

## Article

# Diversity, Distribution, and Chromosomal Rearrangements of TRIP1 Repeat Sequences in *Escherichia coli*

Zhan Li <sup>1</sup>, Xiong Liu <sup>2</sup>, Nianzhi Ning <sup>1</sup>, Tao Li <sup>1</sup> and Hui Wang <sup>1,\*</sup>

<sup>1</sup> State Key Laboratory of Pathogens and Biosecurity, Beijing Institute of Microbiology and Epidemiology, No. 20 Dongda Street, Fengtai District, Beijing 100071, China; yexi19881214@126.com (Z.L.); ningnianzhi@163.com (N.N.); litaobmi@126.com (T.L.)

<sup>2</sup> Chinese PLA Center for Disease Control and Prevention, Dongda Street 20#, Fengtai District, Beijing 100071, China; liuxiong714@163.com

\* Correspondence: geno0109@vip.sina.com

**Abstract:** The bacterial genome contains numerous repeated sequences that greatly affect its genomic plasticity. The *Escherichia coli* K-12 genome contains three copies of the TRIP1 repeat sequence (TRIP1a, TRIP1b, and TRIP1c). However, the diversity, distribution, and role of the TRIP1 repeat sequence in the *E. coli* genome are still unclear. In this study, after screening 6725 *E. coli* genomes, the TRIP1 repeat was found in the majority of *E. coli* strains (96%: 6454/6725). The copy number and direction of the TRIP1 repeat sequence varied in each genome. Overall, 2449 genomes (36%: 2449/6725) had three copies of TRIP1 (TRIP1a, TRIP1b, and TRIP1c), which is the same as *E. coli* K-12. Five types of TRIP1 repeats, including two new types (TRIP1d and TRIP1e), are identified in *E. coli* genomes, located in 4703, 3529, 5741, 1565, and 232 genomes, respectively. Each type of TRIP1 repeat is localized to a specific locus on the chromosome. TRIP1 repeats can cause intra-chromosomal rearrangements. A total of 156 rearrangement events were identified, of which 88% (137/156) were between TRIP1a and TRIP1c. These findings have important implications for future research on TRIP1 repeats.

**Keywords:** TRIP1 repeats; *Escherichia coli*; chromosomal rearrangement



**Citation:** Li, Z.; Liu, X.; Ning, N.; Li, T.; Wang, H. Diversity, Distribution, and Chromosomal Rearrangements of TRIP1 Repeat Sequences in *Escherichia coli*. *Genes* **2024**, *15*, 236. <https://doi.org/10.3390/genes15020236>

Academic Editor: Silvia Turrone

Received: 17 January 2024

Revised: 7 February 2024

Accepted: 10 February 2024

Published: 13 February 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

DNA repeats have been defined as sequences with high similarity to other sequences found in the same genome, which were first identified in a bacterial genome during the early 1980s [1]. It has been discovered that repeat sequences are a common feature of bacterial genomes. Bacterial genomes contain numerous repeat sequences that can be classified based on several criteria [2]. These include their direction (direct or inverted), size (simple short or long), distribution (contiguous or interspersed), and copy number with the genome. The large number of repeat sequences found in the bacterial genome has attracted much attention regarding their origin, distribution [3], rearrangement [4], and influence on the bacterial genome [5].

DNA repeat sequences have divergent functions in complex DNA processes, such as replication, recombination, repair, and transcription [6]. Through their involvement in these processes, DNA repeat sequences influence the genetic instability and evolution of DNA molecules [7]. DNA repeat sequences that are rich in GCs, as triple hydrogen bonds, could stabilize DNA structures to facilitate bacterial survival [8]. DNA repeats are inserted into the transcription units of surrounding sequences, affecting their transcription and translation [9]. The removal of these DNA sequences led to improper folding and instability of the protein structure [10]. Therefore, DNA repeat sequences play a critical role in an organism's physiology and cellular architecture.

DNA repeats vary in length from dinucleotides to thousands of nucleotides and are widespread in bacterial genomes. DNA repeats of the length of at least 100 nucleotides are

called long DNA repeats [11]. Long identical DNA repeats are often distributed in a non-random manner across bacterial genomes. Their distribution differs in the overall number, orientation, and proximity to the origin of replication. In addition to some functional DNA repeat sequences [12–14], many long DNA repeats have been found to be closely related to the structural evolution of bacterial genomes [15–18]. Long DNA repeats are major players in mediating rearrangements. Long DNA repeats can lead to intrachromosomal rearrangements by acting as substrates to the bacterial recombination machinery. Rearrangement results were found to depend on the location and direction of the associated repeats, which change the order of genes on the chromosome [15,18,19]. Certain mobile elements have also targeted DNA repeat sequences [20–22]. DNA repeat analysis in bacterial genomes can provide valuable information on genomic plasticity among microbial species; moreover, changes in genome structure affect function and biological processes [23–25]. The analysis can reveal important insights into the genetic makeup of these organisms.

TRIP1 repeat sequences are 146 to 273 bp in length, and three copies (TRIP1a, TRIP1b, and TRIP1c) of TRIP1 repeat sequences are found in the complete *Escherichia coli* K-12 strain genome sequence [26]. The three copies are all located at different chromosomal locations. TRIP1a is 31 bp downstream of the apparent *aqpZ* transcription terminator. TRIP1b is 41 bp downstream of the *putP* stop codon and is a partial form of TRIP1a, with 125 bp internal deletion. TRIP1c is 44 bp downstream of the *yohI* gene, in an opposite orientation relative to TRIP1a. The copy number, distribution, and influence of the TRIP1 repeat sequence on the *E. coli* genome are not very clear. In this study, a total of 6725 *E. coli* genomes were scanned for the presence of the TRIP1 repeat sequence, and the types, distribution, and influence of TRIP1 on the *E. coli* genome were investigated.

## 2. Materials and Methods

### 2.1. Genome Sequences

*E. coli* genome sequences were downloaded from the NCBI ftp site (<https://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/> accessed on 8 October 2016). In total, 6725 *E. coli* genome sequences were available for analysis and were downloaded using Wget V1.18 software (<https://eternallybored.org/misc/wget/> accessed on 8 October 2016).

### 2.2. Identification and Localization of the TRIP1 Repeat Sequences

The query sequence for the BLAST search consisted of all three copies of the TRIP1 (TRIP1a, TRIP1b, and TRIP1c) repeat sequence found in *E. coli* K-12 (NC\_000913.3). TRIP1a was located on the top strand from 915016 to 915281, TRIP1b from 1080855 to 1081000, and TRIP1c from 2229122 to 2229394. The gap opening cost was set to 5, and gap the extension was set to 2.

