetyltransferases of this type participate in various metabolic pathways, including lipid biosynthesis, and in regulatory systems. Except for *B. subtilis*, other bacteria have less than half the number of these enzymes with respect to the number found in *D. radiodurans*. Like the acetylases in other bacteria, these enzymes are likely to participate in detoxification of antibitotics and possibly of toxic products that arise upon DNA damage, as well as in regulatory protein acetylation. A *Deinococcus*-specific family of acetyltransferases, which consists of at least 11 proteins, is most similar to acetyltransferases involved in peptide antibiotic resistance, such as streptothricin acetyltransferase of *Streptomyces* (98). These acetyltransferases might aid the survival of *Deinococcus* in the presence of peptide antibiotics secreted by other bacteria, with which it has to compete for nitrogen and carbon sources as a part of its heterotrophic life-style.

Enzymes of the α/β hydrolase superfamily are mainly neutral lipases or acetyl esterases, but some of them have unusual substrate specificity, e.g., heroin esterase from *Rhodococcus* (169) and antibiotic bialaphos acetyl esterase from *Streptomyces* (167); other proteins of this superfamily possess unexpected activities, e.g., metal ion-free oxidoreductase from *Streptomyces* (91). The expanded families of α/β hydrolases in *Deinococcus* could be exploited for xenobiotic metabolism and/or the biogenesis of the complex cell envelopes (see above).

In several cases, expansion of specific subfamilies within common protein families appears to be important. *Deinococcus* encodes three paralogous proteins (DR0202, DR0494, and DR2273) related to the FlaR protein from gram-positive bacteria. One of these proteins has been shown to affect DNA topology and is osmoregulated when expressed in *E. coli* (173). It also influences the expression of supercoiling-sensitive promoters and is considered to be a chromatin-associated protein (173). Topological changes of DNA could play a role in DNA repair of *Deinococcus*, and the FlaR homologs might be involved in these processes. The FlaR subfamily belongs to the P-loop-containing kinase superfamily that includes nucleotide, gluconate, and shikimate kinases (224). *Deinococcus* encodes three paralogous proteins (DR0609, DR2467, and DR2139) that belong to another uncharacterized subfamily of these kinases which is also represented in several other bacteria.

Another interesting case is the LigT protein family, which is found in several bacteria, archaea, and eukaryotes and includes RNA ligases and predicted 2′,5′-cyclic nucleotide phosphodiesterases. In addition to the LigT ortholog (DR2339), *Deinococcus* encodes two predicted phosphodiesterases of this family (DR1000 and DR1814) that may participate in RNA metabolism or signaling.

Expansion of several other protein families is consistent with the unusual stress resistance capabilities of *D. radiodurans*. For example, *Deinococcus* encodes seven small nuclease domains related to the McrA endonuclease of *E. coli* (94). The McrA-like nuclease domain is part of three multidomain protein architectures that seem to be unique to *Deinococcus* (see below). This previously unreported propagation of McrA-like nucleases could make a contribution to the repair potential of *Deinococcus*. In evolutionary terms, the McrA domain, like the MutT domain, apparently has been expanded in *Deinococcus* through a recent duplication (DR1312 and DR2483 are 50% identical), as well as through acquisition of genes by horizontal gene transfer.

Expansion of proteins of the TerDEXZ/CABP family in *Deinococcus* is interesting because some of these proteins could confer resistance to a variety of DNA-damaging agents, including heavy-metal cations, methyl methanesulfonate, mitomycin C and UV (21, 103), and other forms of stress (11). Two members of this family, CABP1 and CABP2, are expressed during starvation in *Dictyostelium* and form a heterodimer that binds cyclic AMP (cAMP) (78), suggesting that other members of the family also bind various small-molecule ligands.

Deinococcus encodes the largest number of the pathogenesis-related 1 (PR1) family proteins (five members) among bacteria. These secreted proteins are widespread in eukaryotes but sporadic in bacteria (195); unlike the eukaryotic members of this family, the bacterial PR1-related proteins lack the disulfide bond-forming cysteines (68). Since they are predicted to be secreted, the bacterial PR1 family proteins might play a role in inhibiting extracellular enzymes or in interacting with other cells, as suggested by the known activities of their eukaryotic homologs (106).

The second largest protein expansion in *Deinococcus* is the family of uncharacterized small proteins whose prototype is *B. subtilis* DinB, a DNA damage-inducible gene product (39). Among bacteria, *Deinococcus* encodes the greatest number of these proteins, although comparable independent expansions are seen in *B. subtilis* and the actinomycetes (Fig. 3). Examination of the multiple alignment of this family (Fig. 5) reveals three conserved histidines that could form a catalytic triad of a novel metal-dependent enzyme, perhaps a hydrolase. The prediction of enzymatic activity of these proteins raises the possibility that they could be nucleases directly involved in DNA degradation, which begins in *Deinococcus* immediately after DNA damage (23, 211). This protein family may be particularly amenable to experimental studies, given its expansion in *B. subtilis*, a model for many DNA repair studies.

Several families of *Deinococcus* proteins are highly diverged and, in the initial analysis, appeared to have no homologs in other species. Database searches with individual sequences of these proteins failed to show statistically significant similarity to any proteins other than their paralogs from *Deinococcus*. Only profiles that included information on all of the paralogs (see the description of methods above) allowed the identification of homologs from other organisms. An example of such a family is a distinct group of six HTH-containing DNA-binding proteins predicted to function as transcriptional regulators.

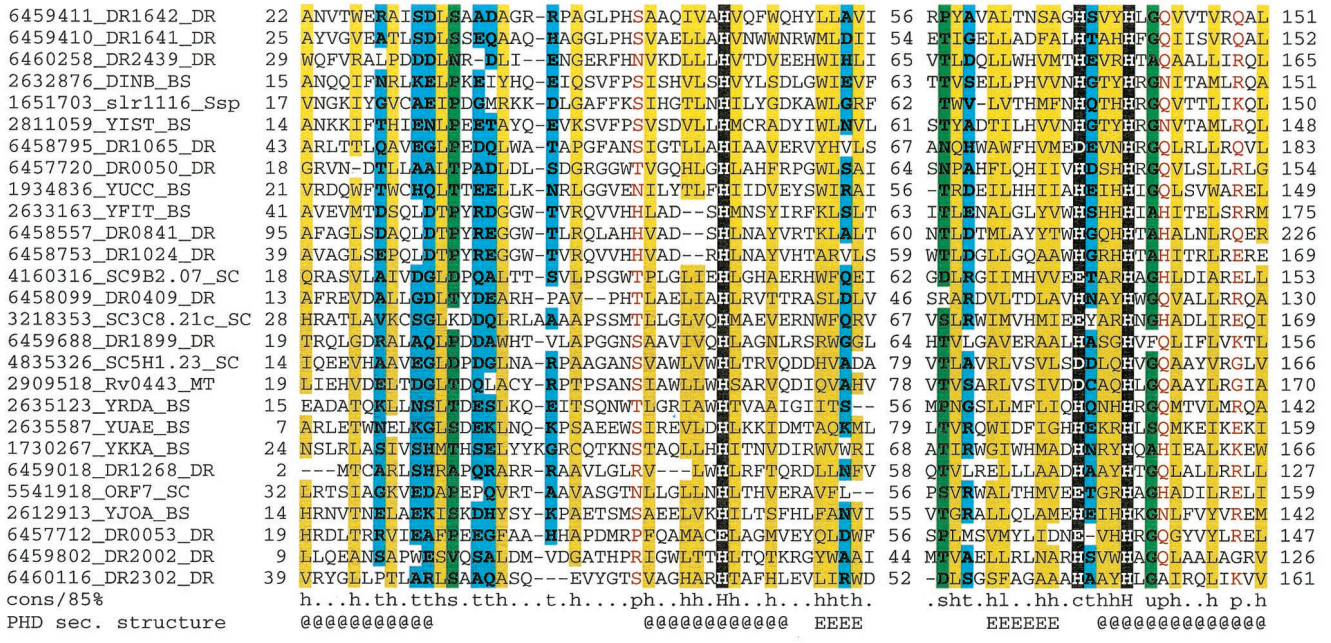

FIG. 5. Multiple alignment of the conserved core of the DinB/YfiT protein family. The alignment was generated by parsing the PSI-BLAST HSPs and realigning them with the ALITRE program (181). The numbers between aligned blocks indicate the lengths of variable inserts that are not shown; the numbers at the end of each sequence indicate the distances from the protein termini to the proximal and distal aligned blocks. The shading of conserved residues is according to the 85% consensus. The three predicted metal ligand residues are shown in inverse shading (white against a black background); Consensus sequence was obtained by a "consensus" program (http://www.bork.embl-heidelberg.de/Alignment/consensus.html) with default amino acid grouping assignments (h, s, t, p, +, etc.). The coloring of conserved position is as follows: h, hydrophobic residues (yellow background); s, small residues (bold with green background); t, turn-like residues (bold with cyan background); +, positively charged and polar (red). In front of each sequence, the GenBank identifier number (GI) and a two-letter code of species are shown. DR, *D. radiodurans*; BS, *B. subtilis*; SC, *Streptomyces coelicolor*; MT, *M. tuberculosis*, Ssp, *Synecocystis* sp.

This family is of particular interest because at least one of its members (DR0171, or IrrI [211]) appears to be associated with radiation resistance (22).

About 720 proteins encoded in the *D. radiodurans* genome have no detectable homologs in the current databases. Most of these are predicted membrane or nonglobular proteins that tend to evolve rapidly, and this impedes the detection of sequence similarity. Nevertheless, we identified 26 families with at least two members each that appear to be *Deinococcus* specific (Table 5). Some of these families have conserved sequence and structural features that, in spite of the absence of significant overall similarity to any other proteins, are reminiscent of well-characterized domains. For example, the DR2457-like and DR2241-like families contain pairs of conserved cysteines that resemble Zn ribbons present in different enzymes and nucleic acid-binding proteins. Several other uncharacterized globular proteins in *Deinococcus* (e.g., DR1088 and DR1486) also contain such cysteine pairs, which suggests metal binding or perhaps nucleic acid binding. An intriguing possibility is that, similarly to better-characterized families, these unique protein families have emerged as a result of adaptation and may be involved in novel mechanisms of DNA repair or stress response specific to *Deinococcus*.

Proteins with Unusual Domain Architectures

Combining several domains into one protein may give rise to novel protein functions, enhance the cooperation between existing functionally linked protein activities, facilitate regulation, and/or result in modification of substrate specificity (74, 128, 192). Thus, it seems reasonable to assume that, like expansion of paralogous families, unique domain architectures are lineage-specific adaptations to a particular life-style. The *D. radiodurans* genome encodes over 20 multidomain proteins with unusual domain combinations that have not been detected in other species (Fig. 1). The two phenomena appear to be linked since there are several examples where the unusual domain architectures are present in members of expanded protein families. The unique combination of the Nudix hydrolase domain with a methyltransferase and a cytosine deaminase has already been described. The McrA-like nuclease domain is a part of three unusual domain arrangements, at least two of which are suggestive of repair functions (Fig. 1). A particularly good example of such a functionally interpretable association is the DRA0131 protein, where the endonuclease domain is combined with a RAD25-like helicase, which in eukaryotes is involved in nucleotide excision repair of UV-damaged DNA (84, 196). In DR1533, an McrA-like endonuclease is linked to a SAD domain, which so far has been detected only in eukaryotic chromatin-associated proteins (71). A third protein, DRA0057, also contains a domain of the TerDEXZ/CABP family (see above) and is likely to be related to stress response.

One of the *Deinococcus* α/β hydrolases is fused to a flavin-containing monooxygenase domain, also a unique domain configuration (Fig. 1). The well-established role of flavin-contain-

TABLE 5. Unique protein families in *D. radiodurans*

ORFs in family	Range of identity (%)	Approx length[a]	Sequence features and comments
DRA0346, DRB0145	33	400	α/β proteins
DR1261, DR1348	31	80	DQE/H-rich proteins, predominantly β-strand proteins; present in *Caulobacter crescentus* unfinished genome
DR1022, DR2185	43	150	Predominantly α-helix proteins; N-terminal domain in DR1022 (C-terminal domain is a MazG-like protein, related to phosphoribosyl-ATP pyrophosphatase)
DR0082, DR2593, DR1748	31–35	160	Repetitive sequences (GRhGG repeats); coiled-coil
DR2532, DR2457	43	120	α/β proteins, tryptophan-rich
DR0871, DR1920, DR2360	36–44	120	Membrane proteins; CXPXXXC motif; DR0871 has duplication of the domain
DR1814, DR1000	30	150	Predominantly α-helical proteins
DR2179, DR1611	71	150	Possible recent duplication; α/β proteins
DR1251, DR1319, DR1545	26–31	180	Secreted α/β proteins with a single conserved cysteine
DR1530, DR0419	43	130	Predominantly α-helical proteins; contain a glycine-rich loop
DRA0012, DR2241	43	450	A 90-amino-acid N-terminal repeat; contain CXXC and CXXXC motifs; predominantly α-helical and coiled-coil proteins
DR0481, DR1195, DR1301	31–44	170	Predominantly α-helical proteins; some have transmembrane segments
DR0387 (DR2038 + DR2039)	38–46	260	Predominantly α-helical proteins

[a] Number of amino acids.

ing monooxygenases in xenobiotic transformation and oxygen reactivity (177) strengthens the hypothesis that the two domains function together in the metabolism of some environmental compound or secondary metabolite. There are other such fusions that point to potential novel metabolic functions. For example, the DRA0304 protein contains a metallo-β-lactamase-like domain fused to a C-terminal rhodanese-like domain. Proteins that consist of a single rhodanese-like domain are involved in different forms of stress response. For example, *E. coli* PspE (phage shock protein E) is induced in response to heat, ethanol, osmotic shock, and phage infection; din1 and sen1 proteins from plants are dark inducible and senescence associated, and the 67B2 protein of *Drosophila* is also heat shock inducible (96, 109). Proteins with the same domain composition but with the order of the domains reversed are encoded in the gas vesicle plasmid of the archaeon *Halobacterium halobium* (154) (GenBank ID number [GI], 2822321 and 2822327), which suggests that the hydrolase domain and rhodanese-like domain cooperate in their chaperone or metabolic functions. Another unique domain fusion (DR1207) with a possible role in the metabolism of some amino group-containing compounds includes a cytosine deaminase domain and a PP-loop ATPase similar to the cell cycle protein MesJ.

