

Article

Structural Differences between the Genomes of *Deinococcus radiodurans* Strains from Different Laboratories

Ksenija Zahradka , Davor Zahradka  and Jelena Repar *

Division of Molecular Biology, Ruđer Bošković Institute, 10000 Zagreb, Croatia; kvlahovi@irb.hr (K.Z.); zahradka@irb.hr (D.Z.)

* Correspondence: jelena.repar@irb.hr

Abstract: The bacterium *Deinococcus radiodurans* is known to efficiently and accurately reassemble its genome after hundreds of DNA double-strand breaks (DSBs). Only at very large amounts of radiation-induced DSBs is this accuracy affected in the wild-type *D. radiodurans*, causing rearrangements in its genome structure. However, changes in its genome structure may also be possible during the propagation and storage of cell cultures. We investigate this possibility by listing structural differences between three completely sequenced genomes of *D. radiodurans* strains with a recent common ancestor—the type strain stored and sequenced in two different laboratories (of the ATCC 13939 lineage) and the first sequenced strain historically used as the reference (ATCC BAA-816). We detected a number of structural differences and found the most likely mechanisms behind them: (i) transposition/copy number change in mobile interspersed repeats—insertion sequences and small non-coding repeats, (ii) variable number of monomers within tandem repeats, (iii) deletions between long direct DNA repeats, and (iv) deletions between short (4–10 bp) direct DNA repeats. The most surprising finding was the deletions between short repeats because it indicates the utilization of a less accurate DSB repair mechanism in conditions in which a more accurate one should be both available and preferred. The detected structural differences, as well as SNPs and short indels, while being important footprints of deinococcal DNA metabolism and repair, are also a valuable resource for researchers using these *D. radiodurans* strains.

Keywords: structural variation; genome rearrangements; mobile elements; IS; SNR; alternative end-joining



Citation: Zahradka, K.; Zahradka, D.; Repar, J. Structural Differences between the Genomes of *Deinococcus radiodurans* Strains from Different Laboratories. *Genes* **2024**, *15*, 847. <https://doi.org/10.3390/genes15070847>

Academic Editor: Silvia Turrone

Received: 24 May 2024

Revised: 24 June 2024

Accepted: 25 June 2024

Published: 27 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The bacterium *Deinococcus radiodurans* can survive and accurately repair hundreds of simultaneous DNA double-strand breaks (DSBs) caused by γ -radiation, desiccation, or other DNA damaging agents [1–4]. For this reason, *D. radiodurans* has long been the subject of DSB repair studies. Despite the potent and accurate repair, mistakes do seem to happen at extreme levels of DSBs, implying inaccurate DNA repair at very high levels of DSBs [5,6].

DSB repair based on RecA-dependent homologous DNA matching is highly accurate but can cause genome structural changes in the presence of interspersed repetitive sequences by homologously matching ectopic repetitive sequences [7,8]. *D. radiodurans* contains a high number of interspersed repetitive sequences—mobile genetic elements, which consist of both insertion sequences (ISs with a length of ~500 to ~3000 bp, encoding transposases) and small non-coding repeats (SNRs with a length of 60–215 bp) [9,10]. However, the presence of repeats does not seem to notably hinder the ability of γ -irradiated *D. radiodurans* to accurately repair hundreds of simultaneous DSBs.

The efficient and accurate repair of DSBs in *D. radiodurans*, which is not hindered by the presence of repetitive sequences, is considered a possible adaptation to DSB-causing dry environments such as deserts [2,3,11]. It follows that *D. radiodurans* would hypothetically produce fewer mistakes than, for example, *Escherichia coli*, under equivalent DSB-inducing

conditions even though the two species contain a comparable number of interspersed repetitive sequences [12].

The main mechanism of DSB repair in *D. radiodurans*, i.e., the extended-synthesis-dependent strand annealing (ESDSA), is dependent on homologous DNA matching but is also atypical for a bacterium. ESDSA relies on the existence of multiple genome copies within the same cell, i.e., on the existence of an unbroken homologous DNA stretch overlapping each DSB; this unbroken stretch of DNA is then used as a template for massive DNA synthesis [13]. The accuracy of the process is strongly dependent on the (long) homology between DNA strands (template- and DSB site-strand), which leads to the extension and attachment of DNA strands after DSBs. Classical homologous recombination (HR) has also been detected in *D. radiodurans* as a final stage of DSB repair during which the ESDSA-produced long linear intermediates are matured into unit-size circular chromosomes [13].

Alternatively, more inaccurate mechanisms have been proposed to act as backups in DSB repair in *D. radiodurans*: single-strand annealing (SSA) [14–16] and alternative end-joining (A-EJ) [17]. They utilize shorter homologies for strand reattachment (>25 bp and 2–20 bp, respectively) and are, therefore, more prone to mistakes and the production of genome structural change. Different mechanisms can usually be distinguished by the features of a resulting structural change, features such as the length of homologies at the ends of deleted DNA, or the linear distance between misattached DNA ends [18–21]. These less accurate mechanisms of DSB repair are thought to only be employed as a backup in case of an overwhelming number of DSBs.

Spontaneous DSBs, i.e., DSBs not caused by external agents, can also appear in bacteria, although not many at once, during normal growth and DNA replication [22–24]. It follows that genome structural change can also spontaneously arise in normally growing bacterial populations. In fact, spontaneous genome change in laboratory strains has been previously detected for several bacteria, including *Vibrio cholerae* and *E. coli* [25,26]. The frequency of a genome change in the population can be increased by selection or drift. Interestingly, selection has been known to act in the genome of *E. coli* under specific laboratory conditions [25,27]. However, how laboratory storage and propagation will influence the generation and frequency of genome change in a laboratory strain population is often unclear; here, we characterize such genome changes in different strains of *D. radiodurans*.

In this paper, we analyze differences between the genomes of *D. radiodurans* strains descended from the first *D. radiodurans* isolate obtained in 1956 [28]. The samples of this isolate have been separated at an unknown point in time and stored and propagated in different laboratories, and their genomes were subsequently completely sequenced and published independently by different groups. We use the published genome sequences resulting from this “accidental” diversification experiment to ask: Has the efficient and accurate repair of DSBs in *D. radiodurans* protected its genome structure from rearrangements during divergence in different laboratories?