The BLAST search results were analyzed as follows. First, all matching sequences with length coverage greater than 80% were compared and clustered using USEARCH manual v8.1 software (available at <https://drive5.com/usearch/manual8.1/> accessed on 25 August 2017) [27]. Two previously undiscovered TRIP1 types (TRIP1d and TRIP1e) were identified in the chromosomes of other *E. coli* strains. TRIP1d was located on the top strand from 2251149 to 2251412 in *E. coli* 2012C-4502 (CP027440.1) and TRIP1e from 1232542 to 1232806 in *E. coli* RHB34-C04 (CP057178.1).

The five types of TRIP1 repeat sequences were compared and aligned using BioEdit 7.7 [28] and ClustalW 7.7 [29] software. Second, only the highest-scoring TRIP1 type was used for subsequent analysis when multiple types of TRIP1 repeat sequences were matched at the same position in a genome. Third, to determine the position of the TRIP1 repeat sequence in each genome, the 1000 bp DNA sequences flanking both sides were extracted and annotated using the KEGG gene database (<https://www.genome.jp/kegg/> accessed on 17 October 2023).

### 2.3. RNA Structure Prediction of the TRIP1 Repeat Sequences

The RNAfold WebServer (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWeb-Suite/RNAfold.cgi/> accessed on 27 October 2023) was used to depict the most stable folded RNAs hypothetically expressed by the TRIP1a, TRIP1b, TRIP1c, TRIP1d, or TRIP1e repeat sequences.

### 2.4. Identification of Chromosomal Arrangement Caused by TRIP1 Repeat Sequences

To determine the chromosomal arrangement resulting from the TRIP1 repeat sequences, all genome sequences were analyzed as follows. First, for each of the identified TRIP1 loci, the two flanking genes were searched in the entire genome sequence. If an arrangement event occurred between two of the loci, the relative position and distance of the four associated genes would change, generating a new permutation between the flanking genes. All genomes with novel permutations between the locus-flanking genes were recorded. Second, the rearrangement events between the complete genome sequences were analyzed using Mauve 2.3.0 software (<https://darlinglab.org/mauve/mauve/> accessed on 1 November 2023).

## 3. Results

### 3.1. The Distribution and Copy Number of TRIP1 Repeats in *E. coli* Genomes

TRIP1 repeat sequences have been found with three copies in *E. coli* K-12 [26]. With the development of high-throughput sequencing technology [30], thousands of *E. coli* genome sequences are now available. Our primary objective is to determine the copy number of TRIP1 in all these strains. Out of the total 6725 genomes analyzed, 6454 (96%) were identified as having the TRIP1 repeat sequences (Figure 1A). The copy number of the TRIP1 repeat sequences found in each genome varied, ranging from zero to five copies. Most genomes had three copies (2449: 36%), which is the same as *E. coli* K-12. Others had one copy (1521: 23%), two copies (1433: 21%), or four copies (1048: 16%). Only three genomes had five TRIP1 repeat sequences in their genomes. Out of the 6725 genomes, only 353 of them had complete genome sequences. TRIP1 repeat sequences were identified in all 353 complete genomes (Figure 1B). The distribution of TRIP1 copy numbers in the complete genomes was similar to that of the total genomes. Most of the complete genomes had three copies (164: 46%), while others had one copy (62: 18%), two copies (64: 18%), or four copies (62: 18%); only one complete genome had five TRIP1 repeat sequences. The copy number of TRIP1 repeat sequences varied in each genome, indicating that they were dynamically created and removed during the genome evolution of the *E. coli* genome.

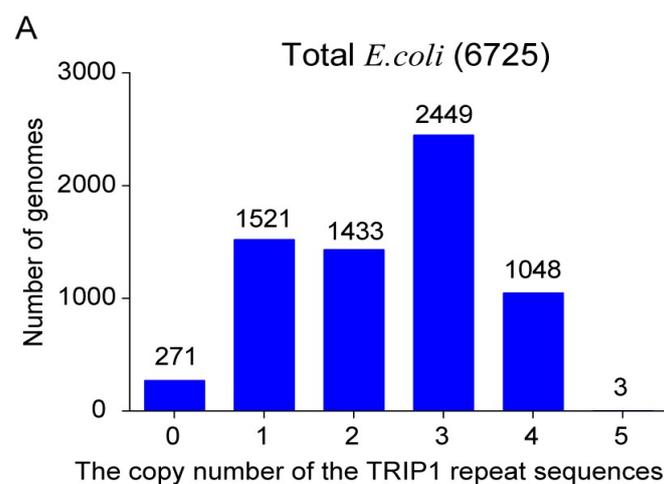
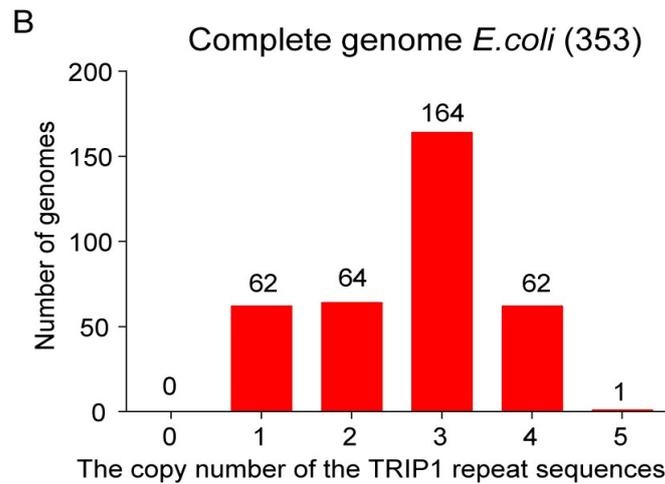