DRB0098 contains a phosphatase domain and a polynucleotide kinase domain and is another example of an independent origin of a multidomain proteine with analogous domain architectures in distant taxa. Proteins combining these (predicted) enzymatic activities have been found only in *Deinococcus*, bacteriophage T4, and some eukaryotes, including humans, *Caenorhabditis elegans*, and *Schizosaccharomyces pombe*. The phosphatase domain of the phage T4 and eukaryotic proteins belongs to the haloacid dehalogenase superfamily (12, 110), whereas the one from *Deinococcus* belongs to the HD hydrolase superfamily (15). By analogy to the eukaryotic proteins that function in DNA repair following ionizing radiation and oxidative damage (102), the deinococcal enzyme may be implicated in a similar process.

Most of the HTH-containing proteins predicted to function as transcriptional regulators in *Deinococcus* share the domain architecture with their bacterial and archaeal homologs. However, one of these proteins, DR2199, has an unusual combination of domains (Fig. 1). In addition to a C-terminal HTH domain, this protein contains (i) a distinct N-terminal domain homologous to the eukaryotic developmental regulator *schlafen* (180) and to several uncharacterized bacterial and archaeal proteins and (ii) another, uncharacterized domain shared with several bacterial and archaeal proteins. The unusual domain architecture of DR2199 is conserved in two proteins from the archaeon *Pyrococcus abyssi* (GI, 5459605 and 5458925).

Horizontal Gene Transfer

Numerous recent observations support the notion that horizontal gene transfer has played a major role in the evolution of bacteria and archaea (18, 57, 153). *Deinococcus* is no exception to this trend since it apparently acquired a significant number of genes by horizontal transfer from various sources. The most notable of these genes are listed in Table 6. Several genes found in *Deinococcus* previously have been detected only in eukaryotes and/or archaea. One of these encodes topoisomerase IB, an enzyme that is highly characteristic of eukaryotes and is present in *D. radiodurans* in addition to the typical bacterial topoisomerases IA and II. The recent demonstration of a structural and mechanistic relationship between topoisomerase IB and site-specific recombinases (38) makes a role in recombination plausible for the *D. radiodurans* enzyme. A knockout mutant with this gene deleted is substantially more sensitive to UV (254 nm) but not to ionizing radiation than is the wild type (Daly et al., unpublished). Notably, in the *Deinococcus* genome, the gene for topoisomerase IB is adjacent to a gene that encodes uracil-DNA glycosylase with a clear eukaryotic phylogenetic affinity. It appears likely that the two genes were simultaneously transferred from a eukaryotic source, possibly a large DNA virus because both enzymes are encoded by poxviruses (182), although a virus in which these genes were adjacent has not yet been detected.

TABLE 6. Examples of horizontally transferred genes in *D. radiodurans*

Protein	Gene name	Taxons where homologs are found	Best BLAST hit: species, gene identifier, and e-value	Comments
Topoisomerase IB	DR0690	*Eucarya* and double-stranded DNA viruses	Orf virus, gil521138, 2×10^{-11}	Belongs to eukaryotic type I topoisomerases; performs ATP-independent breakage of single-stranded DNA, followed by passage and rejoining; the first finding of a topoisomerase of this family in bacteria
Yellow protein (*Drosophila*) or royal jelly protein (honeybee)	DR1790	*Insecta*	*Drosophila subobscura*, gil2222667, 1×10^{-14}	Required for cuticular pigmentation in *Drosophila* and important component of royal jelly of honeybee
Acyl coenzyme A-binding protein (ACBP)	DR0166	*Eucarya*	*Caenorhabditis elegans*, gil2088729, 2×10^{-17}	Binds medium- and long-chain acyl coenzyme A esters with very high affinity
Ro RNA-binding protein	DR1262	*Eucarya*	*Xenopus laevis*, gil1173109, 4×10^{-86}	Ribonucleoproteins complexed with several small RNA molecules; involved in UV-resistance in *Deinococcus*
LEA14-like desiccation-induced protein	DR1372	*Plantae* and *Archaea*	*Lycopersicon esculentum*, gil1684830, 1×10^{-3}	Protein induced in leaves by desiccation, ethylene, or abscisic acid
Desiccation-induced protein	DRB0118	*Craterostigma plantagineum* (plants)	*Craterostigma plantagineum*, gil118926, 4×10^{-19}	Protein induced in leaves by desiccation or abscisic acid
LEA76/LEA26-like desiccation-induced protein	DR1172	*Eucarya* (mostly plants)	*C. elegans*, gil2353333, 2×10^{-26}	In plants, protein induced in leaves by desiccation, ethylene, or abscisic acid
Protein kinase of RIO1 family	DR2209	*Eucarya* and *Archaea*	*Schizosaccharomyces pombe*, gil2661615, 1×10^{-12}	Protein kinase SudD, RIO1 family member, is a suppressor of *bimD* genes, which are involved in cell cycle control in *Emericella nidulans*
Peroxidase	DRA0145	*Polyporaceae* spp. (fungi)	*Polyporaceae* spp., gil2160705, 4×10^{-34}	
Tryptophan-2,3-dioxygenase	DRA0339	*Eucarya*	*Drosophila simulans*, gil881370, 5×10^{-18}	Converts L-tryptophan to L-formylkynurenine; binds heme; can utilize other substrates
L-Kynurenine hydrolase	DRA0338	Orthologs only in *Eucarya*	*Saccharomyces cerevisiae*, gil1532216, 4×10^{-37}	Belongs to pyridoxal-dependent aminotransferase family; hydrolyzes L-kynurenine to anthranilate and L-alanine
Serine carboxypeptidase	DR0964	*Eucarya*	*Homo sapiens*, >gil2098347, 9×10^{-7}	
Tungsten formylmethanofuran dehydrogenase, subunit E (FwdE)	DRA0267	*Archaea*	*Pyrococcus horikoshii*, >gil3257655, 2×10^{-13}	Involved in methanogenesis; operon encoding all subunits of this enzyme contains six genes, *fwdEFACDB*, most of which are absent in this genome
Homolog of a tymocyte protein cThy28kD	DR0566	*Eucarya, Archaea*, and cyanobacteria	*Synechocystis* spp., gil1653325, 2×10^{-28}	Bacterial proteins show significantly greater similarity to each other and to eukaryotic homologs than to archaeal homologs, which suggests horizontal transfer between bacteria and eukaryotes
Uncharacterized protein	DR0376	Cyanobacteria and *Aquifecales*	*Synechocystis* spp., gil2708801, 4×10^{-44}	Probable enzymatic domain with a conserved glutamate; *Synechocystis* encodes at least 35 proteins of this family; *Deinococcus* has 3 of them

Another typical eukaryotic protein encoded by *Deinococcus* is a highly conserved ortholog of the eukaryotic RNA-binding protein Ro. This protein has a distinct RNA-binding domain that is shared with the RNA-binding subunits of eukaryotic telomerases such as TP-1 and p80. In eukaryotes, Ro binds specific small RNA molecules (Y RNAs) of ribonucleoprotein particles that are found both in the cytoplasm and in the nucleus (79) and has been proposed to play a role in the "quality control" of large-scale 5S rRNA biosynthesis (79). In *Deinococcus*, the Ro ortholog is involved in the regulation of UV repair, in which the eukaryote-type topoisomerase IB is believed to participate. It binds to several small RNAs analogous to the Y-RNAs that are encoded by genes upstream of the Ro gene (37). Interestingly, an independent transfer of

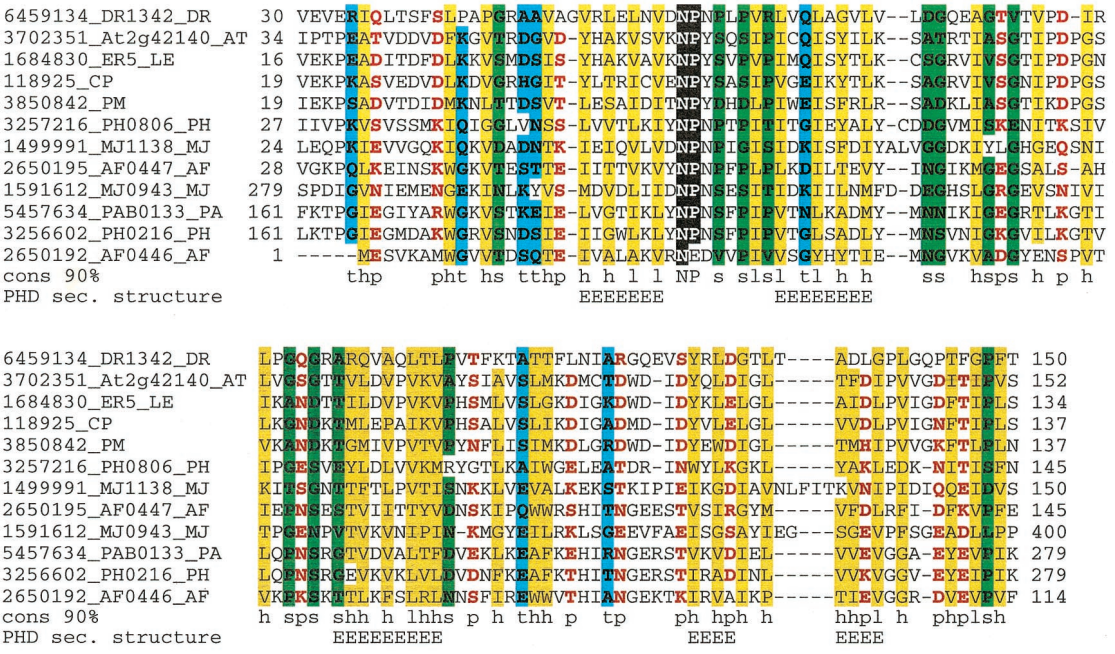

FIG. 6. Multiple alignment of the selected members of the LEA14 family of desiccation related proteins. The numbers and coloring in this alignment are the same as in Fig. 5. The two-letter code for species is as follows: AF, *Archaeaoglobus fulgidus*; MJ, *Methanococcus jannaschii*; PH, *Pyrococcus horikoshii*; PA, *Pyrococcus abyssi*; PH, *Pyrococcus furiosus*; AT, *Arabidopsis thaliana*; LE, *Lycopersicon esculentum*; CP, *Craterostigma plantagineum*; PM, *Pseudotsuga menziesii*.

another eukaryotic member of this family, related to the telomerase RNA-binding subunit, into the genome of *Streptomyces* suggests a more widespread accquisition of Ro-like RNA-binding proteins by bacteria (L. Aravind, unpublished data).

The gene for a predicted protein kinase of the RIO1 family, which previously has been detected in archaea and eukaryotes but not in bacteria (121), also appears to have been transferred into the genome of *Deinococcus*.

Four *Deinococcus* proteins whose plant homologs are induced by desiccation are of particular interest; this is the first report of bacterial homologs of plant desiccation resistance-associated proteins. The DR1372 protein belongs to the Lea-14 (late embryogenesis abundant) family of group 4 of LEA proteins, one of the best-studied plant desiccation response-associated protein families (73, 124, 225, 226). Using iterative database searching, we detected additional homologs of LEA14-like proteins in many archaeal species (Fig. 6). In plants, these proteins are cytosolic. However, DR1372 and some of the archaeal homologs contain a signal peptide, and this suggests that in *Deinococcus* and in archaea, the subcellular localization of these proteins could be different.

