



Communication

Comparative Genome Analysis Reveals Accumulation of Single-Nucleotide Repeats in Pathogenic *Escherichia* Lineages

Koji Ishiya and Nobutaka Nakashima *

Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Sapporo 062-8517, Japan; koji.ishiya@aist.go.jp

* Correspondence: n-nakashima@aist.go.jp; Tel.: +81-11-857-9576

Abstract: Homopolymeric tracts (HPTs) can lead to phase variation and DNA replication slippage, driving adaptation to environmental changes and evolution of genes and genomes. However, there is limited information on HPTs in *Escherichia*; therefore, we conducted a comprehensive cross-strain search for HPTs in *Escherichia* genomes. We determined the HPT genomic distribution and identified a pattern of high-frequency HPT localization in pathogenic *Escherichia* lineages. Notably, HPTs localized near transcriptional regulatory genes. Additionally, excessive repeats accumulated in toxin-coding genes. Moreover, the genomic localization of some HPTs might be derived from exogenous DNA, such as that of bacteriophages. Altogether, our findings may prove useful for understanding the role of HPTs in *Escherichia* genomes.

Keywords: comparative genomics; *Escherichia*; homopolymeric tracts; single-nucleotide repeats; transcriptional regulation; pathogenic lineages



Citation: Ishiya, K.; Nakashima, N. Comparative Genome Analysis Reveals Accumulation of Single-Nucleotide Repeats in Pathogenic *Escherichia* Lineages. *Curr. Issues Mol. Biol.* **2022**, *44*, 498–504. <https://doi.org/10.3390/cimb44020034>

Academic Editor: Pedro Escoll

Received: 2 November 2021

Accepted: 17 January 2022

Published: 20 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Homopolymeric tracts (HPTs), also referred to as single-nucleotide repeats, are nucleotide repeats consisting of one of the nucleotides (A, T, G, or C). HPTs are a feature of phase variation and can lead to changes in bacterial gene expression [1–7]. In bacteria, the ability to regulate gene expression according to environmental conditions is important for survival. Repetitive DNA sequences, such as HPTs, cause polymerase slippage during DNA replication, resulting in DNA strand expansion and contraction [8]. These slippage events can be a driving force for the evolution of genes and genomes [9–11]. Therefore, the genomic distribution of repetitive DNA sequences involved in acquisition mechanisms of HPTs is an important topic from an evolutionary standpoint. Although HPTs have been explored in several bacterial species, limited comprehensive analyses of HPTs have been performed within the *Escherichia* genus, which is the most widely used genus for generating genetically engineered bacteria and shows pathogenic variation among strains.

Here, we performed a comprehensive comparative genome analysis for identifying and comparing HPTs across members belonging to the genus *Escherichia*. Our results provide a comprehensive picture of the distribution of HPTs across *Escherichia* genomes and reveal their strain-specific profiles.

2. Materials and Methods

Our study mainly consisted of the following three steps to determine the status of HPTs in *Escherichia* strains: data collection, data sorting, and HPT detection (Figure S10).

2.1. Data Collection

All published *Escherichia* genome sequences were obtained from the National Center for Biotechnology Information (NCBI) GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>; last accessed on 21 August 2020) and RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>; last

accessed on 21 August 2020) using the NCBI Genome Downloading Scripts ver. 0.2.12 (<https://github.com/kblin/ncbi-genome-download>; last accessed on 12 April 2020). To unify the completeness of the genome assembly, downloaded sequences were limited to those at the complete assembly level. As a result, 1346 assembly sequences (FASTA format) were collected along with the coding nucleotide sequences (FASTA format) and annotation tables (GFF). These data were grouped into 140 taxonomies according to the Taxonomy IDs in NCBI (<https://www.ncbi.nlm.nih.gov/taxonomy>; last accessed on 21 August 2020). The complete assembly genomes of the *Propionibacterium* and *Helicobacter* were also collected for interspecies comparison in their HPTs. The collected assembly sequences are summarized in Table S5.