Our starting hypothesis was that there would be no traces of inaccurate repair of spontaneous DSBs in these recently diverged *D. radiodurans* strains due to the high accuracy of the main DSB repair mechanism. More precisely, there would be no genome rearrangements (deletions, duplications, inversions, or translocations) bordered (and, presumably, caused) by repeated sequences that may have caused the derailment of the DSB repair mechanism. To test this hypothesis, we specifically studied structural genome changes (i.e., changes affecting regions of DNA of approximately 1 kb or longer) because they carry potential implications for DSB repair in this bacterium. We detect, characterize, and classify structural changes and make conclusions about the mechanisms by which they came about. We discuss the detected mechanisms in the context of the expected extreme accuracy of DSB repair in *D. radiodurans*.

Additionally, we characterize the single nucleotide polymorphisms (SNPs) and short insertions/deletions (indels) and identify affected genes. We expect that the comprehensive list of differences between the deinococcal genomes will benefit researchers studying specific genes in the investigated *D. radiodurans* strains.

2. Materials and Methods

The genome of *D. radiodurans* is multipartite. The sequences used in this paper are available in GenBank under IDs (i) AE000513.1, AE001825.1, AE001826.1, and AE001827.1 (chromosome I, chromosome II, megaplasmid MP1, and plasmid CP1, respectively, from the first completely sequenced genome of *D. radiodurans* [29] of the strain ATCC BAA-816) and (ii) CP038663.1, CP038664.1, CP038665.1, and CP038666.1 (chromosome I, chromosome II, megaplasmid MP1, and plasmid CP1, respectively, of the ATCC 13939 lineage, obtained in 2002 from the laboratory of Suzanne Sommer (Université Paris-Sud, Orsay, France), subsequently stored at -80°C in 15% glycerol and PacBio sequenced and assembled by our laboratory [17]). The first genome is referred to as genome-1999 throughout the paper, and the second one is genome-2021, referring to the years in which they have been published. The listed sequences for genome-1999 and genome-2021 are available in the NCBI Assembly database under IDs ASM856v1 and ASM2054668v1 and have been evaluated internally by NCBI to have completeness scores of 97.17% and 98.67%, respectively. Additionally, we have included a third published and completely sequenced *D. radiodurans* genome in our analysis [30], the genome elements of which are available in the GenBank under IDs CP015081.1, CP015082.1, CP015083.1, and CP015084.1. Its NCBI genome assembly ID is ASM163882v1 and is marked to have a completeness score of 98.81%. This genome will be referred to as genome-2016, referring to the year in which it was published. To simplify the analysis and alignment to the other two genomes, genome-2016 was fully circled (i.e., the very long bordering DNA overlaps present in some genome elements were cut), and the genome reverse complement was calculated for the genome elements oriented differently than the corresponding elements in the other two genomes.

The phylogenetic distance between the different *D. radiodurans* strains was estimated using chromosome I of each genome and by calculating their distance using the *and*i software v.0.12 [31]. Based on the distances, a neighbor-joining phylogenetic tree was constructed using MEGA v. 11.0.13 [32]. The *Deinococcus deserti* chromosome I that we used is available under GenBank ID CP001114.1.

Differences between the genomes were detected using the methodology described in our previous paper [17]. Briefly, the genomes were aligned and compared using the program DNAdiff v.1.3. from the Mummer package [33]. DNAdiff reports both structural differences and SNPs between pairs of strains. Based on this report of structural differences, the sequences encompassing the nucleotides reported as affected by the difference, i.e., ± 100 bp, were extracted from the genomes and used for structural change characterization to identify footprints of possible mechanisms that caused the change. Blastn was used to identify structural changes stemming from the transpositional activity; the affected genome regions were matched to the ISFinder database of IS sequences using the blastn option of ISFinder (available as an online tool accessed during April of 2024) [34] and to the previously characterized SNR (small non-coding repeats) sequences [9] using blastn (BLAST v2.9.0+, [35]) locally. The results were used to detect which of these changes corresponded to a movement of an insertion sequence or SNR sequence. The sequences pinpointed by DNAdiff were also examined by Tandem Repeats Finder (TRF) [36] as well as Inverted Repeats Finder (IRF) [37], online tools accessed in April 2024. The structural changes that could not be explained by the transposition of repetitive sequences or variation in the number of tandem repeats were examined manually through the blastn alignment [35] to the corresponding change from the other *D. radiodurans* strain or to itself.

The SNP output of DNAdiff was mapped by a custom Java script to the GenBank annotations of genes in genome-1999 in order to examine the affected genes, i.e., genes whose sequence differed between the two genomes.

3. Results

We compared the genomes of three *D. radiodurans* strains that have diverged during processes of growth and storage in different laboratories. These genomes will be referred to as genome-1999, genome-2016, and genome-2021 according to the year in which they were

published (see Section 2). The phylogenetic tree shows that genome-2021 and genome-2016 are very closely related and genome-1999 is more distant (Figure 1A). However, these are very small distances, which become unnoticeable when these genomes are compared to a closely related outgroup such as *Deinococcus deserti* (Figure 1B). This is expected since the three investigated strains of *D. radiodurans* originate from the same recent (1956) isolate. Throughout Sections 3 and 4, if genome-2016 is not specifically mentioned, it means that its structure corresponds to the structure of genome-2021, as would be expected from their phylogenetic closeness (Figure 1).