The Lea-76 family belongs to group 3 of LEA proteins, which also are well-characterized and widespread desiccation-induced proteins in plants (46, 58, 99, 140). The main sequence feature in these proteins is a tandem repeat of a distinct 11-mer motif, in which the amino acids at positions 1, 2, 5, and 9 are nonpolar and the rest are charged or amide residues (e.g., AAQKTKDYASD in the Lea-76 protein from soybean; GI, 421875) (58). Besides plants, at least two proteins of this family are present in the nematode *C. elegans* (GI, 2353333 and 3924824). This motif is conserved in two *Deinococcus* proteins, DR0105 and DR1172, which show significant similarity to Lea-76 proteins. More generally, several other families of late embryogenesis-abundant and/or water stress resistance-related proteins are rich in repeats and/or have biased amino acid composition (62, 97, 185), complicating the identification of homologs. Therefore, it is possible that some as yet uncharacterized *Deinococcus* proteins containing compositionally biased sequences are also relevant to desiccation resistance. The DRB0118 protein is a homolog of a desiccation-related protein from *Craterostigma plantagineum* (GI, 118622), an extremely desiccation-resistant plant from the Asteridae class. In this plant, several water stress response proteins have been identified (161), with the protein homologous to DRB0118 being the only one that has no homologs in other plants. A positive correlation between resistance to desiccation and radioresistance has recently been established by examining a series of *D. radiodurans* radiosensitive mutants for desiccation resistance (135). It is possible, therefore, that the homologs of plant desiccation resistance-associated genes have been acquired by *Deinococcus* via horizontal gene transfer; the products of these genes may be generally important to the resistance phenotype (135).

Other apparent horizontal transfers to *Deinococcus* from eukaryotes are not easily interpretable. For example, DR1790 is a highly conserved member of a protein family that includes the yellow protein of *Drosophila* and royal jelly protein from the honeybee and so far has not been detected outside the insects (4). This seems to point to a rather precise source of this horizontally transferred gene, but the biochemical function of its product is not known. Based on its role in cuticular pigmentation in *Drosophila* (111), it may be speculated that it

TABLE 7. Distribution of insertion sequences in the *D. radiodurans* genome

Name	Family	Length (bp)	Copy no. in:				Total length (bp)
			Plasmid	DR177	DR412	DR_MAIN	
IS*2621*	IS*4*	1,322	0	6	1	6	17,186
IS*2621* (5′ fragment)		25	0	1	2	4	NA
IS*4*_DR	IS*4*	1,207	4	6	0	3	15,942
IS*605*_DR	IS*605*	~1,060	0	0	0	8	8,480
TCL9	Tc1/mariner	1,048	0	1	0	4	5,250
TCL121	Tc1/mariner	1,073	0	2	0	1	3,210
TCL23	Tc1/mariner	1,069	1	1	0	1	3,207
AXL_DR	Tc1/mariner	912	1	0	0	1	1,824
IS*3*_DR	IS*3*	1,304	0	1	0	0	1,300
TNPA2_DR	TNPA	~600	0	0	0	1	600
VCL_DR	IS*15*	~500	1	0	0	0	1,500
DNIV_DR	DNA invertase	~600	1	0	0	0	600
TNPA1_DR	TNPA	~3,000	1	0	0	0	3,000
Total			9	17	1	25	62,099
No. of copies per 10,000 nucleotides			1.97	0.96	0.02	0.09	

could be an enzyme required for the metabolism of certain pigments.

Mobile Genetic Elements

The genome of *D. radiodurans* contains a number of predicted mobile elements of different classes. These are of particular interest because of the role some of them could play in recombinational repair.

Inteins. Two inteins, protein splicing elements that are typically inserted in genes involved in DNA metabolism and other nucleotide-utilizing enzymes (162), were identified in *D. radiodurans*. One of these is inserted in the ribonucleotide reductase and is similar to the inteins inserted in orthologous enzymes from *B. subtilis*, pyrococci, and chilo iridiscent virus. This intein contains an inserted cro-like HTH domain (14) followed by a homing endonuclease of the LAGLI-DAG family (47). The second intein is inserted between the P-loop motif and the Mg^{2+}-binding (Walker B) motif of a SWI2/SNF2 family ATPase, which is involved in chromatin remodeling; this is the first documented instance of an intein interrupting a protein of this family. The most unusual feature of this intein that it is encoded by two distinct adjacent ORFs (DR1258 and DR1259), each of which also encodes a portion of the ATPase split by the intein. Recently, it has been proposed and then shown experimentally that the split intein in the *Synechocystis* DNA polymerase III α-subunit assembles from the two separately translated ORFs and splices out to form a fully functional protein (66, 77, 222). A similar protein transsplicing mechanism is likely to generate an active SWI2/SNF2 ATPase in *Deinococcus*.

Insertional sequences. Insertional sequences (ISs) in the *D. radiodurans* genome were identified during the genome annotation by the presence of ORFs homologous to transposases of several different IS families (34). Several of these ORFs exist in multiple copies. For most of these elements (IS4_DR, TCL9, TCL121, TCL23, IS3_DR, and AXL_DR), the precise length could be determined. All of these elements have the typical features of ISs identified in other species (72). In particular, they contain one or two ORFs that encode a transcriptional regulator and a transposase, as well as inverted terminal repeats and/or internal repeats (data not shown). Three elements (TCL9, TCL121, and TCL23) of the Tc1-mariner family are closely related to each other and are likely to be the product of a recent duplication, probably specific to the *Deinococcus* lineage (data not shown).

Overall, we detected 52 IS elements in the *D. radiodurans* genome (Table 7). The three most abundant ISs are IS4_DR (13 copies), IS2621_DR (11 copies), and IS200_DR (8 copies). IS elements are unevenly distributed on the chromosomes and plasmids. The number of copies per 10,000 nucleotides in the plasmid and the megaplasmid is more than 10 times greater

TABLE 8. Number of repeats in bacterial genomes

Species	Genome size (Mb)	No. of IS elements	No. of SNRs
D. radiodurans	3.3	52	295
B. subtilis	4.2	0	36
E. coli	4.6	37	263
M. tuberculosis	4.4	32	252
Synechocystis spp.	3.6	NA[a]	118
A. fulgidus	2.2	13	NA

[a] NA, not applicable.

TABLE 9. Distribution of SNRs in the *D. radiodurans* genome

Name	Length (bp)	Copy no. in:			
		Plasmid	DR177	DR412	DR_MAIN
SRE	160	0	3	4	32
SNR1	139	0	0	1	39
SNR2	114	0	0	8	76
SNR4	147	0	1	2	4
SNR5	215	0	0	1	27
SNR7	140	0	2	0	14
SNR8	131	0	0	1	19
SNR9	105	0	0	1	6
SNR10	60	0	0	0	6
Total no.		0	6	18	223
No. of copies per 10,000 nucleotides		0	0.3	0.4	0.8

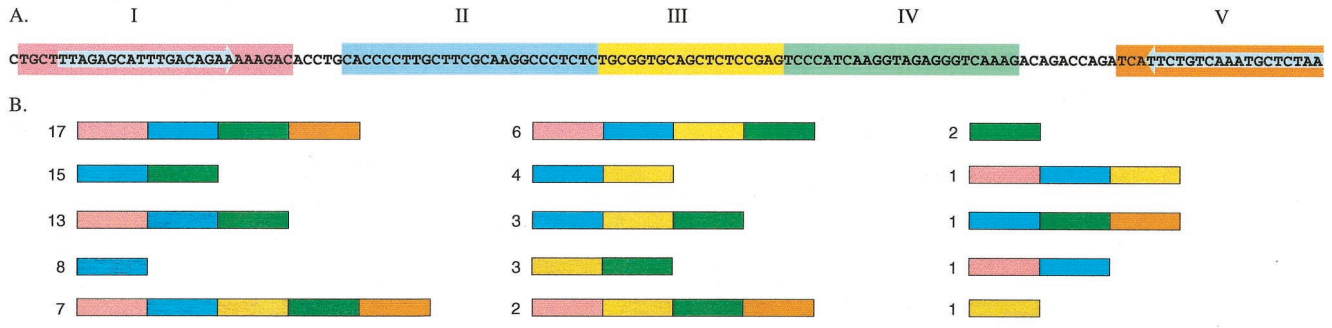

FIG. 7. (A) Structure of the full-length repeated member of the SNR2 family. Inverted repeats are marked by arrows. Roman numerals and different colors mark the five conserved modules. (B) Number of SNR2 members with the indicated modular configuration. Each class of module is represented by a different color.

than the number found in chromosomes I and II. Only one IS element is present in the chromosome II, whereas the plasmid contains nine. There are five single-copy IS elements in the *D. radiodurans* genome, three of them on the plasmid. They may be transpositionally inactive, or, alternatively, they could have been only recently acquired by the R1 strain.

Notably, IS elements are significantly more abundant in *D. radiodurans* than in any of the other sequenced bacterial genomes (Table 8). *D. radiodurans* contains 16.3 IS elements per 1,000 genes, whereas *E. coli*, ranking second, has only 8.4. If the number of IS elements is a reflection of transposition activity, this would be expected to cause genome instability and result in high levels of genome rearrangement in *Deinococcus*. There is, however, little direct evidence for any active transposition in *D. radiodurans*. In the entire genome, there is only one example of gene disruption by an IS element, where IS*2621* is inserted into the gene for alkaline serine exoprotease A (aqualysin I). Similarly, only one IS-induced mutation has been detected in *D. radiodurans* (*uvrA* [149]). Nevertheless, the abundance of IS elements in the *Deinococcus* genome is remarkable, and their involvement in genome instability is the subject of ongoing investigations.

Small noncoding repeats. We identified several families of small noncoding repeats (SNRs) in the *D. radiodurans* intergenic regions (Table 9). A comparison to other bacterial genomes showed that, like IS elements, SNRs are more abundant in *D. radiodurans* than in *E. coli* (Table 8). However, the location bias observed for IS elements appears to be reversed for SNRs. There are no SNRs in the plasmid, that contains five IS elements. In contrast, chromosome II, that contains only one IS element, has 18 SNRs.

D. radiodurans SNRs have a complex mosaic configuration, as exemplified by SNR2, that consists of five conserved modules (Fig. 7A). Module I (also shared with the small repetitive element [SRE] family) and module V contain two parts of the inverted repeat present in SNR2. The different configurations of the SNR2 family are shown (Fig. 7B). These data suggest that deletions and insertions are likely to have played an important role in the evolution of SNRs. For example, module III is likely to be missing when both modules II and IV are present.

The distribution of SNRs along *D. radiodurans* chromosome I was tested against the null hypothesis of random occurrence of an SNR in the intergenic regions. The analyses for individual families as well as SNRs together showed that, with a single exception, there is no significant deviation from the random-placement model. The exception is the SNR5 family members that show a tendency ($P < 0.05$) to occur closer to each other than predicted by the random model. There is no significant correlation between the direction of a repeat and the direction of the adjacent gene, nor an apparent relationship between a particular SNR family and the functions of the adjacent genes. Thus, SNRs are not likely to play a direct role in the regulation of transcription or translation. It should be noted in this context that while some *D. radiodurans* SNRs have characteristics similar to the *E. coli* families of small repeats (bacterial interspersed mosaic elements [BIMEs] [76]), SNRs do not share sequence or structural features with *E. coli rho*-independent transcription terminators (Ter repeats [29]). The energy of potential RNA secondary structures predicted for *D. radiodurans* SNRs does not differ from the values obtained for coding regions or other sequence fragments unrelated to SNRs.

A sequence for the SRE from the *D. radiodurans* strain SARK was published previously (120) and has provided an opportunity to compare two evolutionarily distinct but closely related SNRs. A multiple alignment of SRE sequences from

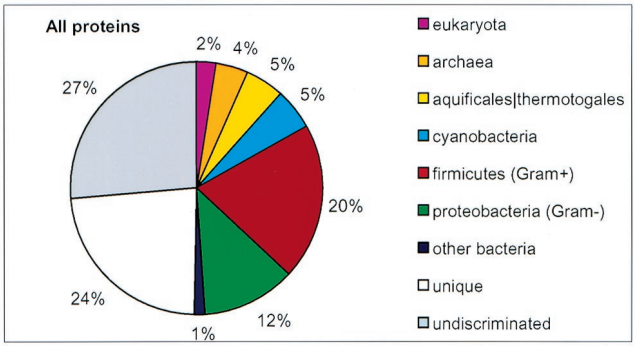

FIG. 8. Taxonomic affinities of *Deinococcus* proteins. We defined a hit to a particular lineage as the best one if it had a BLAST E-value for a protein from this lineage 100 times lower than to any protein from another lineage. Hits to *Thermus-Deinococcus* group species were disregarded.