2.2. Detection of HPTs

In this study, an HPT was defined as a region with six or more nucleotide repeats, according to a previous study on *Propionibacterium acnes* [11]. We explored HPTs across the collected assembly sequences using SeqKit, a toolkit for FASTA/Q file manipulation [12]. To investigate the nucleotide bias in the HPTs, the lengths of A/T/G/C repeats in the HPTs were measured by an in-house Python script. The percentages of genomic features with ≥ 6 repeats were calculated with the pandas package ver. 1.0.3 [13] in Python. To statistically evaluate the differences in the number of repeats among the features, multiple comparisons were performed using Tukey's honestly significant difference test [14] (familywise error rate $\alpha = 0.05$) with the statsmodels package ver. 0.12.2 [15] in Python.

To further investigate the accumulation of HPTs on phage-derived sequences of pathogenic strains, we used PHASTER web server (<http://phaster.ca/>; last accessed on 2 November 2021) [16] to infer bacteriophage-derived regions of their genomes. The detection of HPTs was also performed on the inferred regions. These results are summarized in Table S4.

2.3. Analysis of Intragenic HPTs

Intragenic HPTs were identified using the coding regions of DNA or RNA. All detected intragenic HPTs were subjected to annotation validation based on the GFF files. To reduce the effect of incomplete genome assembly, we focused on genes containing intragenic HPTs in all strains. Furthermore, to evaluate variation in intragenic HPTs, genes with significantly different HPTs across strains were identified according to the coefficient of variation of repeat lengths [17]. To explore strain-specific patterns in intragenic HPTs, a cluster analysis based on Euclidean distance was performed using the seaborn package ver. 0.11.1 (<https://seaborn.pydata.org/>; last accessed on 5 October 2020) [18] in Python.

2.4. Analysis of Intergenic HPTs

To elucidate the characteristic patterns of intergenic HPTs appearing between genes, we identified the nearest neighbor sequences of the detected HPTs. Thus, a BED file was prepared containing the genomic coordinates of 150 bases on both sides of intergenic HPTs, and HPT flanking regions (FASTA format) were extracted. With these FASTA files as input queries, a homology search was performed using the BLASTX [19] algorithm of DIAMOND ver. 0.9.24.125 [20] against all protein-coding sequences for *Escherichia*. Query sequences with more than 90% homology and an E -value $< 1 \times 10^{-9}$ were classified as the nearest neighbor genes of the intergenic HPTs. In addition, the frequency of occurrence of genes located in the neighboring HPT region was calculated for all strains. A cluster analysis based on Euclidean distance was conducted for determining strain-specific features in the neighboring genes.

2.5. Phylogenetic Analysis

To clarify taxonomic profiles of strains and HPT localization patterns, we performed a phylogenetic analysis including all 140 strains of the genus *Escherichia*. The samples used for this phylogenetic analysis are summarized in Table S6. Firstly, core genes among all strains

were identified using Roary ver. 3.11.2 [21]; next, a maximum likelihood phylogenetic tree was constructed from these core gene sequences using iqtree ver. 1.6.12 [22]. Finally, the tree was visualized using iTOL (<https://itol.embl.de/>; last accessed on 10 October 2020) [23].

2.6. Correlation Analysis between Pathogenic Factors and HPTs

We used the PathoFact pipeline ver. 1.0 [24] for HPT-accumulating (O-157) and non-accumulating (K12) strains to search their genomes for potentially pathogenic factors, coding bacterial toxin-related ORFs. To investigate the association between pathogenicity and HPT accumulation, correlation analysis was performed on the predicted number of predicted pathogenic ORFs and the observed number of HPTs in their genomes. These results are summarized in Figure S9.

3. Results and Discussion

HPTs have been identified in the genomes of several prokaryote species [2–5,7,25–28]. However, the genomic localization pattern of HPTs in *Escherichia* strains, including common species in biotechnology or clinical pathogens, is not well understood. Therefore, we performed a comprehensive survey of 1346 genomes from all 140 *Escherichia* strains and found a total of 14,335,513 HPTs with six or more nucleotides. The data for each nucleotide are shown in Table 1. Analysis of the nucleotide bias in the HPTs showed that *Escherichia* had a high proportion of A and T repeats relative to C and G repeats, with an (A + T)/(G + C) ratio of 11.60 (Table 1; Figure S1). Although this result may be partially due to poly-T tracts in Rho-independent terminators [11], the bias in this ratio may also result from their genomic nucleotide composition. Indeed, in a supplemental analysis of the genomes of other species, the (A + T)/(G + C) ratios of HPTs in *Propionibacterium* and *Helicobacter* were 0.05 and 10.0, respectively (Table S1), suggesting that the nucleotide bias of HPTs may be influenced by their genomic nucleotide composition.