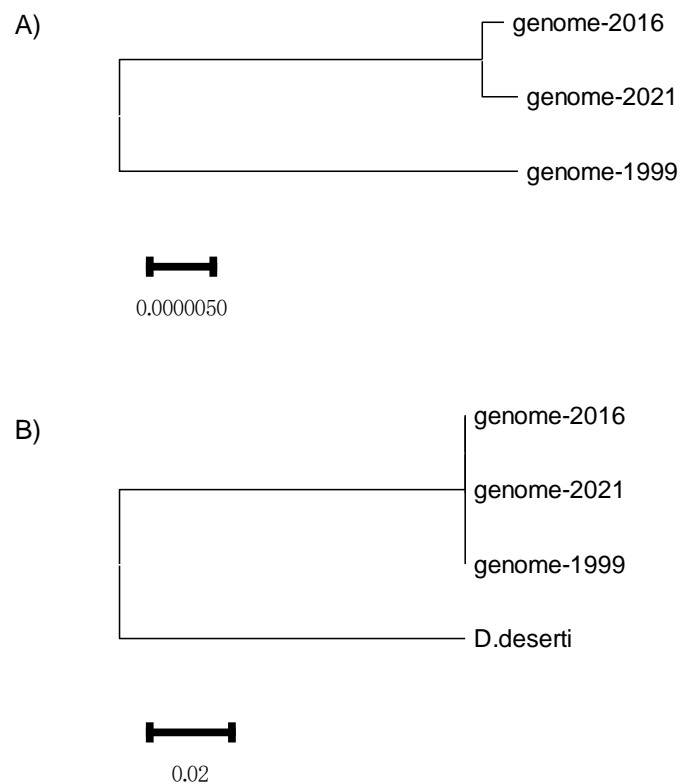


Figure 1. Phylogenies computed from the complete chromosome I based on the distance d_a ([31], see Methods) between the (A) *Deinococcus radiodurans* genomes used in this paper and (B) the same genomes with the *Deinococcus deserti* genome as an outgroup.

A comparison of the genomes of three *D. radiodurans* strains revealed changes in the genome structure that we characterized and classified into four types according to their probable mechanism of origin (Figure 2).

First, the strains differ in the transposition/copy number of mobile interspersed repetitive sequences (Tables 1 and 2). We detected the movement of two types of mobile interspersed repetitive sequences in the genome of *D. radiodurans*: insertion sequences (ISs) and small non-coding repeats (SNRs). Most of these were detected when comparing genome-1999 and genome-2021 (Table 1). In total, there were only two additional structural changes when comparing genome-2021 to genome-2016, i.e., two additional transpositions (Table 2). Of the 12 families of ISs previously found in genome-1999 [9,10], we detected positional changes of five, namely, ISDra2, ISDra3, ISDra5, ISDra6, and IS2621 (Tables 1 and 2). All five of these ISs have been previously shown to be transpositionally active [38,39]. Some of these IS changes do not include a variation in the copy number—e.g., one copy of ISDra3 is found in a different position in genome-1999 when compared to the genome-2021 (Table 1). However, some of the changes do include a variation in the copy number—e.g., there are six more copies of ISDra2 in genome-1999 than in genome-2021 (Table 1). In such cases, it is unclear whether ISs were deleted in one or have multiplied in the other genome. Of the nine types of SNRs previously found in genome-1999 [9,10], we

detected an additional copy of SNR4 when comparing it to genome-2021, demonstrating a copy number change and potential mobility of SNR4 (Table 1).

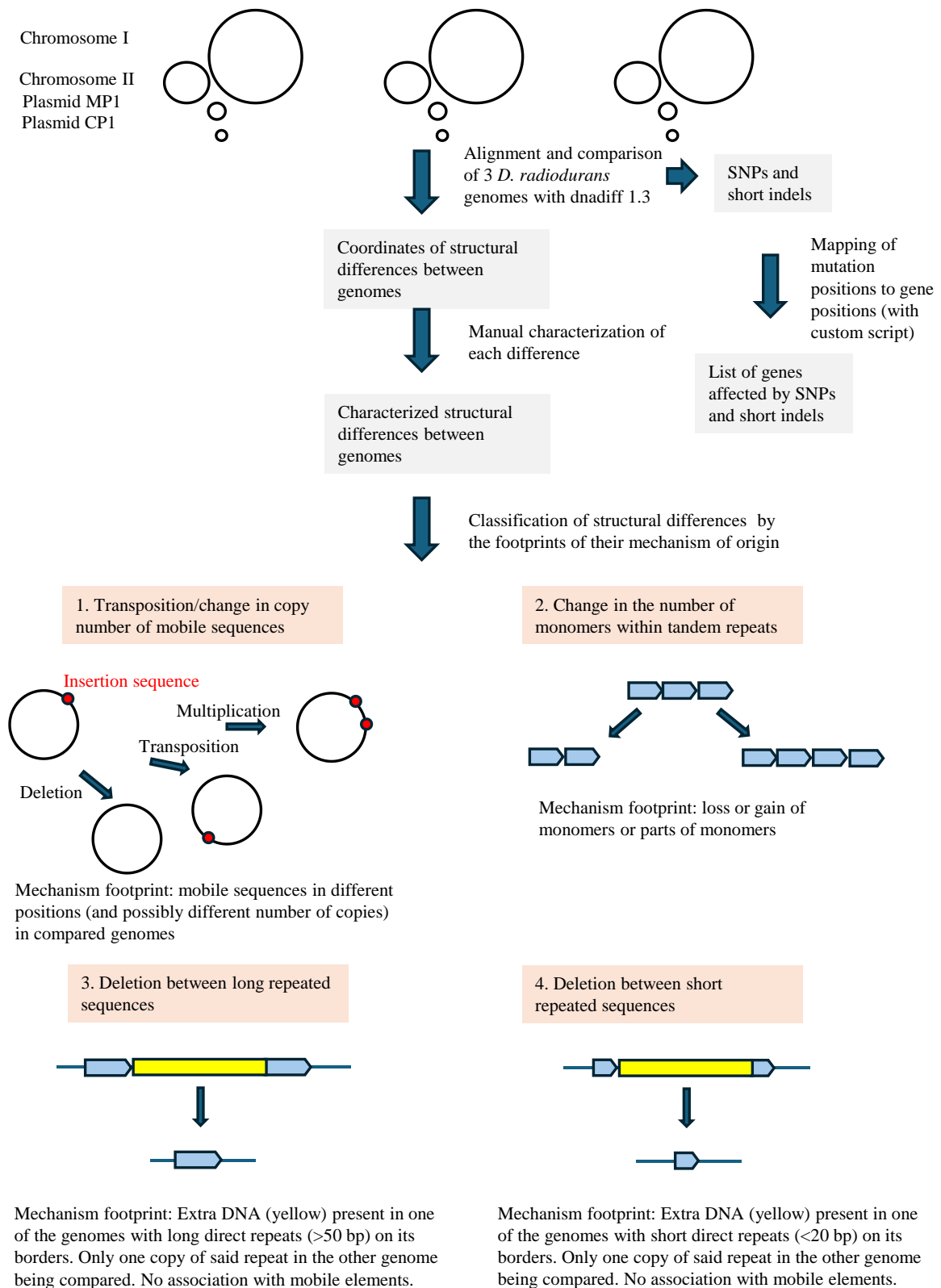


Figure 2. Workflow of our analysis, with graphical representations of detected types of structural genome change in *D. radiodurans*.