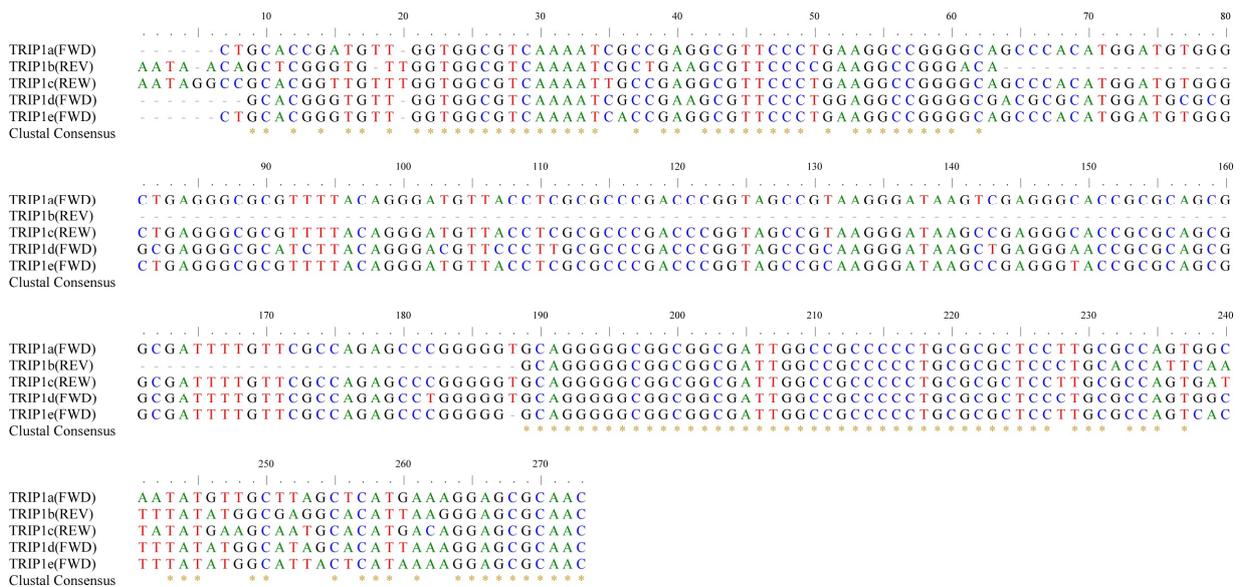
Figure 1. Cont.



**Figure 1.** The copy number of the TRIP1 repeat sequences in the *E. coli* genome. The TRIP1 repeat sequences were searched in 6725 available genomes, and the copy per genome was recorded. The distribution of TRIP1 repeat sequence copies is summarized for both total *E. coli* (A) and the complete genome of *E. coli* (B).

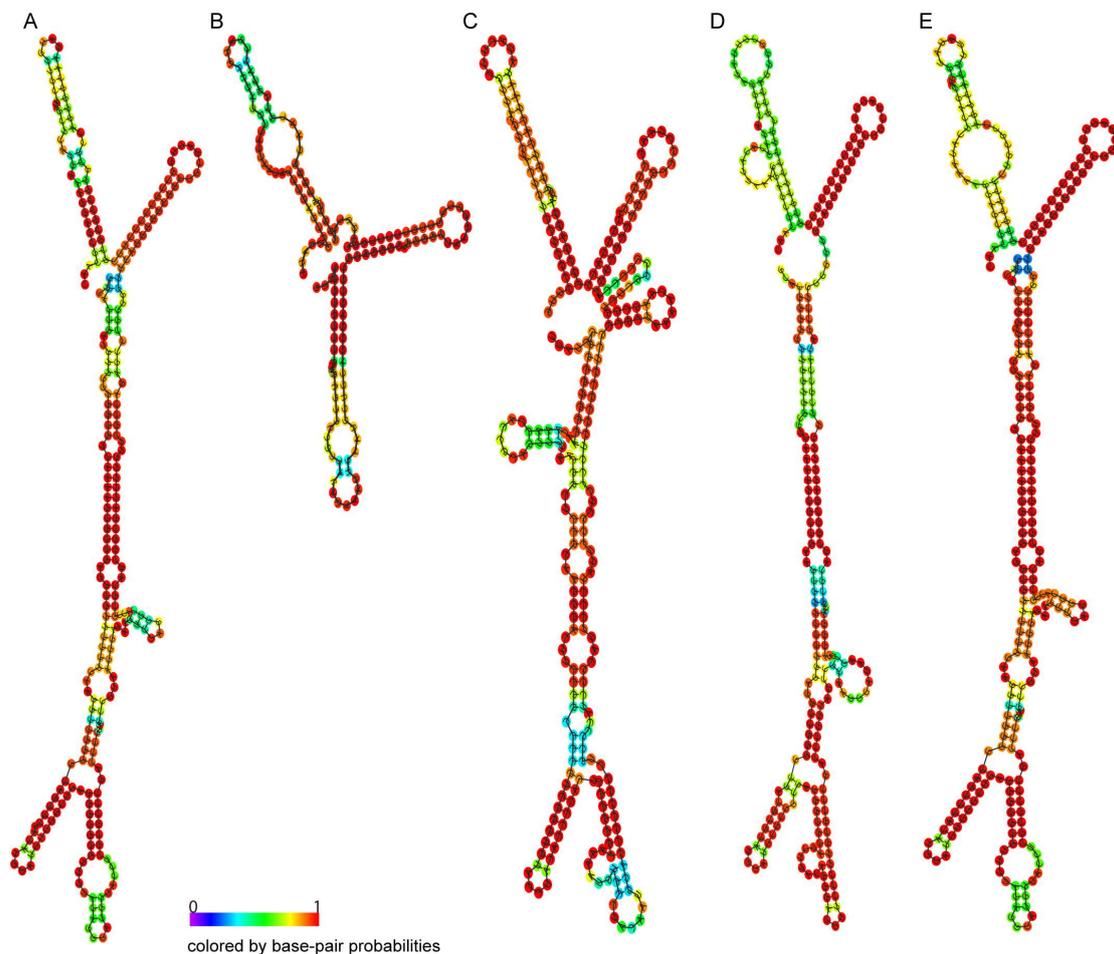
### 3.2. Distribution and Genome Localization of the Five Types of TRIP1 Repeats in *E. coli*

Further, we investigated the type and genomic localization of TRIP1 repeats in *E. coli*. The sequences of TRIP1 repeats were compared and analyzed, and the flanking sequences of TRIP1 repeats were extracted and annotated. Finally, we identified five types of TRIP1 repeats in *E. coli*, including three previously reported types (TRIP1a, TRIP1b, and TRIP1c) [26] and two newly discovered types (TRIP1d, and TRIP1e) (Figure 2). All five TRIP1 sequences are GC-rich (62–65% G + C). The TRIP1a repeat sequence is 266 bp in length. TRIP1b is 146 bp in length, which is a partial form of TRIP1a, sharing 84% identity with TRIP1a, and missing an internal 125 bp region. TRIP1c is 273 bp in length and has a 94% identity to TRIP1a. TRIP1d is 264 bp in length and has an 89% identity to TRIP1a. TRIP1e is 265 bp in length and has a 94% identity to TRIP1a.



**Figure 2.** Multiple alignment of the TRIP1 repeats family. The TRIP1a-1e repeat sequences were obtained from NCBI in FASTA format and compared to multiple sequences using BioEdit 7.7 software. The invariant bases are indicated with asterisks. “FWD” stands for the positive direction of gene transcription and “REW” stands for the reverse direction of gene transcription.