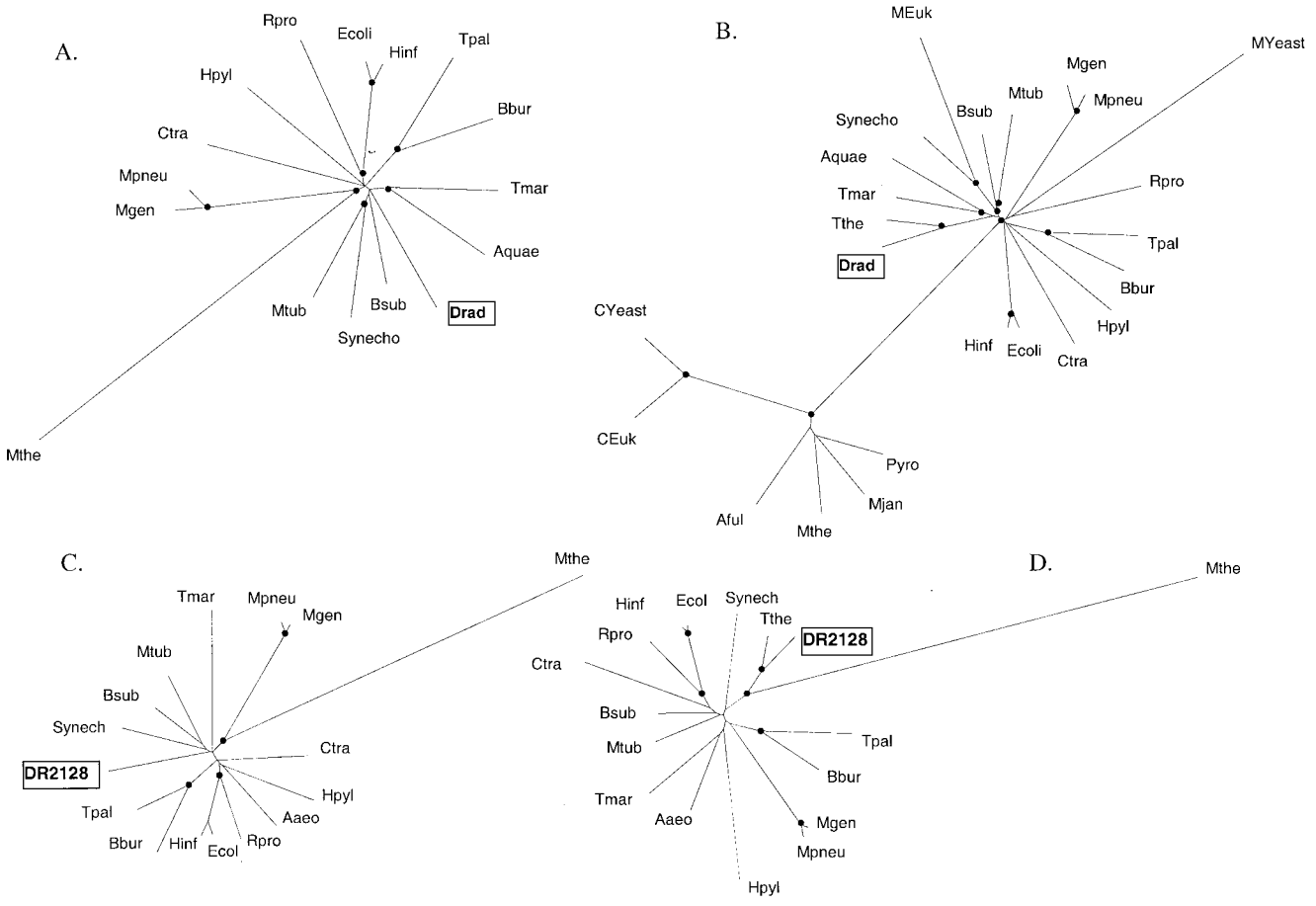

FIG. 9. Phylogenetic trees. (A) All ribosomal proteins shared by selected organisms with a completely sequenced genome; (B) all ribosomal proteins shared by *Thermus* and selected organisms; (C) RNA polymerase subunit A; (D) fragment of RNA polymerase subunit A shared by *Thermus* and selected organisms. Proteins were aligned by CLUSTALW. Alignments were checked manually, and unaligned fragments were removed. Subsequently, alignments were used for tree reconstruction using the PHYLIP program (default parameters throughout). Abbreviations of species in the trees: Mthe, *Methanobacterium thermoautotrophicum*; Bsub, *B. subtilis*; Mpneu, *Mycoplasma pneumoniae*; Mgen, *Mycoplasma genitalium*; Mtub, *Mycobacterium tuberculosis*; Ecol, *E. coli*; Hinf, *Haemophilus influenzae*; Rpro, *Rickettsia prowazekii*; Bbur, *Borrelia burgdorferi*; Tpal, *Treponema pallidum*; Hpyl, *Helicobacter pylori*; Ctra, *Chlamydia trachomatis*; Synech, *Synechocystis* sp.; Aaeo, *Aquifex aeolicus*; Tmar, *Thermotoga maritima*; The, *Thermus thermophilus*; Scer, *Saccharomyces cerevisiae*.

both strains (not shown) showed that most of the strain-specific substitutions were located in the central regions of two pairs of inverted repeats. In strain SARK, these SRE inverted repeats may form a pair of hairpin-like structures (120). However, the substitutions seen in R1 may disrupt these hairpins. The predicted free energy for the consensus hairpin I in SARK is −11.2 kcal/mol, but that in R1 is only −6.1 kcal/mol, as estimated by the Mfold program (134); for hairpin II, it is −17.0 and −11.0 kcal/mol, respectively. The nonrandom clustering of the strain-specific nucleotide substitutions in strain R1 is tantalizing. One could speculate that there is a strain-specific selection pressure for either strengthening (in SARK) or disrupting (in R1) these hairpins within this particular repeat family, and this would suggest a specific function for these repeats. The second possibility is that the multiple substitutions in the hairpin represent regions of the SRE that are hot spots for spontaneous mutagenesis.

The first SRE detected in *D. radiodurans* was within a cloned mitomycin C-inducible gene of strain SARK (120). Interestingly, a comparison with the corresponding region of the R1 strain shows that the repeat is missing in R1, demonstrating the mobility of SRE (127). The propensity of *D. radiodurans* to amplify DNA sequences that are flanked by direct repeats (31, 51, 190) is relevant to the large number of repeats, both ISs and SNRs, in its chromosomes and plasmids (4 to 10 copies per cell). The abundance of such repeated sequences flanking genes and operons throughout the genome could provide the potential for expansion and regulation of genomic regions in response to environmental challenges.

Prophages. Two prophages unrelated to one another are present in the *Deinococcus* genome. One of these is located on chromosome I (between positions 518499 and 547679), and the other is located on chromosome II (between positions 80554 and 113236). Some of the proteins encoded in these prophages are distantly related to several phage proteins from other bacteria, but most ORFs have no detectable homologs. These prophages contain some genes definitely acquired from a bacterial genome, e.g., a serine/threonine protein kinase (DR0534)

TABLE 10. *Deinococcus-Thermus* shared features

Protein name and function	(gene_ID)	T.t (GI)	Comment
Uncharacterized[a]	DR1981	2624409	Diverged HD hydrolase domain (most similar to the HD domain of GLND, a uridylyltransferase)
Uncharacterized[a]	DR1423	1872145	Homolog of the nitrogen regulatory protein P-II, GlnB; highly conserved between *Deinococcus, Thermus*, and *B. subtilis*
Uncharacterized[a]	DR0972	1781362	Secreted protein
S-layer-like protein[a]	DR0383, DR1185, DR1115, DR1124	993026	Present also in *Thermotoga maritima*
ArgB, acetylglutamate kinase[b]	DR0383	2696108	Archaeal form
ArgC, *N*-acetyl-γ-glutamyl phosphate reductase[b]	DR0963	1781360	Archaeal form
UppS, undecaprenyl diphosphate synthase[b]	DR2447	3724366	Archaeal form
IdsA, geranyl geranyl diphosphate synthase[b]	DR1395	1549220	Eukaryotic-archaeal form, bifunctional enzyme
ProC, pyrroline-5-carboxylate reductase[b]	DR1522	473555	Eukaryotic-archaeal form

[a] Proteins (nearly) unique for the *Deinococcus-Thermus* clade.
[b] "Archaeal" and "eukaryotic" proteins.

and a MotB/OMPA family protein (DR0536), and therefore are possible vectors for horizontal gene transfer.

Evolutionary Relationships to Other Bacteria and Phylogeny

A specific relationship between *Thermus* and *Deinococcus* has been established by both traditional microbiological (32, 146) and molecular phylogenetic (156) approaches. These species currently comprise a bacterial group without a clear relationship to other major branches of bacteria. Previous attempts to clarify these relationships (80) have led to the proposition that the *Thermus-Deinococcus* group is an intermediate between gram-positive and gram-negative bacteria. Furthermore, on the basis of phylogenetic trees developed for several protein families (HSP70, HSP40, FtsZ, RecA, and some translation elongation factors) and rRNA, an affinity of this group with cyanobacteria has been proposed (reference 80 and references therein).

Sequence analysis on the complete genome scale has revealed a major role of horizontal gene transfer in the evolution of bacteria and archaea. It appears that for most bacterial genomes, at least 10 to 15% of genes have been involved in horizontal transfer (18, 153; K. S. Makarova, L. Aravind, and E. V. Koonin, unpublished data). As discussed above, this level of horizontal gene transfer is consistent with our findings in *Deinococcus*. The taxonomic distribution of the best BLAST hits for all proteins in the *Deinococcus* genome is shown in Fig. 8. More than half of the genes did not show specific affinity to any major bacterial branch, archaea, or eukaryotes. Some of these genes were unique to *Deinococcus*, but the majority appear to be more or less equidistant from their homologs from other major taxa. Among the remaining genes, the greatest

TABLE 11. *Deinococcus-Thermus* differences

Protein name	*Deinococcus* gene_ID	*Thermus* GI	Comment
Absent in *D. radiodurans*			
Aspartokinase α-2		1616998	
Aspartokinase β-2		1616997	
Adenine-*N6*-DNA methyltransferase		1942357	
SAM-dependent methyltransferase		1655696	Archaeal form
Dioxygenase		281495	Specific for thermophiles, archaea, and eukaryotes
Xylose isomerase		94736	
Site-specific deoxyribonuclease		77598	
Site-specific DNA-methyltransferase		77594	
Restriction endonuclease		2665832	
IS element		217182	
		217181	
Dissimilar orthologs			
DNA polymerase III, gamma and tau subunit	DR2410	2583049	Significant differences in protein length (variation in C-terminal tail)
DNA polymerase X	DR0467	1526547	*D. radiodurans* contains an additional PHP domain, which is present also in *B. subtilis* and *M. thermoautotrophicum*; the polymerase domain in *Deinococcus* appears to be inactivated
Acetolactate synthase, large subunit	DR1516	1311482	The similarity between the *Deinococcus* and *Thermus* proteins is low compared to the similarity between each of them and orthologs from other bacteria

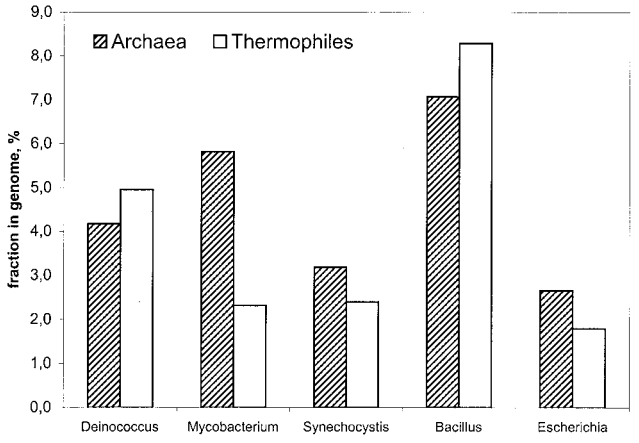

FIG. 10. Comparison of shared "thermophilic" genes in different species.

fraction was most similar to homologs from gram-positive bacteria (Fig. 8), but even in this case it is difficult to distinguish a genuine phylogenetic signal from preferential horizontal gene transfer. Therefore, this form of analysis does not yield a specific phylogenetic placement for the *Thermus-Deinococcus* group.

For phylogenetic reconstruction, we used a nearly complete set of ribosomal proteins (50 sequences) and three RNA polymerase subunits that are shared among all bacterial species. A slightly smaller protein set was used to additionally include *Thermus aquaticus* (Fig. 9). All of these proteins are subunits of large, coevolving, macromolecular complexes, and therefore the respective genes are less prone to horizontal transfer. Furthermore, the large amount of sequence information included in this analysis helped to minimize the effects of possible horizontal transfer events or fluctuations in the evolutionary rate that could affect the tree topology for individual protein families. Tree A and tree C (Fig. 9) have essentially the same topology, indicating that the *Thermus-Deinococcus* group is a deeply rooted bacterial branch with a marginal, but not necessarily reliable, affinity to the cluster of gram-positive bacteria, cyanobacteria, and bacterial thermophiles. Tree B and tree D clearly confirm the strong relationship between *Thermus* and *Deinococcus*. These analyses did not detect any evidence for the previously suggested specific relationship between the *Thermus-Deinococcus* group and cyanobacteria.

Derived shared characteristics between *Thermus* and *Deinococcus* were used for a preliminary assessment of the possible genome organization and physiological features of their common ancestor. A comparison of all available protein sequences from *Thermus* to those encoded in the *Deinococcus* genome showed several features that are unique to this clade (Table 10). The conservation of a distinct S-layer-like protein in the *Thermus-Deinococcus* group suggests that the last common ancestor already possessed the unique membrane structure observed in both organisms (146). Another shared protein unique to these organisms contains a predicted signal peptide that could be involved in the formation of the characteristic infrastructure of their outer membranes. Further, the conservation of two proteins that are distantly related to nitrogen metabolism regulators suggests that a derived state of this system evolved before the divergence of *Thermus* and *Deinococcus* (Table 10) (33). Several proteins that are highly conserved in *Thermus* and *Deinococcus* show a clear affinity to archaea and/or eukaryotes, and this may have arisen by ancient horizontal gene transfer events.