Table 1. List of mobile interspersed repetitive sequences—insertion sequences (ISs) and small non-coding repeats (SNRs) differently positioned when comparing genome-1999 and genome-2021. The list is ordered by position on the genome element. Extra DNA means DNA (in this case, a copy of IS or SNR) present in one genome but not the other.

IS/ SNR	The Investigated Genome in Which Extra DNA Is Present	Extra DNA Starts at bp:	Genome Element	Extra DNA Overlaps an Annotated Gene at Its Edge (Location and GenBank Annotation *); Genes Are Numbered for Clarity as There Are Instances of Multiple Genes Being Affected.
ISDra3	genome-1999	252969	Chromosome I	[i]<1..>1070,23S ribosomal RNA, DR_r02
ISDra2/ IS8301	genome-1999	674926	Chromosome I	None
ISDra6	genome-2021	694904	Chromosome I	[i]<1..16, E5E91_03580, hypothetical protein
IS2621	genome-1999	881456	Chromosome I	[i]<1..31, hypothetical protein, DR_0869
SNR4	genome-2021	969707	Chromosome I	None
ISDra2/ IS8301 **	genome-1999	994752	Chromosome I	[i]<1..70, DR_0977, AAF10554.1, phosphoenolpyruvate carboxykinase, [ii]1705..>1743, DR_0980, AAF10557.1, glutamate dehydrogenase, putative, loss of 6 kb DNA
IS2621	genome-1999	1337894	Chromosome I	None
ISDra2/ IS8301	genome-1999	1387636	Chromosome I	[i]<1..71, DR_1380, AAF10959.1, hypothetical protein
ISDra2/ IS8301	genome-1999	1611598	Chromosome I	[i]complement(1675..>1745), DR_1594, AAF11161.1, hypothetical protein
IS2621	genome-1999	1638129	Chromosome I	[i]complement(1302..>1322), DR_1619, conserved hypothetical protein
ISDra3	genome-2021	1745642	Chromosome I	[i]complement(<1), E5E91_08855, incomplete, partial on complete genome, missing N-terminus, DUF4238 domain-containing protein
ISDra2/ IS8301	genome-1999	1953199	Chromosome I	[i]complement(<1..38), DR_1931, conserved hypothetical protein, [ii]complement(1673..>1742), DR_1934, hypothetical protein
ISDra2/ IS8301	genome-1999	2321426	Chromosome I	[i] complement(<1..38), DR_2322, serine protease, subtilase family, C-terminal fragment, [ii]complement(1673..>1742), DR_2325, serine protease, subtilase family, N-terminal fragment
IS2621	genome-2021	374859	Chromosome II	[i] <1..3, E5E91_15400, incomplete, partial on the complete genome, missing C-terminus, tetratricopeptide repeat protein

* The GenBank annotation includes the position of the affected gene in a specific GenBank format, which is best understood through examples, e.g., 1302..>1322. In this example, a region with a length of 1322 bp is investigated. Within the region, there is a gene with a starting position at nucleotide 1302, and the gene extends through (“..”) and beyond (“>”) the last nucleotide of the investigated region, which is at position 1322. ** The appearance of this IS in the new position is correlated with a long (>6 kb) deletion in the same position.

Table 2. List of mobile interspersed repetitive sequences—insertion sequences (ISs) differently positioned when comparing genome-2016 and genome-2021. The list is ordered by the position on the genome element. Extra DNA means DNA (in this case, a copy of IS) present in one genome but not the other.

IS	The Investigated Genome in Which Extra DNA Is Present	Extra DNA Starts at bp:	Genome Element	Extra DNA Overlaps an Annotated Gene at Its Edge
ISDra5	genome-2016	656554	Chromosome I	None
IS2621	genome-2016	1753516	Chromosome I	None

The second type of change in the genome structure that we detected is a variable number of monomers within tandem repeats (Table 3). Generally, tandem repeats consist of different numbers of adjacent copies of a monomer sequence. The multiplication/deletion of monomers probably occurs by misalignments during replication [37], but given the length of monomers in Table 3, deletions of monomers could also conceivably occur through the SSA mechanism of DSB repair or A-EJ.

Table 3. Features of detected differences in tandem repeats between the two genomes and their statistics, reported by Tandem Repeats Finder [36].

Monomer Size	Copy Number in Genome-1999	Starting Position (bp) in Genome-1999	Copy Number in Genome-2021	Starting Position (bp) in Genome-2021	Percent Matches, Percent Indels	Genome Element
24	3.1	1118441	5.1	1119952	100,0	Chromosome I
21	1	2457333	3.9	2453220	79,9	Chromosome I
218	2	45488	1	1	97,0	Plasmid CP1

The third type of change in the genome structure that we detected is deletions between long interspersed direct repeats, in which one copy of the repeat was also deleted (Table 4). The probable mechanism of origin for these events is ESDSA or classical DSB repair via homologous recombination. However, the possibility of local misassembly in the genome-1999 sequence should also be considered (see Section 4).

The fourth type of change in genome structure that we detected is deletions between short (<20 bp) interspersed direct repeats, in which one copy of the repeat was also deleted (Table 5). The distinction between long and short DNA repeats was made on the basis of the implicated mechanism responsible for the structural change. We were specifically interested in repeats <20 bp as they imply the alternative end-joining (A-EJ) mechanism of DSB repair.

We also detected SNPs and short indels when comparing the three genomes. We mapped the SNPs and indels to the genes annotated in genome-1999 to create a resource for researchers studying specific genes in either one of the *D. radiodurans* strains. We report the full list of genes in Table S1 (for a comparison of genome-1999 vs. genome-2021) and Table S2 (for a comparison of genome-1999 vs. genome-2016), and the summary of annotations of affected genes is given in Table 6. We also included several proteins of general interest in Table 6. A frameshift SNP in the SSB gene (Table 6), as well as one in the neighboring intergenic sequence (Table S1), was previously reported as a misassembly in genome-1999 [40].