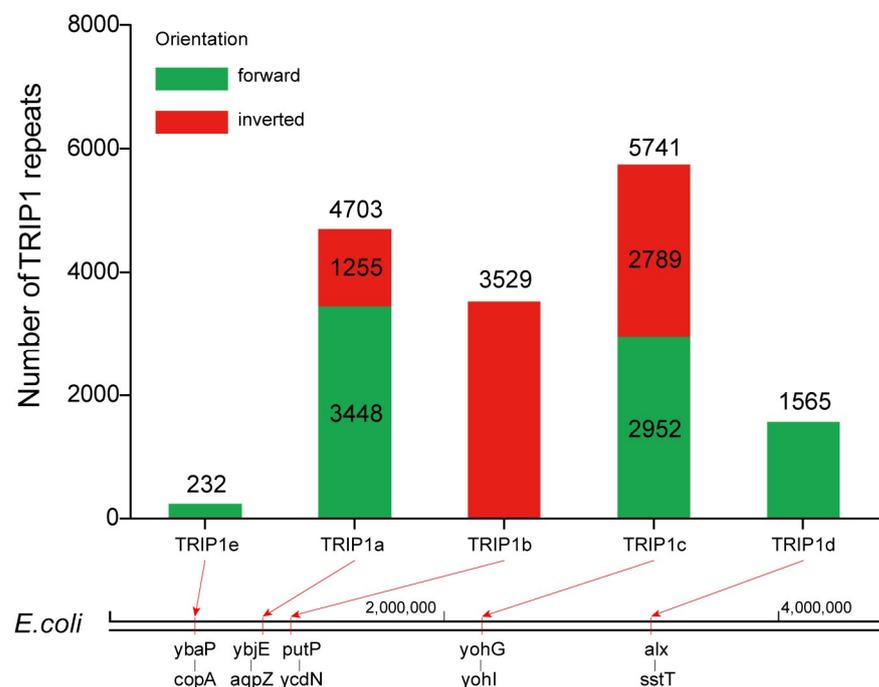
The RNA structure of the five types of TRIP1 repeats was predicted, with higher values indicating greater structural confidence (Figure 3A–E). All five TRIP1 sequences are capable of forming very stable secondary structures in RNA with a conserved double-Y structure. The free energy of the thermodynamic ensemble is predicted to be  $-143.99$  kcal/mol for TRIP1a (Figure 3A),  $-65.21$  kcal/mol for TRIP1b (Figure 3B),  $-131.36$  kcal/mol for TRIP1c (Figure 3C),  $-145.37$  kcal/mol for TRIP1d (Figure 3D), and  $-149.59$  kcal/mol for TRIP1e (Figure 3E).



**Figure 3.** The predicted RNA structure of TRIP1a (A), TRIP1b (B), TRIP1c (C), TRIP1d (D), and TRIP1e (E). The free energy of the thermodynamic ensemble is predicted to be  $-143.99$  kcal/mol for TRIP1a (A),  $-65.21$  kcal/mol for TRIP1b (B),  $-131.36$  kcal/mol for TRIP1c (C),  $-145.37$  kcal/mol for TRIP1d (D), and  $-149.59$  kcal/mol for TRIP1e (E). Centroid structure drawings encoding base-pair probabilities of TRIP1a-1e repeats were generated using the RNAfold website: <http://ma.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi/> accessed on 27 October 2023.

The majority of the TRIP1 repeats (15,770/15,941, 99%) were found in five loci. Each type of TRIP1 repeat is localized to a specific chromosomal position (Figure 4). TRIP1a is located 158 bp upstream of the *ybjE* gene start codon and 70 bp downstream of the *aqpZ* gene stop codon. The *ybjE* gene encodes a transmembrane transporter that mediates the export of L-Lysine [31,32]. The *aqpZ* gene encodes a water channel (or aquaporin) that allows bidirectional passive diffusion of water in *E. coli* [33]. TRIP1b is located 41 bp downstream of the *putP* gene stop codon and 355 bp upstream from the *ycdN* gene start codon. The protein encoded by the *putP* gene is a member of the SSS family of sodium/solute transporters [34]. The *ycdN* gene encodes a ferrous iron transmembrane transporter [35,36]. TRIP1c is located 363 bp upstream of the *yohG* gene start codon and 43 bp downstream from the *yohI* gene stop codon. The *yohG* gene encodes an outer membrane efflux protein

that associates with multidrug resistance [37]. The *yohI* gene encodes a protein that binds to FMN and has tRNA dihydrouridine synthetase activity [38]. TRIP1d is located 96 bp downstream of the *alx* gene stop codon and 255 bp upstream from the *sstT* gene start codon. The protein encoded by the *alx* gene is a putative membrane-bound redox modulator [39]. The protein encoded by the *sstT* gene is a member of the DAACS family of sodium ion-coupled serine/threonine complexes with neutral L-amino acid transmembrane transporter activity [40]. TRIP1e is located 124 bp upstream of the *ybaP* gene start codon and 38 bp downstream from the *copA* gene stop codon. The *ybaP* gene encodes a TraB family protein whose function is unknown [41]. The *copA*-encoded protein is a member of the P-type ATPase cation transporter family and participates in copper homeostasis primarily as an ATP-dependent monovalent copper exporter under aerobic conditions [42]. These results showed that most of the genes upstream and downstream of the TRIP1 repeat sequences encode transmembrane transporters associated with the pathogenicity and resistance of *E. coli*.



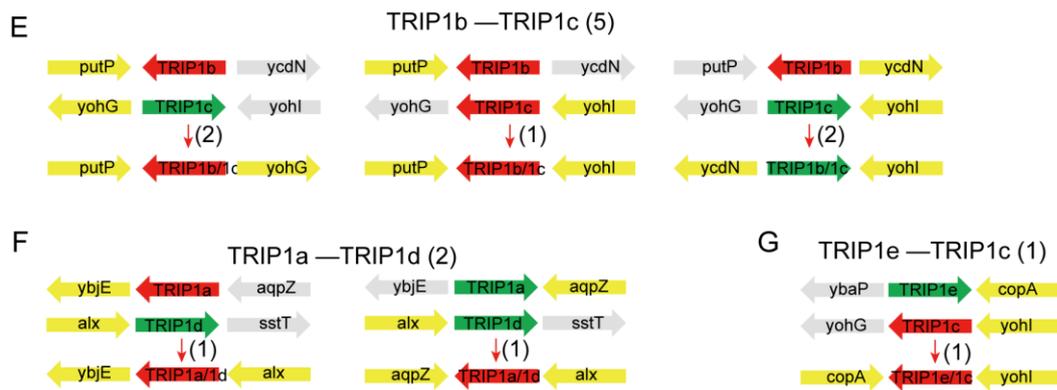
**Figure 4.** Distribution and genome localization of five types of TRIP1 repeats in *E. coli*. The bottom section demonstrates the position of each type of TRIP1 repeats in the *E. coli* K-12 genome.