We also estimated gene flow and gene loss rates in these moderately related bacteria from the same clade. *Deinococcus* has twice as many genes as *Thermus* (http://www.nlm.nih.gov:80/PMGifs/Genomes/bact.html), yet we found a significant number of genes that are present in *Thermus* but not in *Deinococcus* (Table 11). This probably reflects the distinct metabolic repertoires of these bacteria, as well as the presence in *Thermus* of genes associated with thermophilicity.

The phylogenetic affinity between *Thermus* and *Deinococcus* raises the issue of whether their common ancestor was a thermophile. We compared the fractions of genes shared by archaeal and bacterial thermophiles for all bacteria with completely sequenced large genomes. Perhaps not unexpectedly, the greatest fraction was seen in *B. subtilis*, because many species of the *Bacillus-Clostridium* group are thermophiles and *Thermotoga* may be a highly derived member of this group; *Deinococcus* had the second greatest fraction (Fig. 10). The number of common genes between these thermophiles and *Deinococcus* is consistent with the hypothesis that the ancestor of the *Thermus-Deinococcus* group also was at least a moderate thermophile with the descendent clades evolving in different directions and accquiring different sets of genes via horizontal transfer. The complete genome sequence of *Thermus* and, ideally, other members of this clade would be required for a definitive evaluation of this hypothesis.

CONCLUSIONS

The analysis of the *D. radiodurans* genome resulted in the identification and preliminary characterization of a number of unusual features. For example, the expanded Nudix hydrolase superfamily and the homologs of plant desiccation resistance-associated proteins are likely to contribute to both the extreme radiation and the desiccation resistance of *Deinococcus*. A variety of other proteins, particularly those that belong to expanded families, are likely to be involved in the unusual phenotype of this bacterium. Furthermore, the unexpectedly numerous nucleotide repeats may also play a role in stress response. The genome analysis yielded many functional predictions that can be tested experimentally and that could prove particularly significant if considered in an evolutionary context. For example, knockouts of the typically eukaryotic genes for TopoIB and Ro protein that were identified in *D. radiodurans* were generated, and preliminary data were obtained on the DNA repair capabilities and resistance phenotypes of the mutants (37; unpublished observations). In addition to detecting a variety of single horizontal gene transfer events, there is evidence for transfer of entire gene systems. For example, we identified several *Deinococcus* genes encoding pilus-associated functions: pilus biogenesis regulation operon, several pilins and prepilins, prepilin peptidase, PilT ATPase, and the fimbrial assembly protein PilM. Remarkably, there is no experimental evidence that *D. radiodurans* is capable of producing any pili, but it seems likely that the products of these genes contribute to the formation of other surface structures, espe-

cially those that could be involved in secretory systems similar to the type III secretion pathway. As illustrated repeatedly, the genome promises to open up many new areas for experimental work, and these are likely to further expand as genomes of other species from the same clade are sequenced and analyzed.

The sobering conclusion from this study is that the fundamental questions underlying the extreme resistance phenotype of *D. radiodurans* remain unanswered. It seems most likely that this phenotype is very complex and is determined collectively by some of the features revealed by this genome analysis, as well as by many more subtle structural peculiarities of proteins and DNA that are not readily inferred from the sequences, at least not with the current limited collection of genomes available for comparative analysis. This is parallel to the results of comparative analysis of the genomes of archaeal and bacterial thermophiles, which provided many tantalizing clues in terms of genes that are shared by these organisms, to the exclusion of mesophiles, and their possible functions but so far have failed to establish an unequivocal molecular basis for thermophilicity (18, 126, 153). We expect that a comprehensive understanding of the mechanisms of damage repair in *Deinococcus* will arise from a combination of further comparative genomic analysis and prediction-driven experiments.

AVAILABILITY OF COMPLETE RESULTS

The annotation of *D. radiodurans* protein-coding genes is available at ftp://ncbi.nlm.nih.gov/pub/koonin/Deinococcus/.

ACKNOWLEDGMENTS

This research was funded largely by grant DE-FG02-98ER62583 from the Microbial Genome Program, Office of Biological and Environmental Research, Department of Energy (DOE), and grant 5R01-GM39933-09 from the National Institutes of Health. Some of this work was also supported by grants FG02-98ER62492 and FG02-97ER20293 from the DOE.

We are grateful to John Battista (Louisiana State University) and Owen White (The Institute for Genomic Research) for numerous helpful discussions and critical reading of the manuscript.

ADDENDUM IN PROOF

After the manuscript was submitted for publication, we became aware of several recent findings that provide new insights into *Deinococcus* gene functions. In particular, an alternative, α-animoadipate pathway of lysine biosynthesis (in contrast to the diaminopimelate pathway, which is typical of most other bacteria) was discovered in *Thermus thermophilus* (N. Kobashi, M. Nishiyama, and M. Tanokura, J. Bacteriol. 181:1713–1718, 1999). As described above, *Deinococcus* can grow on minimal media without lysine, and it now appears most likely that it also produces lysine via the α-animoadipate pathway. The following *Deinococcus* genes are orthologs of the *Thermus* genes encoding enzymes of this pathway: DR1238 (homocitrate synthase), DR1610 or DR1778 (large subunit of 3-isopropylmalate dehydratase), DR1784 and DR1614 (small subunit of 3-isopropylmalate dehydratase), DR1674 (isocitrate dehydrogenase), and DR2194 (glutaminyl transferase); the pathway also could include additional, still unidentified enzymes. However, in *Deinococcus* these genes do not form a cluster as in *T. thermophilus* and *Pyrococcus horokoshii* (N. Nishida, M. Nishiyama, N. Kobashi, T. Kosuge, T. Hoshino, and H. Yamane, Genome Res. 9:1175–1183, 1999). All 21 Nudix hydrolase genes from *Deinococcus* were cloned, and some novel enzymatic activities (UDP-glucose pyrophosphatase and CoA pyrophosphatase) were identified (W. Xu, J. Shen, C. A. Dunn, S. Desai, and M. Bessman, Mol. Microbiol. 39:286–290, 2001). The *mglB*-like genes that are expanded in *Deinococcus* belong to a protein superfamily that also includes dynein light chains of the Roadblock/LC7 class; together with Ras/Rho GTPases, they form a regulatory module which might be involved in the control of some molecular motors of the cell (E. V. Koonin and L. Aravind, Curr. Biol. 10:R774–R776, 2000).

REFERENCES

1. **Agostini, H. J., J. D. Carroll, and K. W. Minton.** 1996. Identification and characterization of *uvrA*, a DNA repair gene of *Deinococcus radiodurans*. J. Bacteriol. **178:**6759–6765.
2. **Aizenman, E., H. Engelberg-Kulka, and G. Glaser.** 1996. An *Escherichia coli* chromosomal "addiction module" regulated by guanosine [corrected] 3′,5′-bispyrophosphate: a model for programmed bacterial cell death. Proc. Natl. Acad. Sci. USA **93:**6059–6063. (Erratum, **93:**9991.)
3. **Al-Bakri, G. H., M. W. Mackay, P. A. Whittaker, and B. E. Moseley.** 1985. Cloning of the DNA repair genes mtcA, mtcB, uvsC, uvsD, uvsE and the leuB gene from *Deinococcus radiodurans*. Gene **33:**305–311.
4. **Albert, S., D. Bhattacharya, J. Klaudiny, J. Schmitzova, and J. Simuth.** 1999. The family of major royal jelly proteins and its evolution. J. Mol. Evol. **49:**290–297.
5. **Altschul, S. F., and E. V. Koonin.** 1998. Iterated profile searches with PSI-BLAST–a tool for discovery in protein databases. Trends Biochem. Sci. **23:**444–447.
6. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
7. **Anderson, A., H. Nordan, R. Cain, G. Parrish, and D. Duggan.** 1956. Studies on a radoresistant micrococcus. I. Isolation, morphology, cultural charchteristoics, and resistance to gamma radiation. Food Technol. **10:**575–578.
8. **Anderson, R.** 1983. Alkylamines: novel lipid constituents in *Deinococcus radiodurans*. Biochim. Biophys. Acta **753:**266–268.
9. **Anderson, R., and K. Hansen.** 1985. Structure of a novel phosphoglycolipid from *Deinococcus radiodurans*. J. Biol. Chem. **260:**12219–12223.
10. **Andersson, A. M., N. Weiss, F. Rainey, and M. S. Salkinoja-Salonen.** 1999. Dust-borne bacteria in animal sheds, schools and children's day care centres. J. Appl. Microbiol. **86:**622–634.
11. **Antelmann, H., J. Bernhardt, R. Schmid, H. Mach, U. Volker, and M. Hecker.** 1997. First steps from a two-dimensional protein index towards a response- regulation map for *Bacillus subtilis*. Electrophoresis **18:**1451–1463.
12. **Aravind, L., M. Y. Galperin, and E. V. Koonin.** 1998. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. Trends Biochem. Sci. **23:**127–129.
13. **Aravind, L., and E. V. Koonin.** 1999. DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. Nucleic Acids Res. **27:**1609–1618.
14. **Aravind, L., and E. V. Koonin.** 1999. DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res. **27:**4658–4670.
15. **Aravind, L., and E. V. Koonin.** 1998. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem Sci. **23:**469–472.
16. **Aravind, L., and C. P. Ponting.** 1999. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol Lett. **176:**111–116.
17. **Aravind, L., and C. P. Ponting.** 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem. Sci. **22:**458–459.
18. **Aravind, L., R. L. Tatusov, Y. I. Wolf, D. R. Walker, and E. V. Kunin.** 1998. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. Trends Genet. **14:**442–444. (Erratum, **15:**41.)
19. **Reference deleted.**
20. **Aravind, L., D. R. Walker, and E. V. Koonin.** 1999. Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Res. **27:**1223–1242.
21. **Azeddoug, H., and G. Reysset.** 1994. Cloning and sequencing of a chromosomal fragment from *Clostridium acetobutylicum* strain ABKn8 conferring chemical-damaging agents and UV resistance to *E. coli* recA strains. Curr. Microbiol. **29:**229–235.
22. **Battista, J. R.** 1997. Against all odds: the survival strategies of *Deinococcus radiodurans*. Annu. Rev. Microbiol. **51:**203–224.
23. **Battista, J. R., A. M. Earl, and M. J. Park.** 1999. Why is *Deinococcus*