Table 4. Features of deletions associated with long repetitive sequences.

Deletion— Length of Unique Sequence Present in Only One Genome (bp)	Genome with Unique Non-Deleted DNA	Genome Element	The Position (bp) at Which Unique Non-Deleted DNA Starts	Repetitive Sequence at Deletion Boundaries (L and R)	Unique Non-Deleted DNA Present in Only One of the Two Investigated Genomes Contains Genes (Location and GenBank Annotations *). Genes Are Numbered for Clarity as There Are Instances of Multiple Genes Being Affected.
6065	genome-2021	Chromosome I	991922	Special case: extra DNA exchanged with one IS sequence (IS2621) (see Section 4)	[i]<1..276, E5E91_05075, phosphoenolpyruvate carboxykinase, [ii] 492..1343, E5E91_05080, hypothetical protein, [iii] 1734..2879, pdhA, E5E91_05085, pyruvate dehydrogenase (acetyl-transferring) E1 component subunit α , [iv] 3132..4157, E5E91_05090, α -ketoacid dehydrogenase subunit β , [v] 4404..5663, E5E91..05095, Glu/Leu/Phe/Val dehydrogenase, [vi] 5728..>6065, E5E91_05100, Glu/Leu/Phe/Val dehydrogenase
24	genome-1999	Chromosome I	1960529	526 bp	[i]complement(<1..>24), DR_1939, putative; polyphosphate kinase, authentic frameshift
917	genome-1999	Chromosome II	288082	600 bp	[i]complement(<1..331), DR_A0268, adenine deaminase-related protein, [ii]complement(524..880), DR_A0269, hypothetical protein
48	genome-1999	Megaplasmid	96697	59 bp	[i]complement(32..>48), DR_B0078, this region contains authentic frameshift

* The GenBank annotation includes the position of the affected gene in a specific GenBank format, which is best understood through examples, e.g., 1302..>1322. In this example, a region with a length of 1322 bp is investigated. Within the region, there is a gene with a starting position at nucleotide 1302, and the gene extends through (“..”) and beyond (“>”) the last nucleotide of the investigated region, which is at position 1322.

Table 5. Features of deletions associated with short repetitive sequences.

Deletion— Length of Unique Sequence (bp)	Unique Non-Deleted DNA in Strain	Genome Element Identity	The Position (bp) at Which Unique Non-Deleted DNA Starts	Repetitive Sequence at Deletion Boundaries (L and R)	Unique Non-Deleted DNA Present in Only One of the Two Investigated Genomes Contains Genes (Location and GenBank Annotations *). Genes Are Numbered for Clarity as There Are Instances of Multiple Genes Being Affected.
1900	genome-1999	Chromosome I	1230616	CGGC	[i]288..1043, E5E91_06355, “regulator”, [ii]1040..>1900, E5E91_06360, EamA family transporter

Table 5. Cont.

Deletion— Length of Unique Sequence (bp)	Unique Non-Deleted DNA in Strain	Genome Element Identity	The Position (bp) at Which Unique Non-Deleted DNA Starts	Repetitive Sequence at Deletion Boundaries (L and R)	Unique Non-Deleted DNA Present in Only One of the Two Investigated Genomes Contains Genes (Location and GenBank Annotations *). Genes Are Numbered for Clarity as There Are Instances of Multiple Genes Being Affected.
147	genome-1999	Chromosome I	1234925	CAGGCGGCGC	[i]<1..>147, E5E91..06375, glycosyltransferase
1178	genome-1999	Chromosome I	1819776	ACCCAGCGGG	[i]74..922, E5E91_09235, hypothetical protein

* The GenBank annotation includes the position of the affected gene in a specific GenBank format which is best understood through examples, e.g., 1302..>1322. In this example, a region with a length of 1322 bp is investigated. Within the region, there is a gene with a starting position at nucleotide 1302, and the gene extends through (“..”) and beyond (“>”) the last nucleotide of the investigated region, which is at position 1322.

Table 6. Summary of annotations of genes affected by SNPs and short indels between the two investigated *D. radiodurans* strains. The reported SNPs and indels have been mapped to the gene annotations in genome-1999 and summarized. Separate from the summarized categories of proteins, we included 7 proteins of general interest also affected by detected SNPs/indels (DNA topoisomerase I, FtsE, elongation factor TU, DNA polymerase III (β subunit), FtsK, SSB, and SbcC).

Annotation of the Gene with the SNP/Indel	Number of Positions Affected by SNP/Indel in Genome-1999 Compared to Genome-2021	Number of Positions Affected by SNP/Indel in Genome-1999 Compared to Genome-2016
Hypothetical protein	147	190
Conserved hypothetical protein	33	42
Transposase (putative)	49	49
DNA-binding response regulator	3	2
Intergenic SNPs	189	240
DNA topoisomerase I	1	1
FtsE	3	3
Elongation factor TU	1	0
DNA polymerase III, β subunit	1	1
Cell division protein FtsK (putative)	5	5
Single-stranded DNA-binding protein	1	1
Exonuclease SbcC	11	11
Other	114	135
Total	558	680

4. Discussion

The detection of deletions that have occurred during the divergence of *D. radiodurans* strains in different laboratories is a surprising finding in a bacterium thought to be extremely DSB-repair proficient. This finding contradicts our hypothesis that we would not find any traces of inaccurate DSB repair in these closely related strains.

Moreover, the deletions happen not only between long repetitive sequences but also between short ones (4–10 bp). Previously, deletions between short (4–11 bp) repeated sequences have only been shown after the inactivation of more accurate, RecA-dependent, and long homology-based mechanisms of DSB repair (within the genome-2021 strain [6,17]). However, our present results suggest that such deletions might occur even after spontaneous DSBs in fully repaired proficient cells. This is surprising because deletions between short repeated sequences imply the activity of a less accurate mechanism of DSB repair (A-EJ or SSA, with the shortness of homology implying A-EJ [17,18,21]) in conditions in which a more accurate mechanism (e.g., ESDSA or HR) should be both available and preferred. Indeed, strains used in this research are highly radiation-resistant, showing high capacity for accurate DSB repair.