Moreover, five types of TRIP1 repeats occur at different frequencies in *E. coli* (Figure 4). In the 6725 *E. coli* genomes, TRIP1a is within 4703 genomes, with 3448 in the forward (green) direction and 1255 in the reverse (red) direction. TRIP1b is within 3529 genomes, all of which are in the reverse direction. TRIP1c is within 5741 genomes, with 2952 in the forward direction and 2789 in the reverse direction. TRIP1d is within 1565 genomes, all of which are in the forward direction. TRIP1e is within 232 genomes, all of which are in the forward direction.

### 3.3. Chromosomal Rearrangements Caused by TRIP1 Repeat Sequences

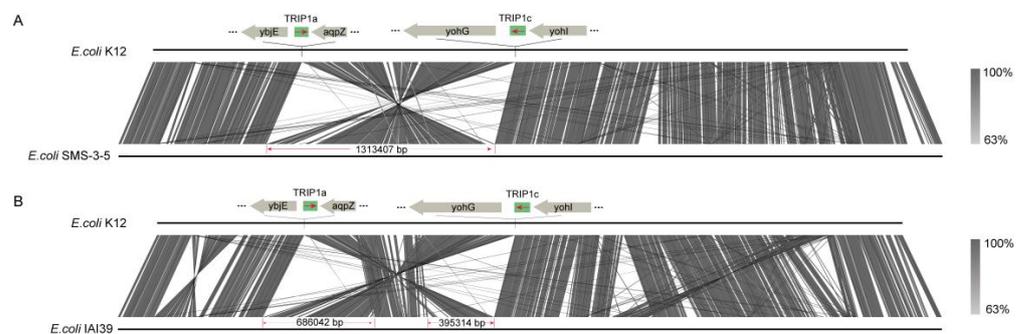
In the *E. coli* genomes, a few TRIP1 repeats were found at the junction of flanking genes from two separate loci. This is a typical result of intra-chromosomal rearrangement caused by repeat sequences [6,43,44]. To determine the frequency and influence of chromosomal arrangement caused by TRIP1 repeats, we scanned all genomes for the relative position of the 10 flanking genes from the five loci. We identified a total of 156 rearrangement events in 129 genomes (2%), including six types (Figure 5A). Eighty-eight percent of the rearrangements (137/156) were found between TRIP1a and TRIP1c (Figure 5B), including





**Figure 5.** Intra-chromosomal rearrangement caused by the TRIP1 repeat sequence. (A) The circle represents the genome of *E. coli* K-12, with the location of each type of TRIP1 repeat indicated. The terms oriC and Ter denote the origin and terminus of chromosome replication, respectively. The line between the two types of TRIP1 repeats represents an intra-chromosomal rearrangement that occurred between them. Six types of rearrangements were identified, with the color and thickness of the line indicating the number of identified rearrangement events. A total of 137 TRIP1a/1c type rearrangement events were found (indicated by the thick red line), while the other five types were much less frequent (indicated by the thin blue line). The number of each rearrangement type is demonstrated. (B–G) The orientation of the flanking genes for each TRIP1 locus is represented by the direction of the arrow. The direction and color of the arrow represent the orientation of the TRIP1 repeats (green for forward, red for inverted, and orange for both directions). The genes linked together by each rearrangement type are highlighted in yellow. The number of each rearrangement type is indicated in brackets. (B) The rearrangements found between TRIP1a and TRIP1c, (C) the rearrangements found between TRIP1a and TRIP1e, (D) the rearrangements found between TRIP1a and TRIP1b, (E) the rearrangements found between TRIP1b and TRIP1c, (F) the rearrangements found between TRIP1a and TRIP1d, and (G) the rearrangements found between TRIP1e and TRIP1c. Please note that the genes and repeats are not drawn to scale.

Moreover, the rearrangement events were analyzed in two complete genome sequences. In *E. coli* SMS-3-5 [45], the whole genome region between TRIP1a and TRIP1c is inverted compared with *E. coli* K-12 (Figure 6A). The region affected by this rearrangement event is 1,313,407 bp. These rearrangements are likely to have a significant impact on the strain’s phenotype. Similarly, in *E. coli* IAI39 [46], parts of the downstream region of TRIP1a and parts of the upstream region of TRIP1c are inverted, while the middle region between them remains unaffected (Figure 6B). These data showed that the TRIP1 repeats could induce large-scale chromosomal rearrangements in the *E. coli* genomes.



**Figure 6.** Intra-chromosomal rearrangement caused by TRIP1 repeats in *E. coli* SMS-3-5 and IAI39. (A) The alignment between *E. coli* K-12 and *E. coli* SMS-3-5, with the entire region between TRIP1a and TRIP1c being inverted. (B) The alignment between *E. coli* K-12 and *E. coli* IAI39, with parts of the intermediate region being inverted. The red arrow marks the orientation of the genes.

#### 4. Discussion

DNA repeats are causes and consequences of genome plasticity [47]. In this study, we systematically analyzed a class of DNA repeat sequences (TRIP1 repeats) that are ubiquitous in the *E. coli* genome but less well-studied. The only recent report on TRIP1 repeats is that of Rudd et al. from 1999 on their discovery and analysis in *E. coli* K12 [26]. Because recombination and selection for DNA repeats vary between genomes, the number and types of DNA repeats are also quite diverse and in line with ecological variables, such as host-dependent associations or population sizes, and with genetic variables, such as the recombination machinery [47,48]. We found that the copy number of TRIP1 repeat sequences varied across the *E. coli* genome, ranging from zero to five copies, with most of the genomes having three copies (36%, 2449/6725), suggesting a high degree of dynamic change in TRIP1 during *E. coli* evolution.

We also found that 271 of the 6725 (not completely sequenced) genomes (4%) did not contain any TRIP1 sequences. However, none of the 353 completely sequenced genomes lacked TRIP1. This suggests that most or all of the 4% of the genomes lacking TRIP1 repeat sequences may be due to incomplete sequencing. A comparison of the frequencies of other copies of TRIP1 repeat sequences in completely and incompletely sequenced genomes supports this hypothesis. The proportion of completely sequenced genomes with one to five copies of TRIP1 repeat sequences was always similar to that of incompletely sequenced genomes: 18%/23%, 18%/21%, 46%/36%, 18%/16%, and 0.28%/0.07%, respectively, in contrast to the situation with zero copies.