radiodurans so resistant to ionizing radiation? Trends Microbiol. **7:**362–365.
24. **Bauer, C. E., S. Elsen, and T. H. Bird.** 1999. Mechanisms for redox control of gene expression. Annu. Rev. Microbiol. **53:**495–523.
25. **Baumeister, W., M. Barth, R. Hegerl, R. Guckenberger, M. Hahn, and W. O. Saxton.** 1986. Three-dimensional structure of the regular surface layer (HPI layer) of *Deinococcus radiodurans*. J. Mol. Biol. **187:**241–250.
26. **Baumeister, W., O. Kubler, and H. P. Zingsheim.** 1981. The structure of the cell envelope of *Micrococcus radiodurans* as revealed by metal shadowing and decoration. J. Ultrastruct. Res. **75:**60–71.
27. **Beard, W. A., and S. H. Wilson.** 1995. Purification and domain-mapping of mammalian DNA polymerase beta. Methods Enzymol. **262:**98–107.
28. **Bessman, M. J., D. N. Frick, and S. F. O'Handley.** 1996. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. J. Biol. Chem. **271:**25059–25062.
29. **Blaisdell, B. E., K. E. Rudd, A. Matin, and S. Karlin.** 1993. Significant dispersed recurrent DNA sequences in the *Escherichia coli* genome. Several new groups. J. Mol. Biol. **229:**833–848.
30. **Bolling, M. E., and J. K. Setlow.** 1966. The resistance of *Micrococcus radiodurans* to ultraviolet radiation. III. A repair mechanism. Biochim. Biophys. Acta **123:**26–33.
31. **Brim, H., S. C. McFarlan, J. K. Fredrickson, K. W. Minton, M. Zhai, L. P. Wackett, and M. J. Daly.** 2000. Engineering *Deinococcus radiodurans* for metal remediation in radioactive mixed waste environments. Nat. Biotechnol. **18:**85–90.
32. **Brooks, B. W., R. G. E. Murray, J. L. Johnson, E. Stackebrandt, C. R. Woese, and G. E. Fox.** 1980. Red-pigmented micrococci: a basis for taxonomy. Int. J. Syst. Bacteriol. **30:**627–646.
33. **Bueno, R., G. Pahel, and B. Magasanik.** 1985. Role of *glnB* and *glnD* gene products in regulation of the *glnALG* operon of *Escherichia coli*. J. Bacteriol. **164:**816–822.
34. **Capy, P., T. Langin, D. Higuet, P. Maurer, and C. Bazin.** 1997. Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? Genetica **100:**63–72.
35. **Carbonneau, M. A., A. M. Melin, A. Perromat, and M. Clerc.** 1989. The action of free radicals on *Deinococcus radiodurans* carotenoids. Arch. Biochem. Biophys. **275:**244–251.
36. **Carroll, J. D., M. J. Daly, and K. W. Minton.** 1996. Expression of *recA* in *Deinococcus radiodurans*. J. Bacteriol. **178:**130–135.
37. **Chen, X., A. M. Quinn, and S. L. Wolin.** 2000. Ro ribonucleoproteins contribute to the resistance of deinococcus radiodurans to ultraviolet irradiation. Genes Dev. **14:**777–782.
38. **Cheng, C., P. Kussie, N. Pavletich, and S. Shuman.** 1998. Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. Cell **92:**841–850.
39. **Cheo, D. L., K. W. Bayles, and R. E. Yasbin.** 1991. Cloning and characterization of DNA damage-inducible promoter regions from *Bacillus subtilis* J. Bacteriol. **173:**1696–1703.
40. **Chervitz, S. A., L. Aravind, G. Sherlock, C. A. Ball, E. V. Koonin, S. S. Dwight, M. A. Harris, K. Dolinski, S. Mohr, T. Smith, S. Weng, J. M. Cherry, and D. Botstein.** 1998. Comparison of the complete protein sets of worm and yeast: orthology and divergence. Science **282:**2022–2028.
41. **Christensen, E. A., and H. Kristensen.** 1981. Radiation-resistance of microorganisms from air in clean premises. Acta Pathol. Microbiol. Scand. Sect. B **89:**293–301.
42. **Chuang, S. E., and F. R. Blattner.** 1993. Characterization of twenty-six new heat shock genes of *Escherichia coli*. J. Bacteriol. **175:**5242–5252.
43. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, and B. G. Barrell.** 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:**537–544.
44. **Counsell, T. J., and R. G. E. Murray.** 1986. Polar lipid profiles of the genus *Deinococcus*. Int. J. Syst. Bacteriol. **36:**202–206.
45. **Curnow, A. W., D. L. Tumbula, J. T. Pelaschier, B. Min, and D. Soll.** 1998. Glutamyl-tRNA(Gln) amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. Proc. Natl. Acad. Sci. USA **95:**12838–12843.
46. **Curry, J., and M. K. Walker-Simmons.** 1993. Unusual sequence of group 3 LEA (II) mRNA inducible by dehydration stress in wheat. Plant Mol. Biol. **21:**907–912.
47. **Dalgaard, J. Z., A. J. Klar, M. J. Moser, W. R. Holley, A. Chatterjee, and I. S. Mian.** 1997. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Res. **25:**4626–4638.
48. **Daly, M. J.** 2000. Engineering radiation-resistant bacteria for environmental biotechnology. Curr. Opin. Biotechnol. **11:**280–285.
49. **Daly, M. J., O. Ling, and K. W. Minton.** 1994. Interplasmidic recombination following irradiation of the radioresistant bacterium *Deinococcus radiodurans*. J. Bacteriol. **176:**7506–7515.
50. **Daly, M. J., and K. W. Minton.** 1996. An alternative pathway of recombination of chromosomal fragments precedes *recA*-dependent recombination in the radioresistant bacterium *Deinococcus radiodurans*. J. Bacteriol. **178:**4461–4471.
51. **Daly, M. J., and K. W. Minton.** 1995. Interchromosomal recombination in the extremely radioresistant bacterium *Deinococcus radiodurans*. J. Bacteriol. **177:**5495–5505.
52. **Daly, M. J., and K. W. Minton.** 1997. Recombination between a resident plasmid and the chromosome following irradiation of the radioresistant bacterium *Deinococcus radiodurans*. Gene **187:**225–229.
53. **Daly, M. J., L. Ouyang, P. Fuchs, and K. W. Minton.** 1994. In vivo damage and *recA*-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium *Deinococcus radiodurans*. J. Bacteriol. **176:**3508–3517.
54. **Davis, N. S., G. J. Silverman, and E. B. Mausurosky.** 1963. Radiation-resistant, pigmented coccus isolated from haddock tissue. J. Bacteriol **86:**294–298.
55. **Davis, S. J., A. V. Vener, and R. D. Vierstra.** 1999. Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic eubacteria. Science **286:**2517–2520.
56. **Dean, C. J., P. Feldschreiber, and J. T. Lett.** 1966. Repair of x-ray damage to the deoxyribonucleic acid in *Micrococcus radiodurans*. Nature **209:**49–52.
57. **Doolittle, W. F.** 1999. Lateral genomics. Trends Cell Biol. **9:**M5–M8.
58. **Dure, L. D.** 1993. A repeating 11-mer amino acid motif and plant desiccation. Plant J. **3:**363–364.
59. **Eddy, S. R.** 1998. Profile hidden Markov models. Bioinformatics **14:**755–763.
60. **Eisen, J. A., and P. C. Hanawalt.** 1999. A phylogenomic study of DNA repair genes, proteins, and processes. Mutat. Res. **435:**171–213.
61. **Embley, T. M., A. G. O'Donnell, R. Wait, and J. Rostron.** 1987. Lipid and cell wall amino acid composition in the classification of members of the genus *Deinococcus*. Syst. Appl. Microbiol. **10:**20–27.
62. **Espelund, M., S. Saeboe-Larssen, D. W. Hughes, G. A. Galau, F. Larsen, and K. S. Jakobsen.** 1992. Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differentially by abscisic acid and osmotic stress. Plant J. **2:**241–252. (Erratum, **2:**639.)
63. **Evans, D. M., and B. E. Moseley.** 1988. Deinococcus radiodurans UV endonuclease beta DNA incisions do not generate photoreversible thymine residues. Mutat. Res. **207:** 117–119.
64. **Evans, D. M., and B. E. Moseley.** 1985. Identification and initial characterisation of a pyrimidine dimer UV endonuclease (UV endonuclease beta) from *Deinococcus radiodurans*: a DNA-repair enzyme that requires manganese ions. Mutat. Res. **145:**119–128.
65. **Evans, D. M., and B. E. Moseley.** 1983. Roles of the *uvsC, uvsD, uvsE*, and *mtcA* genes in the two pyrimidine dimer excision repair pathways of *Deinococcus radiodurans*. J. Bacteriol. **156:**576–583.
66. **Evans, T. C., Jr., D. Martin, R. Kolly, D. Panne, L. Sun, I. Ghosh, L. Chen, J. Benner, X. Q. Liu, and M. Q. Xu.** 2000. Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803. J. Biol. Chem. **275:**9091–9094.
67. **Felsenstein, J.** 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol. **266:**418–427.
68. **Fernandez, C., T. Szyperski, T. Bruyere, P. Ramage, E. Mosinger, and K. Wuthrich.** 1997. NMR solution structure of the pathogenesis-related protein P14a. J. Mol. Biol. **266:**576–593.
69. **Ferreira, A. C., M. F. Nobre, F. A. Rainey, M. T. Silva, R. Wait, J. Burghardt, A. P. Chung, and M. S. da Costa.** 1997. *Deinococcus geothermalis* sp. nov. and *Deinococcus murrayi* sp. nov., two extremely radiation-resistant and slightly thermophilic species from hot springs. Int. J. Syst. Bacteriol. **47:**939–947.
70. **Friedberg, E. C.** 1996. Relationships between DNA repair and transcription. Annu. Rev. Biochem. **65:**15–42.
71. **Fujimori, A., Y. Matsuda, Y. Takemoto, Y. Hashimoto, E. Kubo, R. Araki, R. Fukumura, K. Mita, K. Tatsumi, and M. Muto.** 1998. Cloning and mapping of Np95 gene which encodes a novel nuclear protein associated with cell proliferation. Mamm. Genome **9:**1032–1035.
72. **Galas, D. J., and M. Chandler.** 1989. Mobile DNA. American Society for Microbiology, Washington, D.C.
73. **Galau, G. A., H. Y. Wang, and D. W. Hughes.** 1993. Cotton Lea5 and Lea14 encode atypical late embryogenesis-abundant proteins. Plant Physiol. **101:**695–696.
74. **Galperin, M. Y., and E. V. Koonin.** 1999. Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. Genetica **106:**159–170.
75. **Galperin, M. Y., D. A. Natale, L. Aravind, and E. V. Koonin.** 1999. A specialized version of the HD hydrolase domain implicated in signal transduction. J. Mol. Microbiol. Biotechnol. **1:**303–305.
76. **Gilson, E., W. Saurin, D. Perrin, S. Bachellier, and M. Hofnung.** 1991. The BIME family of bacterial highly repetitive sequences. Res. Microbiol. **142:**217–222.
77. **Gorbalenya, A. E.** 1998. Non-canonical inteins. Nucleic Acids Res. **26:**1741–1748.