We can envision the following explanations for the surprising use of a less accurate DSB repair mechanism after spontaneous DSBs: (i) some laboratory storage conditions may be conducive to DNA damage and other forms of stress (e.g., storage in stabs [25]) and/or (ii) a less accurate DSB repair mechanism may be a significant part of the *D. radiodurans* toolbox for genome repair and maintenance (e.g., see [14,41]).

The transposition/multiplication of ISs is common in bacteria [42] and, therefore, less unexpected, even in *D. radiodurans*. Most of the differently positioned ISs that we detected were inserted into chromosome I (14/15), consistent with the previous observation that chromosome I has the most ISs per bp [9,10]. Notably, a higher copy number of ISDra2 in genome-1999 (Table 1) is possibly due to the expansion of ISDra2, which can be induced by DNA damage [38]. Alternatively, some of the copies might have been deleted in genome-2021.

The first event in Table 4 is a special case of deletion associated with an IS sequence: one transposed sequence of ISDra2 in genome-1999 replaces a ~6 kb sequence present in genome-2021. This deletion is likely a consequence of the ISDra2 transposition; IS elements are known to sometimes cause genome instability [42]. ISDra2 transposes to the right of the target sequence 5-TTGAT-3 [43], which is also the case here. The host sequence on the other side of the transposed ISDra2 varies widely [43], indicating that there are no special sequence requirements for the attachment of the right end of the IS sequence. As this IS sequence requires DNA to be single-stranded during transposition [44], it is easy to imagine how secondary or tertiary structures of single-stranded host DNA might cause the IS to jump over and thus delete some host DNA. However, it is also possible that two ISDra2 transpositions occurred on either side of the ~6 kb deletion fragment and that the deletion between them was mediated by homologous recombination. Interestingly, a similar event was reported previously in *Mycobacterium tuberculosis* [45]. Other than this event, no other deletion was associated with an IS or SNR mobile sequence.

Surprisingly, the length of the deleted DNA in either of the compared genomes varied by up to 6065 bp, and many deletions encompassed coding sequences, including several metabolic genes (see Tables 4 and 5). The effect of the deletion of these genes in the laboratory conditions is unclear. We speculate that it might slightly speed up the replication; the streamlining of bacterial genomes through deletions is expected as an effect of selection pressure for faster growth [46]. Interestingly, the inactivation of some metabolic genes and subsequent accumulation of metabolic intermediates might hypothetically confer higher resistance to γ radiation/desiccation, which happened in *E. coli* subjected to selection pressure for higher radiation resistance [47]. However, it is also possible that the inactivation of a few genes does not significantly influence doubling time and/or radiation survival in the rich medium typically used for *D. radiodurans* growth and experiments. The lack of functional analysis is the main limitation of the present study as we predominantly focus

on structural differences between the genomes. The impact of observed genome differences on function remains to be elucidated in future studies.

When investigating differences between phylogenetically close genome sequences such as the ones presented in this paper, it should be considered whether identified genome differences that involve long repeated sequences might be due to the misassembly of reads during the genome sequencing process. Ironically, the misassembly at long repeated sequences will closely resemble genuine cell events of recombination between ectopic repetitive sequences. Therefore, the third event in Table 3, as well as events 2–4 in Table 4, should be approached with caution. They all involve long repeats in genome-1999, which was sequenced and assembled from reads that were on average much shorter [29] than the PacBio reads used for genome-2021 [17]. Nonetheless, we do not expect the deletion events between short repeats to be due to misassembly as the average length of reads used for the assembly of genome-1999 greatly surpasses the length of repeats in Table 5.

The results of this work showed that *D. radiodurans* strains of nominally the same genotype can hide many genetic differences, including genomic rearrangements. These differences could have arisen over decades due to different ways of storing the same strains in several laboratories around the world. Strains are now stored frozen at $-70\text{ }^{\circ}\text{C}$ or $-80\text{ }^{\circ}\text{C}$ as a standard procedure. This form of storage ensures both the long-term viability of the strains and the stability of their genomes. In 1956, when *D. radiodurans* was first isolated, this was not the standard. Presumably, agar stab cultures or freezing at $-20\text{ }^{\circ}\text{C}$ was predominately used at that time. Storage at $-20\text{ }^{\circ}\text{C}$ is not ideal because the cells lose their viability after a certain time, which then requires more frequent replenishment of stocks and regrowth of strains. It is possible that more frequent cycles of regrowth and freezing open space for spontaneous genomic changes. On the other hand, storage in stabs enables a long lifespan in the stationary phase, which is not an inert state as once thought, but can be the source of various changes in the genome [48]. Therefore, researchers in the field of bacterial genetics and genomics must be constantly aware of the dynamism of the bacterial genome and the possibility that genomic changes can take place at any moment, even in seemingly non-stressful laboratory conditions.

5. Conclusions

Notably, some types of genome structural changes between the three investigated genomes were more expected (transposition/multiplication of ISs and change in length of tandem repeats), and some were unexpected in a bacterium as proficient in DSB repair as *D. radiodurans* (long deletions, especially between short repeated sequences) and indicate the use of a less accurate DSB repair mechanism in conditions in which a more accurate one should be both available and preferred.

Our results demonstrate that presuming isogeneity between commonly shared bacterial strains is not always safe and investigating differences between them can give us new information. Information on common structural changes and their mechanisms of origin in the genomes of bacteria propagated in the laboratory will help navigate further research in the field of structural change and genome rearrangements, especially in the models of DSB repair such as *D. radiodurans*. In addition to being important in terms of inferences about deinococcal DNA metabolism and repair, the detected genome changes are a valuable resource for the researchers using the investigated *D. radiodurans* strains.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes15070847/s1>, Table S1: Annotations of genes affected by SNPs and indels between two *D. radiodurans* strains (genome-1999 was used as the reference DNA sequence for this comparison, and genome-2021 was used as the query). Table S2: Annotations of genes affected by SNPs and indels between two *D. radiodurans* strains (genome-1999 was used as the reference DNA sequence for this comparison, and genome-2016 was used as the query).