Five types of TRIP1 repeats, including two novel types (TRIP1d and TRIP1e), were identified in *E. coli* genomes. Five types of TRIP1 repeats occur at different frequencies in *E. coli*, which were located in 70% (4703/6725), 52% (3529/6725), 85% (5741/6725), 23% (1565/6725), and 3% (232/6725) of the genomes, respectively. Moreover, TRIP1 repeats are present in a non-random manner across bacterial genomes, and each type of TRIP1 repeat is localized to a specific locus on the *E. coli* chromosome.

DNA repeat sequences in bacterial genomes have various functions, including transcribing noncoding RNA [9]. A previous analysis of the RNAseq data available online revealed that out of the 355 DNA repeat elements present in the *E. coli* chromosome, 252 are transcribed [49]. For example, a type of DNA repeat—bacterial Interspersed Mosaic Elements (BIMEs), which are G:C-rich and can form cruciform structures in DNA—can be transcribed into nucleoid-associated non-coding RNA (naRNA) [50]. RNA products of BIMEs are involved in chromosome condensation [9]. Additionally, TRIP1 sequences are also G:C-rich (62–65% G + C) and likely to be capable of forming stable DNA cruciform structures [26]. All five TRIP1 sequences are capable of forming very stable secondary structures in RNA with a conserved double Y structure, suggesting that TRIP1 has the potential to be transcribed into non-coding RNAs with specific physiological functions.

A bacterial transcription unit consists of several components [51]. In addition to the coding region between the start and stop codons, the core components also include the upstream and downstream non-coding regions. The upstream non-coding region, also known as the promoter region, is tens to hundreds of bps in length. It contains regulatory elements such as promoter sequences and transcription factor binding sites that determine the efficiency and specificity of transcription initiation. The downstream non-coding region is located after the coding region and is a few to tens of bps in length. It may contain termination signals and regulatory elements that affect mRNA processing and stability. In this study, we identified five TRIP1 repeat sequences inserted into the noncoding regions of upstream or downstream gene transcription units. For instance, TRIP1a is located 158 bp upstream of the *ybjE* gene start codon and 70 bp downstream of the *aqpZ* gene stop codon. TRIP1b is located 355 bp upstream from the *ycdN* gene start codon and 41 bp downstream of the *putP* gene stop codon. TRIP1c is located 363 bp upstream of the *yohG* gene start codon and 43 bp downstream from the *yohl* gene stop codon. TRIP1d is located 255 bp upstream from the *sstT* gene start codon and 96 bp downstream of the *alx* gene stop codon. TRIP1e is located 124 bp upstream of the *ybaP* gene start codon and 38 bp downstream

from the *copA* gene stop codon. In particular, most of the upstream and downstream genes encode different classes of transmembrane transporters, which are essential for the survival, growth, and adaptation of bacteria in diverse environments. These results suggest that the insertion of these TRIP1 repeat sequences may affect the transcription of these genes, which in turn affects the physiological function of *E. coli*.

DNA repeat sequences could be targeted by recombination processes leading to amplifications, deletions, and rearrangements of genetic material [47,52,53]. In this study, we identified 156 rearrangement events caused by TRIP1 repeat sequences in 6725 *E. coli* genomes (Figure 5). In total, six types of rearrangement events were identified, with the majority (88%, 137/156) occurring between TRIP1a and TRIP1c. Possible reasons for this are the high frequencies of TRIP1a and TRIP1c in the *E. coli* genome, 70% (4703/6725) and 85% (5741/6725), respectively, and their high sequence identity (94%). The region affected by the TRIP1a and TRIP1c rearrangement event can be as long as 1,313,407 bp, as seen in *E. coli* SMS-3-5. Large-scale rearrangements are important in evolution because they can alter chromosome organization and gene expression in ways not possible through point mutations [54]. Thus, we concluded that the large-scale rearrangements between TRIP1 repeats should have a strong influence on the phenotype of these strains, which requires further investigation.

## 5. Conclusions

In summary, we systematically analyzed the diversity, distribution, and chromosomal rearrangements of the TRIP1 repeat sequences in the *E. coli* genome. After screening 6725 *E. coli* genomes, TRIP1 was found to be ubiquitous (96%: 6454/6725) and with many copies in the *E. coli* strains. Five types of TRIP1 repeat sequences were identified in the *E. coli* genomes, and they occurred with varying frequencies in *E. coli*, with TRIP1a and TRIP1c occurring at the highest frequencies. More importantly, the TRIP1 repeat sequence can induce chromosomal rearrangement in the *E. coli* genome. A total of 156 rearrangement events were identified, of which 88% (137/156) were between TRIP1a and TRIP1c. In addition, large-scale rearrangements were also identified between TRIP1a and TRIP1c in *E. coli* genomes, such as strain SMS-3-5. All these findings have important implications for future research on TRIP1 repeats.

**Author Contributions:** Conceptualization and methodology, Z.L.; software and data curation, Z.L., X.L. and N.N.; writing—original draft preparation, Z.L. and T.L.; writing—review and editing, Z.L., T.L. and H.W.; project administration and funding acquisition, T.L. and H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (82072251), the National Key Basic Research Program (973) of China (2015CB554202), and the Beijing Nova Program (Z181100006218111).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data presented are available in the manuscript.

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

## References

1. Smirnov, G.B. Repeats in bacterial genomes: Evolutionary considerations. *Mol. Genet. Mikrobiol. I Virusol.* **2010**, *2*, 11–20. [[CrossRef](#)]
2. Patel, S. Drivers of bacterial genomes plasticity and roles they play in pathogen virulence, persistence and drug resistance. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2016**, *45*, 151–164. [[CrossRef](#)]
3. Čutová, M.; Manta, J.; Porubiaková, O.; Kaura, P.; Šťastný, J.; Jagelská, E.B.; Goswami, P.; Bartas, M.; Brázda, V. Divergent distributions of inverted repeats and G-quadruplex forming sequences in *Saccharomyces cerevisiae*. *Genomics* **2020**, *112*, 1897–1901. [[CrossRef](#)]