78. **Grant, C. E., G. Bain, and A. Tsang.** 1990. The molecular basis for alternative splicing of the CABP1 transcripts in *Dictyostelium discoideum*. Nucleic Acids Res. **18:**5457–5463.
79. **Green, C. D., K. S. Long, H. Shi, and S. L. Wolin.** 1998. Binding of the 60-kDa Ro autoantigen to Y RNAs: evidence for recognition in he major groove of a conserced helix. RNA **4:**750–765.
80. **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.
81. **Gutman, P. D., J. D. Carroll, C. I. Masters, and K. W. Minton.** 1994. Sequencing, targeted mutagenesis and expression of a recA gene required for the extreme radioresistance of *Deinococcus radiodurans*. Gene **141:**31–37.
82. **Gutman, P. D., P. Fuchs, L. Ouyang, and K. W. Minton.** 1993. Identification, sequencing, and targeted mutagenesis of a DNA polymerase gene required for the extreme radioresistance of *Deinococcus radiodurans*. J. Bacteriol. **175:**3581–3590.
83. **Gutman, P. D., H. L. Yao, and K. W. Minton.** 1991. Partial complementation of the UV sensitivity of *Deinococcus radiodurans* excision repair mutants by the cloned *denV* gene of bacteriophage T4. Mutat. Res. **254:**207–215.
84. **Guzder, S. N., P. Sung, V. Bailly, L. Prakash, and S. Prakash.** 1994. RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. Nature **369:**578–581.
85. **Handy, J., and R. F. Doolittle.** 1999. An attempt to pinpoint the phylogenetic introduction of glutaminyl-tRNA synthetase among bacteria. J. Mol. Evol. **49:**709–715.
86. **Hansen, M. T.** 1980. Four proteins synthesized in response to deoxyribonucleic acid damage in *Micrococcus radiodurans*. J. Bacteriol. **141:**81–86.
87. **Hansen, M. T.** 1978. Multiplicity of genome equivalents in the radiation-resistant bacterium *Micrococcus radiodurans*. J. Bacteriol. **134:**71–75.
88. **Harmon, F. G., and S. C. Kowalczykowski.** 1998. RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. Genes Dev. **12:**1134–1144.
89. **Harsojo, S. Kitayama, and A. Matsuyama.** 1981. Genome multiplicity and radiation resistance in *Micrococcus radiodurans*. J. Biochem. (Tokyo) **90:**877–880.
90. **Hartzell, P., and D. Kaiser.** 1991. Function of Mg1A, a 22-kilodalton protein essential for gliding in *Myxococcus xanthus*. J. Bacteriol. **173:**7615–7624.
91. **Hecht, H. J., H. Sobek, T. Haag, O. Pfeifer, and K. H. van Pee.** 1994. The metal-ion-free oxidoreductase from *Streptomyces aureofaciens* has an alpha/beta hydrolase fold. Nat. Struct. Biol. **1:**532–537.
92. **Hecker, M., and U. Volker.** 1998. Non-specific, general and multiple stress resistance of growth-restricted *Bacillus subtilis* cells by the expression of the sigmaB regulon. Mol. Microbiol. **29:**1129–1136.
93. **Higgins, D. G., J. D. Thompson, and T. J. Gibson.** 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. **266:**383–402.
94. **Hiom, K., and S. G. Sedgwick.** 1991. Cloning and structural characterization of the *mcrA* locus of *Escherichia coli*. J. Bacteriol. **173:**7368–7373.
95. **Hiramatsu, T., K. Kodama, T. Kuroda, T. Mizushima, and T. Tsuchiya.** 1998. A putative multisubunit Na^+/H^+ antiporter from *Staphylococcus aureus*. J. Bacteriol. **180:**6642–6648.
96. **Hofmann, K., P. Bucher, and A. V. Kajava.** 1998. A model of Cdc25 phosphatase catalytic domain and Cdk-interaction surface based on the presence of a rhodanese homology domain. J. Mol. Biol. **282:**195–208.
97. **Hollung, K., M. Espelund, and K. S. Jakobsen.** 1994. Another Lea B19 gene (Group1 Lea) from barley containing a single 20 amino acid hydrophilic motif. Plant Mol. Biol. **25**(3)**:**559–564.
98. **Horinouchi, S., K. Furuya, M. Nishiyama, H. Suzuki, and T. Beppu.** 1987. Nucleotide sequence of the strepthothricin acetyltransferase gene from *Streptomyces lavendulae* and its expression in heterologous hosts. J. Bacteriol. **169:**1929–1937.
99. **Hsing, Y. C., Z. Y. Chen, M. D. Shih, J. S. Hsieh, and T. Y. Chow.** 1995. Unusual sequences of group 3 LEA mRNA inducible by maturation or drying in soybean seeds. Plant Mol. Biol. **29:**863–868.
100. **Hubbard, T. J., B. Ailey, S. E. Brenner, A. G. Murzin, and C. Chothia.** 1999. SCOP: a Structural Classification Of Proteins database. Nucleic Acids Res. **27:**254–256.
101. **Ibba, M., A. W. Curnow, and D. Soll.** 1997. Aminoacyl-tRNA synthesis: divergent routes to a common goal. Trends Biochem. Sci. **22:**39–42.
102. **Jilani, A., D. Ramotar, C. Slack, C. Ong, X. M. Yang, S. W. Scherer, and D. D. Lasko.** 1999. Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3′-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. J. Biol. Chem. **274:**24176–24186.
103. **Jobling, M. G., and D. A. Ritchie.** 1988. The nucleotide sequence of a plasmid determinant for resistance to tellurium anions. Gene **66:**245–258.
104. **Jordan, A., and P. Reichard.** 1998. Ribonucleotide reductases. Annu. Rev. Biochem. **67:**71–98.
105. **Kanehisa, M., and S. Goto.** 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. **28:**27–30.
106. **Kitajima, S., and F. Sato.** 1999. Plant pathogenesis-related proteins: molecular mechanisms of gene expression and protein function. J. Biochem. (Tokyo) **125:**1–8.
107. **Kitayama, S., and A. Matsuyama.** 1971. Mechanism for radiation lethality in *M. radiodurans*. Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. **19:**13–19.
108. **Kobatake, M., S. Tanabe, and S. Hasegawa.** 1973. New *Micrococcus* radioresistant red pigment, isolated from *Lama glama* feces, and its use as microbiological indicator of radiosterilization. C. R. Seances Soc. Biol. Fil. **167:**1506–1510. (In French.)
109. **Koonin, E. V., L. Aravind, and M. Y. Galperin.** 2000. A comparative-genomic view of the microbial stress response, p. 417–444. *In* G. Storz and R. Hengge-Aronis (ed.), Bacterial stress responses. ASM Press, Washington, D.C.
110. **Koonin, E. V., and R. L. Tatusov.** 1994. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. J. Mol. Biol. **244:**125–132.
111. **Kornezos, A., and W. Chia.** 1992. Apical secretion and association of the *Drosophila* yellow gene product with developing larval cuticle structures during embryogenesis. Mol. Gen. Genet. **235:**397–405.
112. **Krasin, F., and F. Hutchinson.** 1977. Repair of DNA double-strand breaks in *Escherichia coli*, which requires recA function and the presence of a duplicate genome. J. Mol. Biol. **116:**81–98.
113. **Kreppel, L. K., and G. W. Hart.** 1999. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. J. Biol. Chem. **274:**32015–32022.
114. **Kristensen, H., and E. A. Christensen.** 1981. Radiation-resistant microorganisms isolated from textiles. Acta Pathol. Microbiol. Scand. Sect. B. **89:**303–309.
115. **Kubler, O., and W. Baumeister.** 1978. The structure of a periodic cell wall component (HPI) layer of *Micrococcus radiodurans*. Cytobiologie **17:**1–9.
116. **Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Dusterhoft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Henaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauel, C. Medigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, et al.** 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature **390:**249–256.
117. **Kuzminov, A.** 1999. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. Microbiol. Mol. Biol. Rev. **63:**751–813.
118. **Lancy, P., Jr., and R. G. Murray.** 1978. The envelope of *Micrococcus radiodurans*: isolation, purification, and preliminary analysis of the wall layers. Can. J. Microbiol. **24:**162–176.
119. **Lange, C. C., L. P. Wackett, K. W. Minton, and M. J. Daly.** 1998. Engineering a recombinant *Deinococcus radiodurans* for organopollutant degradation in radioactive mixed waste environments. Nat. Biotechnol. **16:**929–933.
120. **Lennon, E., P. D. Gutman, H. L. Yao, and K. W. Minton.** 1991. A highly conserved repeated chromosomal sequence in the radioresistant bacterium *Deinococcus radiodurans* SARK. J. Bacteriol. **173:**2137–2140.
121. **Leonard, C. J., L. Aravind, and E. V. Koonin.** 1998. Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily. Genome Res. **8:**1038–1047.
122. **Lewis, N. F.** 1971. Studies on a radio-resistant coccus isolated from Bombay duck (*Harpodon nehereus*). J. Gen. Microbiol. **66:**29–35.
123. **Lin, J., R. Qi, C. Aston, J. Jing, T. S. Anantharaman, B. Mishra, O. White, M. J. Daly, K. W. Minton, J. C. Venter, and D. C. Schwartz.** 1999. Whole-genome shotgun optical mapping of *Deinococcus radiodurans*. Science **285:**1558–1562.
124. **Maitra, N., and J. C. Cushman.** 1994. Isolation and characterization of a drought-induced soybean cDNA encoding a D95 family late-embryogenesis-abundant protein. Plant Physiol. **106:**805–806.
125. **Majumdar, S., and A. K. Chandra.** 1985. UV repair and mutagenesis in *Azotobacter vinelandii*. Zentbl. Mikrobiol. **140:**247–254.
126. **Makarova, K. S., L. Aravind, M. Y. Galperin, N. V. Grishin, R. L. Tatusov, Y. I. Wolf, and E. V. Koonin.** 1999. Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. Genome Res. **9:**608–628.
127. **Makarova, K. S., Y. I. Wolf, O. White, K. Minton, and M. J. Daly.** 1999.

Short repeats and IS elements in the extremely radiation-resistant bacterium *Deinococcus radiodurans* and comparison to other bacterial species. Res. Microbiol. **150:** 711–724.
128. **Marcotte, E. M., M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg.** 1999. Detecting protein function and protein-protein interactions from genome sequences. Science **285:**751–753.
129. **Markillie, L. M., S. M. Varnum, P. Hradecky, and K. K. Wong.** 1999. Targeted mutagenesis by duplication insertion in the radioresistant bacterium *Deinococcus radiodurans*: radiation sensitivities of catalase (*katA*) and superoxide dismutase (*sodA*) mutants. J. Bacteriol. **181:**666–669.
130. **Maruta, K., K. Hattori, T. Nakada, M. Kubota, T. Sugimoto, and M. Kurimoto.** 1996. Cloning and sequencing of trehalose biosynthesis genes from *Rhizobium* sp. M-11. Biosci. Biotechnol. Biochem. **60:**717–720.
131. **Masters, C. I., and K. W. Minton.** 1992. Promoter probe and shuttle plasmids for *Deinococcus radiodurans*. Plasmid **28:**258–261.
132. **Masters, C. I., B. E. Moseley, and K. W. Minton.** 1991. AP endonuclease and uracil DNA glycosylase activities in *Deinococcus radiodurans*. Mutat. Res. **254:**263–272.
133. **Masters, C. I., M. D. Smith, P. D. Gutman, and K. W. Minton.** 1991. Heterozygosity and instability of amplified chromosomal insertions in the radioresistant bacterium *Deinococcus radiodurans*. J. Bacteriol. **173:**6110–6117.
134. **Mathews, D. H., T. C. Andre, J. Kim, D. H. Turner, and M. Zuker.** 1998. An updated recursive algorithm for RNA secondary structure prediction with improved free energy parameters. Am. Chem. Soc. Symp. Ser. **682:**246–257.
135. **Mattimore, V., and J. R. Battista.** 1996. Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. J. Bacteriol. **178:**633–637.
136. **Michaels, M. L., and J. H. Miller.** 1992. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). J. Bacteriol. **174:**6321–6325.
137. **Minton, K. W.** 1994. DNA repair in the extremely radioresistant bacterium *Deinococcus radiodurans*. Mol. Microbiol. **13:**9–15.
138. **Minton, K. W.** 1996. Repair of ionizing-radiation damage in the radiation resistant bacterium *Deinococcus radiodurans*. Mutant. Res. **363:**1–7.
139. **Minton, K. W., and M. J. Daly.** 1995. A model for repair of radiation-induced DNA double-strand breaks in the extreme radiophile *Deinococcus radiodurans*. Bioessays **17:**457–464.
140. **Moons, A., A. De Keyser, and M. Van Montagu.** 1997. A group 3 LEA cDNA of rice, responsive to abscisic acid, but not to jasmonic acid, shows variety-specific differences in salt stress response. Gene **191:**197–204.
141. **Morozov, V., A. R. Mushegian, E. V. Koonin, and P. Bork.** 1997. A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases. Trends Biochem. Sci. **22:**417–418.
142. **Moseley, B. E., and D. M. Evans.** 1983. Isolation and properties of strains of *Micrococcus* (*Deinococcus*) *radiodurans* unable to excise ultraviolet light-induced pyrimidine dimers from DNA: evidence for two excision pathways. J. Gen. Microbiol. **129:**2437–2445.
143. **Moseley, B. E., and J. K. Setlow.** 1968. Transformation in *Micrococcus radiodurans* and the ultraviolet sensitivity of its transforming DNA. Proc. Natl. Acad. Sci. USA **61:**176–183.
144. **Mun, C., J. Del Rowe, M. Sandigursky, K. W. Minton, and W. A. Franklin.** 1994. DNA deoxyribophosphodiesterase and an activity that cleaves DNA containing thymine glycol adducts in *Deinococcus radiodurans*. Radiat. Res. **138:**282–285.
145. **Murray, R. G. E.** 1992. The family Deinococcaceae, p. 3732–3744. *In* A. Balows, H. G. Trüper, M. Dworkin, W. Harder, and K. H. Schleifer (ed.), The prokaryotes, vol. 4. Springer-Verlag, New York, N.Y.
146. **Murray, R. G. E.** 1986. Family II. *Deinococcaceae*, p. 1035–1043. *In* P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 2. The Williams & Wilkins Co., Baltimore, Md.
147. **Reference deleted.**
148. **Myers, C. R., and J. M. Myers.** 1997. Outer membrane cytochromes of *Shewanella putrefaciens* MR-1: spectral analysis, and purification of the 83-kDa c-type cytochrome. Biochim. Biophys. Acta **1326:**307–318.
149. **Narumi, I., K. Cherdchu, S. Kitayama, and H. Watanabe.** 1997. The *Deinococcus radiodurans* uvr A gene: identification of mutation sites in two mitomycin-sensitive strains and the first discovery of insertion sequence element from deinobacteria. Gene **198:**115–126.
150. **Narumi, I., K. Satoh, M. Kikuchi, T. Funayama, S. Kitayama, T. Yanagisawa, H. Watanabe, and K. Yamamoto.** 1999. Molecular analysis of the *Deinococcus radiodurans* recA locus and identification of a mutation site in a DNA repair-deficient mutant, rec30. Mutat. Res. **435:**233–243.
151. **Naudet, R.** 1976. The Oklo nuclear reactors: 1800 million years ago. Interdiscip. Sci. Rev. **1:**74–84.
152. **Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger.** 1996. *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
153. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, C. M. Fraser, et al.** 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature **399:**323–329.
154. **Ng, W. V., S. A. Ciufo, T. M. Smith, R. E. Bumgarner, D. Baskin, J. Faust, B. Hall, C. Loretz, J. Seto, J. Slagel, L. Hood, and S. DasSarma.** 1998. Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? Genome Res. **8:**1131–1141.
155. **O'Halloran, T. V.** 1993. Transition metals in control of gene expression. Science **261:**715–725.
156. **Olsen, G. J., and C. R. Woese.** 1993. Ribosomal RNA: a key to phylogeny. FASEB J. **7:**113–123.
157. **Overbeek, R., N. Larsen, G. D. Pusch, M. D'Souza, E. Selkov, Jr., N. Kyrpides, M. Fonstein, N. Maltsev, and E. Selkov.** 2000. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucleic Acids Res. **28:**123–125.
158. **Oyaizu, H.** 1987. A radiation-resistant rod-shaped bacterium, *Deinobacter grandis* gen. nov., sp. nov., with peptidoglycan containing ornithine. Int. J. Syst. Bacteriol. **37:**62–67.
159. **Parkinson, J. S., and E. C. Kofoid.** 1992. Communication modules in bacterial signaling proteins. Annu. Rev. Genet. **26:**71–112.
160. **Pask-Hughes, R. A., and N. Shaw.** 1982. Glycolipids from some extreme thermophilic bacteria belonging to the genus *Thermus*. J. Bacteriol. **149:**54–58.
161. **Piatkowski, D., K. Schneider, F. Salamini, and D. Bartels.** 1990. Characterization of five abscisic acid-responsive cDNA clones isolated from the desiccation-tolerant plant *Craterostigma plantagineum* and their relationship to other water-stress genes. Plant Physiol. **94:**1682–1688.
162. **Pietrokovski, S.** 1998. Identification of a virus intein and a possible variation in the protein-splicing reaction. Curr. Biol. **8:**R634–R635.
163. **Ponting, C. P., L. Aravind, J. Schultz, P. Bork, and E. V. Koonin.** 1999. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. J. Mol. Biol. **289:**729–745.
164. **Punita, S. J., M. A. Reddy, and H. K. Das.** 1989. Multiple chromosomes of *Azotobacter vinelandii*. J. Bacteriol. **171:**3133–3138.
165. **Quintela, J. C., F. Garcia-del Portillo, E. Pittenauer, G. Allmaier, and M. A. de Pedro.** 1999. Peptidoglycan fine structure of the radiotolerant bacterium *Deinococcus radiodurans* Sark. J. Bacteriol. **181:**334–337.
166. **Quintela, J. C., E. Pittenauer, G. Allmaier, V. Aran, and M. A. de Pedro.** 1995. Structure of peptidoglycan from *Thermus thermophilus* HB8. J. Bacteriol. **177:**4947–4962.
167. **Raibaud, A., M. Zalacain, T. G. Holt, R. Tizard, and C. J. Thompson.** 1991. Nucleotide sequence analysis reveals linked *N*-acetyl hydrolase, thioesterase, transport, and regulatory genes encoded by the bialaphos biosynthetic gene cluster of *Streptomyces hygroscopicus*. J. Bacteriol. **173:**4454–4463.
168. **Rainey, F. A., M. F. Nobre, P. Shumann, E. Stackebrandt, and M. S. da Costa.** 1997. Phylogenetic diversity of the deinococci as determined by 16S ribosomal DNA sequence comparison. Int. J. Syst. Bacteriol. **47:**510–514.
169. **Rathbone, D. A., P. J. Holt, C. R. Lowe, and N. C. Bruce.** 1997. Molecular analysis of the *Rhodococcus* sp. strain H1 *her* gene and characterization of its product, a heroin esterase, expressed in *Escherichia coli*. Appl. Environ. Microbiol. **63:**2062–2066.
170. **Rebeyrotte, N.** 1983. Induction of mutation in *Micrococcus radiodurans* by *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine. Mutat. Res. **108:**57–66.
171. **Richmond, R. C., R. Sridhar, and M. J. Daly.** 1999. Physicochemical survival pattern for the radiophile *Deinococcus radiodurans*: a polyextremophile model for life on Mars. SPIE **3755:**210–222.
172. **Sadoff, H. L., B. Shimel, and S. Ellis.** 1979. Characterization of *Azotobacter vinelandii* deoxyribonucleic acid and folded chromosomes. J. Bacteriol. **138:**871–877.
173. **Sanchez-Campillo, M., S. Dramsi, J. M. Gomez-Gomez, E. Michel, P. Dehoux, P. Cossart, F. Baquero, and J. C. Perez-Diaz.** 1995. Modulation of DNA topology by flaR, a new gene from *Listeria monocytogenes*. Mol. Microbiol. **18:**801–811.
174. **Sandigursky, M., and W. A. Franklin.** 1999. Thermostable uracil-DNA glycosylase from *Thermotoga maritima*, a member of a novel class of DNA repair enzymes. Curr. Biol. **9:**531–534.
175. **Schaffer, A. A., Y. I. Wolf, C. P. Ponting, E. V. Koonin, L. Aravind, and S. F. Altschul.** 1999. IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. Bioinformatics **15:**1000–1011.
176. **Schleifer, K. H., and O. Kandler.** 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol. Rev. **36:**407–477.
177. **Schlenk, D.** 1998. Occurrence of flavin-containing monooxygenases in non-mammalian eukaryotic organisms. Comp. Biochem. Physiol. Ser. C **121:**185–195.
178. **Schultz, J., R. R. Copley, T. Doerks, C. P. Ponting, and P. Bork.** 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. **28:**231–234.
179. **Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting.** 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA **95:**5857–5864.