Author Contributions: Conceptualization, K.Z. and J.R.; methodology, J.R.; formal analysis, J.R., K.Z. and D.Z.; data curation, J.R.; software, J.R.; writing—original draft preparation, J.R.; writing—review and editing, K.Z. and D.Z.; project administration, K.Z.; and funding acquisition, K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Croatian Science Foundation project IP-2022-10-7476.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Acknowledgments: We thank Tin Klanjšček for useful comments during manuscript preparation.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Battista, J.R. Against All Odds: The survival strategies of *Deinococcus radiodurans*. *Annu. Rev. Microbiol.* **1997**, *51*, 203–224. [[CrossRef](#)] [[PubMed](#)]
- Cox, M.M.; Battista, J.R. *Deinococcus radiodurans*—The consummate survivor. *Nat. Rev. Microbiol.* **2005**, *3*, 882–892. [[CrossRef](#)] [[PubMed](#)]
- Slade, D.; Radman, M. Oxidative stress resistance in *Deinococcus radiodurans*. *Microbiol. Mol. Biol. Rev.* **2011**, *75*, 133–191. [[CrossRef](#)] [[PubMed](#)]
- Daly, M.J. The scientific revolution that unraveled the astonishing DNA repair capacity of the Deinococcaceae: 40 years on. *Can. J. Microbiol.* **2023**, *69*, 369–386. [[CrossRef](#)] [[PubMed](#)]
- Repar, J.; Supek, F.; Klanjšček, T.; Warnecke, T.; Zahradka, K.; Zahradka, D. Elevated rate of genome rearrangements in radiation-resistant bacteria. *Genetics* **2017**, *205*, 1677–1689. [[CrossRef](#)] [[PubMed](#)]
- Repar, J.; Cvjetan, S.; Slade, D.; Radman, M.; Zahradka, D.; Zahradka, K. RecA Protein assures fidelity of DNA repair and genome stability in *Deinococcus radiodurans*. *DNA Repair* **2010**, *9*, 1151–1161. [[CrossRef](#)] [[PubMed](#)]
- Shen, P.; Huang, H.V. Homologous recombination in *Escherichia coli*: Dependence on substrate length and homology. *Genetics* **1986**, *112*, 441–457. [[CrossRef](#)] [[PubMed](#)]
- Kowalczykowski, S.C.; Dixon, D.A.; Eggleston, A.K.; Lauder, S.D.; Rehrauer, W.M. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* **1994**, *58*, 401–465. [[CrossRef](#)] [[PubMed](#)]
- Makarova, K.S.; Wolf, Y.I.; White, O.; Minton, K.; Daly, M.J. Short repeats and IS elements in the extremely radiation-resistant bacterium *Deinococcus radiodurans* and comparison to other bacterial species. *Res. Microbiol.* **1999**, *150*, 711–724. [[CrossRef](#)]
- Makarova, K.S.; Aravind, L.; Wolf, Y.I.; Tatusov, R.L.; Minton, K.W.; Koonin, E.V.; Daly, M.J. Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 44–79. [[CrossRef](#)]
- Mattimore, V.; Battista, J.R. Radioresistance of *Deinococcus radiodurans*: Functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *J. Bacteriol.* **1996**, *178*, 633–637. [[CrossRef](#)] [[PubMed](#)]
- Repar, J.; Zahradka, D.; Zahradka, K. Accuracy of genome reassembly in γ -irradiated *Escherichia coli*. *Food Technol. Biotechnol.* **2013**, *51*, 327–337.
- Zahradka, K.; Slade, D.; Bailone, A.; Sommer, S.; Averbeck, D.; Petranovic, M.; Lindner, A.B.; Radman, M. Reassembly of shattered chromosomes in *Deinococcus radiodurans*. *Nature* **2006**, *443*, 569–573. [[CrossRef](#)] [[PubMed](#)]
- Daly, M.J.; Minton, K.W. An alternative pathway of recombination of chromosomal fragments precedes RecA-dependent recombination in the radioresistant bacterium *Deinococcus radiodurans*. *J. Bacteriol.* **1996**, *178*, 4461–4471. [[CrossRef](#)] [[PubMed](#)]
- Xu, G.; Lu, H.; Wang, L.; Chen, H.; Xu, Z.; Hu, Y.; Tian, B.; Hua, Y. DdrB stimulates single-stranded DNA annealing and facilitates RecA-independent DNA repair in *Deinococcus radiodurans*. *DNA Repair* **2010**, *9*, 805–812. [[CrossRef](#)]
- Ithurbide, S.; Bentchikou, E.; Coste, G.; Bost, B.; Servant, P.; Sommer, S. Single strand annealing plays a major role in RecA-independent recombination between repeated sequences in the radioresistant *Deinococcus radiodurans* bacterium. *PLoS Genet.* **2015**, *11*, e1005636. [[CrossRef](#)]
- Repar, J.; Zahradka, D.; Sović, I.; Zahradka, K. Characterization of gross genome rearrangements in *Deinococcus radiodurans* recA Mutants. *Sci. Rep.* **2021**, *11*, 10939. [[CrossRef](#)]
- McVey, M.; Lee, S.E. MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. *Trends Genet. TIG* **2008**, *24*, 529–538. [[CrossRef](#)] [[PubMed](#)]
- Sfeir, A.; Symington, L.S. Microhomology-mediated end joining: A back-up survival mechanism or dedicated pathway? *Trends Biochem. Sci.* **2015**, *40*, 701–714. [[CrossRef](#)]
- Bhargava, R.; Onyango, D.O.; Stark, J.M. Regulation of single-strand annealing and its role in genome maintenance. *Trends Genet. TIG* **2016**, *32*, 566–575. [[CrossRef](#)]