4. Waters, E.V.; Tucker, L.A.; Ahmed, J.K.; Wain, J.; Langridge, G.C. Impact of Salmonella genome rearrangement on gene expression. *Evol. Lett.* **2022**, *6*, 426–437. [[CrossRef](#)]
5. Delilhas, N. Impact of small repeat sequences on bacterial genome evolution. *Genome Biol. Evol.* **2011**, *3*, 959–973. [[CrossRef](#)]
6. Brazda, V.; Fojta, M.; Bowater, R.P. Structures and stability of simple DNA repeats from bacteria. *Biochem. J.* **2020**, *477*, 325–339. [[CrossRef](#)] [[PubMed](#)]
7. Bowater, R.P.; Bohálová, N.; Brázda, V. Interaction of Proteins with Inverted Repeats and Cruciform Structures in Nucleic Acids. *Int. J. Mol. Sci.* **2022**, *23*, 6171. [[CrossRef](#)] [[PubMed](#)]
8. Porubiaková, O.; Havlík, J.; Indu, Šedý, M.; Přepechalová, V.; Bartas, M.; Bidula, S.; Šťastný, J.; Fojta, M.; Brázda, V. Variability of Inverted Repeats in All Available Genomes of Bacteria. *Microbiol. Spectr.* **2023**, *11*, e0164823. [[CrossRef](#)] [[PubMed](#)]
9. Qian, Z.; Adhya, S. DNA repeat sequences: Diversity and versatility of functions. *Curr. Genet.* **2017**, *63*, 411–416. [[CrossRef](#)] [[PubMed](#)]
10. Mahfooz, S.; Shankar, G.; Narayan, J.; Singh, P.; Akhter, Y. Simple sequence repeat insertion induced stability and potential ‘gain of function’ in the proteins of extremophilic bacteria. *Extrem. Life Under Extrem. Cond.* **2022**, *26*, 17. [[CrossRef](#)] [[PubMed](#)]
11. Malhotra, N.; Seshasayee, A.S.N. Replication-Dependent Organization Constrains Positioning of Long DNA Repeats in Bacterial Genomes. *Genome Biol. Evol.* **2022**, *14*, evac102. [[CrossRef](#)]
12. Fozo, E.M.; Kawano, M.; Fontaine, F.; Kaya, Y.; Mendieta, K.S.; Jones, K.L.; Ocampo, A.; Rudd, K.E.; Storz, G. Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.* **2008**, *70*, 1076–1093. [[CrossRef](#)] [[PubMed](#)]
13. Liang, W.; Rudd, K.E.; Deutscher, M.P. A role for REP sequences in regulating translation. *Mol. Cell* **2015**, *58*, 431–439. [[CrossRef](#)] [[PubMed](#)]
14. Wen, J.; Fozo, E.M. sRNA antitoxins: More than one way to repress a toxin. *Toxins* **2014**, *6*, 2310–2335. [[CrossRef](#)] [[PubMed](#)]
15. Achaz, G.; Coissac, E.; Netter, P.; Rocha, E.P. Associations between inverted repeats and the structural evolution of bacterial genomes. *Genetics* **2003**, *164*, 1279–1289. [[CrossRef](#)]
16. Eisen, J.A.; Heidelberg, J.F.; White, O.; Salzberg, S.L. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* **2000**, *1*, Research0011. [[CrossRef](#)]
17. Hughes, D. Evaluating genome dynamics: The constraints on rearrangements within bacterial genomes. *Genome Biol.* **2000**, *1*, Reviews0006. [[CrossRef](#)] [[PubMed](#)]
18. Tillier, E.R.; Collins, R.A. Genome rearrangement by replication-directed translocation. *Nat. Genet.* **2000**, *26*, 195–197. [[CrossRef](#)]
19. Suyama, M.; Bork, P. Evolution of prokaryotic gene order: Genome rearrangements in closely related species. *Trends Genet. TIG* **2001**, *17*, 10–13. [[CrossRef](#)]
20. Choi, S.; Ohta, S.; Ohtsubo, E. A novel IS element, IS621, of the IS110/IS492 family transposes to a specific site in repetitive extragenic palindromic sequences in Escherichia coli. *J. Bacteriol.* **2003**, *185*, 4891–4900. [[CrossRef](#)]
21. Partridge, S.R.; Hall, R.M. The IS1111 family members IS4321 and IS5075 have subterminal inverted repeats and target the terminal inverted repeats of Tn21 family transposons. *J. Bacteriol.* **2003**, *185*, 6371–6384. [[CrossRef](#)] [[PubMed](#)]
22. Tobes, R.; Pareja, E. Bacterial repetitive extragenic palindromic sequences are DNA targets for Insertion Sequence elements. *BMC Genom.* **2006**, *7*, 62. [[CrossRef](#)]
23. van Belkum, A. Short sequence repeats in microbial pathogenesis and evolution. *Cell. Mol. Life Sci. CMLS* **1999**, *56*, 729–734. [[CrossRef](#)] [[PubMed](#)]
24. Dame, R.T.; Rashid, F.M.; Grainger, D.C. Chromosome organization in bacteria: Mechanistic insights into genome structure and function. *Nat. Rev. Genet.* **2020**, *21*, 227–242. [[CrossRef](#)] [[PubMed](#)]
25. Lioy, V.S.; Junier, I.; Boccard, F. Multiscale Dynamic Structuring of Bacterial Chromosomes. *Annu. Rev. Microbiol.* **2021**, *75*, 541–561. [[CrossRef](#)]
26. Rudd, K.E. Novel intergenic repeats of Escherichia coli K-12. *Res. Microbiol.* **1999**, *150*, 653–664. [[CrossRef](#)]
27. Edgar, R. Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ* **2018**, *6*, e5030. [[CrossRef](#)]
28. Alzohairy, A.M. BioEdit: An important software for molecular biology. *GERF Bull Biosci* **2011**, *2*, 60–61.
29. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)]
30. Loman, N.J.; Pallen, M.J. Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* **2015**, *13*, 787–794. [[CrossRef](#)]
31. Dubey, S.; Majumder, P.; Penmatsa, A.; Sardesai, A.A. Topological analyses of the L-lysine exporter LysO reveal a critical role for a conserved pair of intramembrane solvent-exposed acidic residues. *J. Biol. Chem.* **2021**, *297*, 101168. [[CrossRef](#)] [[PubMed](#)]
32. Pathania, A.; Sardesai, A.A. Distinct Paths for Basic Amino Acid Export in Escherichia coli: YbjE (LysO) Mediates Export of L-Lysine. *J. Bacteriol.* **2015**, *197*, 2036–2047. [[CrossRef](#)] [[PubMed](#)]
33. Tan, H.; Zhao, Y.; Zhao, W.; Xie, H.; Chen, Y.; Tong, Q.; Yang, J. Dynamics properties of membrane proteins in native cell membranes revealed by solid-state NMR spectroscopy. *Biochim. Biophys. Acta Biomembr.* **2022**, *1864*, 183791. [[CrossRef](#)]
34. Bracher, S.; Schmidt, C.C.; Dittmer, S.I.; Jung, H. Core Transmembrane Domain 6 Plays a Pivotal Role in the Transport Cycle of the Sodium/Proline Symporter PutP. *J. Biol. Chem.* **2016**, *291*, 26208–26215. [[CrossRef](#)] [[PubMed](#)]
35. Grosse, C.; Scherer, J.; Koch, D.; Otto, M.; Taudte, N.; Grass, G. A new ferrous iron-uptake transporter, EfeU (YcdN), from Escherichia coli. *Mol. Microbiol.* **2006**, *62*, 120–131. [[CrossRef](#)]
36. Sturm, A.; Schierhorn, A.; Lindenstrauss, U.; Lilie, H.; Brüser, T. YcdB from Escherichia coli reveals a novel class of Tat-dependently translocated hemoproteins. *J. Biol. Chem.* **2006**, *281*, 13972–13978. [[CrossRef](#)]