180. **Schwarz, D. A., C. D. Katayama, and S. M. Hedrick.** 1998. Schlafen, a new family of growth regulatory genes that affect thymocyte development. Immunity **9:**657–668.
181. **Seledtsov, I. A., I. I. Vul'f, and K. S. Makarova.** 1995. Multiple alignment of biopolymer sequences, based on the search for statistically significant common segments. Mol. Biol. (Moscow) **29:**1023–1039. (In Russian.)
182. **Senkevich, T. G., E. V. Koonin, J. J. Bugert, G. Darai, and B. Moss.** 1997. The genome of molluscum contagiosum virus: analysis and comparison with other poxviruses. Virology **233:**19–42.
183. **Setlow, D. M., and D. E. Duggan.** 1964. The resistance of *Micrococcus radiodurans* to ultraviolet radiation: ultraviolet-induced lesions in the cell's DNA. Biochim. Biophys. Acta **87:**664–668.
184. **Shanado, Y., J. Kato, and H. Ikeda.** 1998. *Escherichia coli* HU protein suppresses DNA-gyrase-mediated illegitimate recombination and SOS induction. Genes Cells **3:**511–520.
185. **Silhavy, D., G. Hutvagner, E. Barta, and Z. Banfalvi.** 1995. Isolation and characterization of a water-stress-inducible cDNA clone from *Solanum chacoense*. Plant Mol. Biol. **27:**587–595.
186. **Sleytr, U. B., and A. M. Glauert.** 1982. Bacterial cell walls and membranes, p. 41–76. *In* J. R. Harris (ed.), Electron microscopy of proteins, vol. 3. Academic Press, Ltd., London, United Kingdom.
187. **Sleytr, U. B., M. Kocur, A. M. Glauert, and M. J. Thornley.** 1973. A study by freeze-etching of the fine structure of *Micrococcus radiodurans*. Arch. Mikrobiol. **94:**77–87.
188. **Smith, K. C., and K. D. Martignoni.** 1976. Protection of *Escherichia coli* cells against the lethal effects of ultraviolet and x irradiation by prior x irradiation: a genetic and physiological study. Photochem. Photobiol. **24:** 515–523.
189. **Smith, M. D., R. Abrahamson, and K. W. Minton.** 1989. Shuttle plasmids constructed by the transformation of an *Escherichia coli* cloning vector into two *Deinococcus radiodurans* plasmids. Plasmid **22:**132–142.
190. **Smith, M. D., E. Lennon, L. B. McNeil, and K. W. Minton.** 1988. Duplication insertion of drug resistance determinants in the radioresistant bacterium *Deinococcus radiodurans*. J. Bacteriol. **170:**2126–2135.
191. **Smith, M. D., C. I. Masters, E. Lennon, L. B. McNeil, and K. W. Minton.** 1991. Gene expression in *Deinococcus radiodurans*. Gene **98:**45–52.
192. **Snel, B., P. Bork, and M. Huynen.** 2000. Genome evolution. Gene fusion versus gene fission. Trends Genet. **16:**9–11.
193. **Sorenson, J. A.** 1986. Perception of radiation hazards. Semin. Nucl. Med. **16:**158–170.
194. **Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis.** 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science **282:**754–759.
195. **Subramanian, G., E. V. Koonin, and L. Aravind.** 2000. Comparative genome analysis of the pathogenic spirochetes *Borrelia burgdorferi* and *Treponema pallidum*. Infect. Immun. **68:**1633–1648.
196. **Sung, P., S. N. Guzder, L. Prakash, and S. Prakash.** 1996. Reconstitution of TFIIH and requirement of its DNA helicase subunits, Rad3 and Rad25, in the incision step of nucleotide excision repair. J. Biol. Chem. **271:**10821–10826.
197. **Sweet, D. M., and B. E. Moseley.** 1974. Accurate repair of ultraviolet-induced damage in *Micrococcus radiodurans*. Mutat. Res. **23:**311–318.
198. **Sweet, D. M., and B. E. Moseley.** 1976. The resistance of *Micrococcus radiodurans* to killing and mutation by agents which damage DNA. Mutat. Res. **34:**175–186.
199. **Tan, S. T., and R. B. Maxcy.** 1986. Simple method to demonstrate radiation-inducible radiation resistance in microbial cells. Appl. Environ. Microbiol. **51:**88–90.
200. **Tanaka, A., H. Hirano, M. Kikuchi, S. Kitayama, and H. Watanabe.** 1996. Changes in cellular proteins of *Deinococcus radiodurans* following gamma-irradiation. Radiat. Environ. Biophys. **35:**95–99.
201. **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28:**33–36.
202. **Tatusov, R. L., E. V. Koonin, and D. J. Lipman.** 1997. A genomic perspective on protein families. Science **278:**631–637.
203. **Taylor, B. L., and I. B. Zhulin.** 1999. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol. Mol. Biol. Rev. **63:**479–506.
204. **Thompson, B. G., R. Anderson, and R. G. E. Murray.** 1980. Unusual polar lipids of *Micrococcus radiodurans* strain SARK. Can. J. Microbiol. **26:**1408–1411.
205. **Thompson, B. G., and R. G. Murray.** 1981. Isolation and characterization of the plasma membrane and the outer membrane of *Deinococcus radiodurans* strain Sark. Can. J. Microbiol. **27:**729–734.
206. **Thompson, B. G., and R. G. E. Murray.** 1982. The association of the surface array and the outer membrane of *Deinococcus radiodurans*. Can. J. Microbiol. **28:**1081–1088.
207. **Thornley, M. J.** 1963. Radiation resistance among bacteria. J. Appl. Bacteriol. **26:**539–547.
208. **Thornley, M. J., R. W. Horne, and A. M. Glauert.** 1965. The fine structure of *Micrococcus radiodurans*. Arch. Mikrobiol. **51:**267–289.
209. **Tsusaki, K., T. Nishimoto, T. Nakada, M. Kubota, H. Chaen, S. Fukuda, T. Sugimoto, and M. Kurimoto.** 1997. Cloning and sequencing of trehalose synthase gene from *Thermus aquaticus* ATCC 33923. Biochim. Biophys. Acta **1334:**28–32.
210. **Tzamarias, D., and K. Struhl.** 1995. Distinct TPR motifs of Cyc8 are involved in recruiting the Cyc8-Tup1 corepressor complex to differentially regulated promoters. Genes Dev. **9:**821–831.
211. **Udupa, K. S., P. A. O'Cain, V. Mattimore, and J. R. Battista.** 1994. Novel ionizing radiation-sensitive mutants of *Deinococcus radiodurans*. J. Bacteriol. **176:**7439–7446.
212. **Venkateswaran, A., S. C. McFarlan, D. Ghosal, K. W. Minton, A. Vasilenko, K. S. Makarova, L. P. Wackett, and M. J. Daly.** 2000. Physiologic determinants of radiation resistance in *Deinococcus radiodurans*. Appl. Environ. Microbiol. **66:**2620–2626.
213. **Vukovic-Nagy, B., B. W. Fox, and M. Fox.** 1974. The release of a deoxyribonucleic acid fragment after x-irradiation of *Micrococcus radiodurans*. Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. **25:**329–337.
214. **Walker, D. R., and E. V. Koonin.** 1997. SEALS: a system for easy analysis of lots of sequences. Ismb **5:**333–339.
215. **Wang, P., and H. E. Schellhorn.** 1995. Induction of resistance to hydrogen peroxide and radiation in *Deinococcus radiodurans*. Can. J. Microbiol. **41:** 170–176.
216. **Welsh, D. T., and R. A. Herbert.** 1999. Osmotically induced intracellular trehalose, but not glycine betaine accumulation promotes desiccation tolerance in *Escherichia coli*. FEMS Microbiol. Lett. **174:**57–63.
217. **Werneburg, B. G., J. Ahn, X. Zhong, R. J. Hondal, V. S. Kraynov, and M. D. Tsai.** 1996. DNA polymerase beta: pre-steady-state kinetic analysis and roles of arginine-283 in catalysis and fidelity. Biochemistry **35:**7041–7050.
218. **White, O., J. A. Eisen, J. F. Heidelberg, E. K. Hickey, J. D. Peterson, R. J. Dodson, D. H. Haft, M. L. Gwinn, W. C. Nelson, D. L. Richardson, K. S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J. J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K. S. Makarova, L. Aravind, M. J. Daly, K. W. Minton, R. D. Fleischmann, K. A. Ketchum, K. E. Nelson, S. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science **286:**1571–1577.
219. **Wolf, Y. I., S. E. Brenner, P. A. Bash, and E. V. Koonin.** 1999. Distribution of protein folds in the three superkingdoms of life. Genome Res. **9:**17–26.
220. **Wootton, J. C.** 1994. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput. Chem. **18:**269–285.
221. **Work, E., and H. Griffiths.** 1968. Morphology and chemistry of cell walls of *Micrococcus radiodurans*. J. Bacteriol. **95:**641–657.
222. **Wu, H., Z. Hu, and X. Q. Liu.** 1998. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. Proc. Natl. Acad. Sci. USA **95:**9226–9231.
223. **Yajima, H., M. Takao, S. Yasuhira, J. H. Zhao, C. Ishii, H. Inoue, and A. Yasui.** 1995. A eukaryotic gene encoding an endonuclease that specifically repairs DNA damaged by ultraviolet light. EMBO J. **14:**2393–2399.
224. **Yan, H., and M. D. Tsai.** 1999. Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. Adv. Enzymol. Relat. Areas Mol. Biol. **73:**103–134.
225. **Zegzouti, H., B. Jones, P. Frasse, C. Marty, B. Maitre, A. Latch, J. C. Pech, and M. Bouzayen.** 1999. Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. Plant J. **18:**589–600.
226. **Zegzouti, H., B. Jones, C. Marty, J. M. Lelievre, A. Latche, J. C. Pech, and M. Bouzayen.** 1997. ER5, a tomato cDNA encoding an ethylene-responsive LEA-like protein: characterization and expression in response to drought, ABA and wounding. Plant Mol. Biol. **35:**847–854.