21. Sallmyr, A.; Tomkinson, A.E. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *J. Biol. Chem.* **2018**, *293*, 10536–10546. [[CrossRef](#)] [[PubMed](#)]
22. Mei, Q.; Fitzgerald, D.M.; Liu, J.; Xia, J.; Pribis, J.P.; Zhai, Y.; Nehring, R.B.; Paiano, J.; Li, H.; Nussenzweig, A.; et al. Two mechanisms of chromosome fragility at replication-termination sites in bacteria. *Sci. Adv.* **2021**, *7*, eabe2846. [[CrossRef](#)] [[PubMed](#)]
23. Mehta, A.; Haber, J.E. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016428. [[CrossRef](#)]
24. Sinha, A.K.; Possoz, C.; Leach, D.R.F. The roles of bacterial DNA double-strand break repair proteins in chromosomal DNA replication. *FEMS Microbiol. Rev.* **2020**, *44*, 351–368. [[CrossRef](#)] [[PubMed](#)]
25. Desroches, M.; Royer, G.; Roche, D.; Mercier-Darty, M.; Vallenet, D.; Médigue, C.; Bastard, K.; Rodriguez, C.; Clermont, O.; Denamur, E.; et al. The odyssey of the ancestral Escherich strain through culture collections: An example of allopatric diversification. *mSphere* **2018**, *3*, e00553-17. [[CrossRef](#)] [[PubMed](#)]
26. Dorman, M.J.; Thomson, N.R. “Community evolution”—Laboratory strains and pedigrees in the age of genomics. *Microbiol. Read. Engl.* **2020**, *166*, 233–238. [[CrossRef](#)] [[PubMed](#)]
27. Bleibtreu, A.; Clermont, O.; Darlu, P.; Glodt, J.; Branger, C.; Picard, B.; Denamur, E. The rpoS gene is predominantly inactivated during laboratory storage and undergoes source-sink evolution in *Escherichia coli* species. *J. Bacteriol.* **2014**, *196*, 4276–4284. [[CrossRef](#)] [[PubMed](#)]
28. Anderson, A.W.; Nordan, H.C.; Cain, R.; Parrish, G.; Duggan, D.; Anderson, A.O.; Nordan, H.; Parish, G.; Cullum-Dugan, D. Studies on a radio-resistant micrococcus. I, Isolation, morphology, cultural characteristics, and resistance to gamma radiation. *Food Technol.* **1956**, *10*, 575–578.
29. White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Haft, D.H.; Gwinn, M.L.; Nelson, W.C.; Richardson, D.L.; et al. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **1999**, *286*, 1571–1577. [[CrossRef](#)]
30. Hua, X.; Hua, Y. Improved complete genome sequence of the extremely radioresistant bacterium *Deinococcus radiodurans* R1 obtained using PacBio single-molecule sequencing. *Genome Announc.* **2016**, *4*, e00886-16. [[CrossRef](#)]
31. Haubold, B.; Klötzl, F.; Pfaffelhuber, P. Andi: Fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics* **2015**, *31*, 1169–1175. [[CrossRef](#)] [[PubMed](#)]
32. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [[CrossRef](#)] [[PubMed](#)]
33. Marçais, G.; Delcher, A.L.; Phillippy, A.M.; Coston, R.; Salzberg, S.L.; Zimin, A. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **2018**, *14*, e1005944. [[CrossRef](#)] [[PubMed](#)]
34. Siguier, P.; Perochon, J.; Lestrade, L.; Mahillon, J.; Chandler, M. ISfinder: The reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **2006**, *34*, D32–D36. [[CrossRef](#)] [[PubMed](#)]
35. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
36. Benson, G. Tandem Repeats Finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
37. Warburton, P.E.; Giordano, J.; Cheung, F.; Gelfand, Y.; Benson, G. Inverted repeat structure of the human genome: The X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.* **2004**, *14*, 1861–1869. [[CrossRef](#)]
38. Mennecier, S.; Servant, P.; Coste, G.; Bailone, A.; Sommer, S. Mutagenesis via IS transposition in *Deinococcus radiodurans*. *Mol. Microbiol.* **2006**, *59*, 317–325. [[CrossRef](#)]
39. Long, H.; Kucukyildirim, S.; Sung, W.; Williams, E.; Lee, H.; Ackerman, M.; Doak, T.G.; Tang, H.; Lynch, M. Background mutational features of the radiation-resistant bacterium *Deinococcus radiodurans*. *Mol. Biol. Evol.* **2015**, *32*, 2383–2392. [[CrossRef](#)]
40. Eggington, J.M.; Haruta, N.; Wood, E.A.; Cox, M.M. The single-stranded DNA-binding protein of *Deinococcus radiodurans*. *BMC Microbiol.* **2004**, *4*, 2. [[CrossRef](#)]
41. Lovett, S.T. Encoded errors: Mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* **2004**, *52*, 1243–1253. [[CrossRef](#)] [[PubMed](#)]
42. Siguier, P.; Gourbeyre, E.; Chandler, M. Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiol. Rev.* **2014**, *38*, 865–891. [[CrossRef](#)] [[PubMed](#)]
43. Islam, S.M.; Hua, Y.; Ohba, H.; Satoh, K.; Kikuchi, M.; Yanagisawa, T.; Narumi, I. Characterization and distribution of IS8301 in the radioresistant bacterium *Deinococcus radiodurans*. *Genes Genet. Syst.* **2003**, *78*, 319–327. [[CrossRef](#)] [[PubMed](#)]
44. Hickman, A.B.; James, J.A.; Barabas, O.; Pasternak, C.; Ton-Hoang, B.; Chandler, M.; Sommer, S.; Dyda, F. DNA recognition and the precleavage state during single-stranded DNA transposition in *D. radiodurans*. *EMBO J.* **2010**, *29*, 3840–3852. [[CrossRef](#)] [[PubMed](#)]
45. Fang, Z.; Doig, C.; Kenna, D.T.; Smittipat, N.; Palittapongarnpim, P.; Watt, B.; Forbes, K.J. IS6110-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. *J. Bacteriol.* **1999**, *181*, 1014–1020. [[CrossRef](#)] [[PubMed](#)]
46. Rocha, E.P.C. The replication-related organization of bacterial genomes. *Microbiology* **2004**, *150*, 1609. [[CrossRef](#)]

-
47. Harris, D.R.; Pollock, S.V.; Wood, E.A.; Goiffon, R.J.; Klingele, A.J.; Cabot, E.L.; Schackwitz, W.; Martin, J.; Eggington, J.; Durfee, T.J.; et al. Directed evolution of ionizing radiation resistance in *Escherichia coli*. *J. Bacteriol.* **2009**, *191*, 5240–5252. [[CrossRef](#)]
 48. Kolter, R.; Siegele, D.A.; Tormo, A. The stationary phase of the bacterial life cycle. *Annu. Rev. Microbiol.* **1993**, *47*, 855–874. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.