37. Sulavik, M.C.; Houseweart, C.; Cramer, C.; Jiwani, N.; Murgolo, N.; Greene, J.; DiDomenico, B.; Shaw, K.J.; Miller, G.H.; Hare, R.; et al. Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. *Antimicrob. Agents Chemother.* **2001**, *45*, 1126–1136. [[CrossRef](#)]
38. Byrne, R.T.; Jenkins, H.T.; Peters, D.T.; Whelan, F.; Stowell, J.; Aziz, N.; Kasatsky, P.; Rodnina, M.V.; Koonin, E.V.; Konevega, A.L.; et al. Major reorientation of tRNA substrates defines specificity of dihydrouridine synthases. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6033–6037. [[CrossRef](#)]
39. Nechooshtan, G.; Elgrably-Weiss, M.; Sheaffer, A.; Westhof, E.; Altuvia, S. A pH-responsive riboregulator. *Genes Dev.* **2009**, *23*, 2650–2662. [[CrossRef](#)]
40. Kim, Y.M.; Ogawa, W.; Tamai, E.; Kuroda, T.; Mizushima, T.; Tsuchiya, T. Purification, reconstitution, and characterization of Na(+)/serine symporter, SstT, of *Escherichia coli*. *J. Biochem.* **2002**, *132*, 71–76. [[CrossRef](#)]
41. Lu, H.; Patil, P.; Van Sluys, M.A.; White, F.F.; Ryan, R.P.; Dow, J.M.; Rabinowicz, P.; Salzberg, S.L.; Leach, J.E.; Sonti, R.; et al. Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PLoS ONE* **2008**, *3*, e3828. [[CrossRef](#)] [[PubMed](#)]
42. Meydan, S.; Klepacki, D.; Karthikeyan, S.; Margus, T.; Thomas, P.; Jones, J.E.; Khan, Y.; Briggs, J.; Dinman, J.D.; Vázquez-Laslop, N.; et al. Programmed Ribosomal Frameshifting Generates a Copper Transporter and a Copper Chaperone from the Same Gene. *Mol. Cell* **2017**, *65*, 207–219. [[CrossRef](#)]
43. Fu, Y.; Zepeda-Gurrola, R.C.; Aguilar-Gutiérrez, G.R.; Lara-Ramírez, E.E.; De Luna-Santillana, E.J.; Rodríguez-Luna, I.C.; Sánchez-Varela, A.; Carreño-López, R.; Moreno-Medina, V.R.; Rodríguez-Pérez, M.A.; et al. The detection of inherent homologous recombination between repeat sequences in *H. pylori* 26695 by the PCR-based method. *Curr. Microbiol.* **2014**, *68*, 211–219. [[CrossRef](#)]
44. Shikov, A.E.; Savina, I.A.; Nizhnikov, A.A.; Antonets, K.S. Recombination in Bacterial Genomes: Evolutionary Trends. *Toxins* **2023**, *15*, 568. [[CrossRef](#)] [[PubMed](#)]
45. Fricke, W.F.; Wright, M.S.; Lindell, A.H.; Harkins, D.M.; Baker-Austin, C.; Ravel, J.; Stepanauskas, R. Insights into the environmental resistance gene pool from the genome sequence of the multidrug-resistant environmental isolate *Escherichia coli* SMS-3-5. *J. Bacteriol.* **2008**, *190*, 6779–6794. [[CrossRef](#)] [[PubMed](#)]
46. Touchon, M.; Hoede, C.; Tenaillon, O.; Barbe, V.; Baeriswyl, S.; Bidet, P.; Bingen, E.; Bonacorsi, S.; Bouchier, C.; Bouvet, O.; et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **2009**, *5*, e1000344. [[CrossRef](#)]
47. Treangen, T.J.; Abraham, A.L.; Touchon, M.; Rocha, E.P. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev.* **2009**, *33*, 539–571. [[CrossRef](#)]
48. MacAlasdair, N.; Pesonen, M.; Brynildsrud, O.; Eldholm, V.; Kristiansen, P.A.; Corander, J.; Caugant, D.A.; Bentley, S.D. The effect of recombination on the evolution of a population of *Neisseria meningitidis*. *Genome Res.* **2021**, *31*, 1258–1268. [[CrossRef](#)]
49. Raghavan, R.; Groisman, E.A.; Ochman, H. Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res.* **2011**, *21*, 1487–1497. [[CrossRef](#)]
50. Qian, Z.; Macvanin, M.; Dimitriadis, E.K.; He, X.; Zhurkin, V.; Adhya, S. A New Noncoding RNA Arranges Bacterial Chromosome Organization. *mBio* **2015**, *6*, e00998-15. [[CrossRef](#)]
51. Cho, B.K.; Zengler, K.; Qiu, Y.; Park, Y.S.; Knight, E.M.; Barrett, C.L.; Gao, Y.; Palsson, B. The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.* **2009**, *27*, 1043–1049. [[CrossRef](#)] [[PubMed](#)]
52. Shaskolskiy, B.; Kravtsov, D.; Kandinov, I.; Dementieva, E.; Gryadunov, D. Genomic Diversity and Chromosomal Rearrangements in *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Int. J. Mol. Sci.* **2022**, *23*, 15644. [[CrossRef](#)] [[PubMed](#)]
53. González-Torres, P.; Rodríguez-Mateos, F.; Antón, J.; Gabaldón, T. Impact of Homologous Recombination on the Evolution of Prokaryotic Core Genomes. *mBio* **2019**, *10*, e02494-18. [[CrossRef](#)] [[PubMed](#)]
54. Raeside, C.; Gaffé, J.; Deatherage, D.E.; Tenaillon, O.; Briska, A.M.; Ptashkin, R.N.; Cruveiller, S.; Médigue, C.; Lenski, R.E.; Barrick, J.E.; et al. Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *mBio* **2014**, *5*, e01377-14. [[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.