Table 1. Repeat length of the observed homopolymeric tracts.

	Count ^a	Mean ^b	SD ^c	Min ^d	Max ^e
A	6,606,099	6.27	0.55	6.00	56
C	569,850	6.22	0.57	6.00	70
G	567,499	6.22	0.58	6.00	68
T	6,592,065	6.27	0.56	6.00	108
All	14,335,513	6.27	0.55	6.00	108

^a Summarized by the type of nucleotide (A/T/G/C), regardless of the length of the homopolymeric tract repeats observed. ^b The average homopolymeric tract repeat length. ^c The standard deviation in the length of the repeat. ^d The minimum length of repeat. ^e The maximum length of repeat.

To investigate the role of HPTs in genomes, detected HPTs were categorized based on the following 10 genomic features: protein coding, pseudogene, ribosomal (r)RNA, transfer (t)RNA, RNase P RNA, antisense RNA, non-coding (nc)RNA, transfer-messenger (tm)RNA, rRNA pseudogene, and other intergenic HPTs located outside the coding region. The protein-coding, pseudogene, rRNA, tRNA, RNase P RNA, antisense RNA, ncRNA, tmRNA, and rRNA pseudogene features accounted for 65.78%, 3.41%, 0.44%, 0.09%, 0.03%, 0.02%, 0.02%, and 0.0003% HPTs, respectively (Figure 1, “ALL”). Similarly, detected HPTs were categorized according to the length of the repeats (Figure 1). The proportion of each genomic feature was different depending on the length of the repeats; multiple comparison tests showed a significant difference in the number of repeats in 84.4% (38/45) of the combinations among the features (adjusted *p*-value < 0.05) (Figure S2, Table S2).

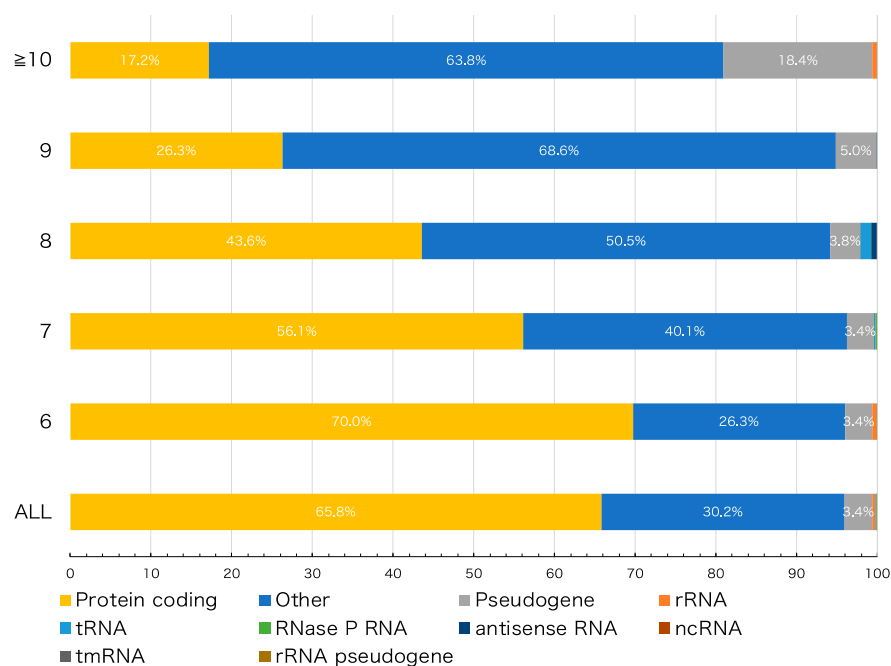


Figure 1. Number of homopolymeric tract (HPT) repeats in various genomic features. The bar plot shows the percentage of the genomic features to which the detected HPTs belong. “ALL” denotes all HPTs with > 6 nucleotide repeats; remaining row labels correspond to the number of nucleotide repeats in HPT. The colors correspond to different genomic features as indicated.

Intergenic HPTs are single-nucleotide polymorphisms present in non-coding regions and their localization patterns across the genome remain largely unexplored. The role of intragenic and intergenic HPTs might be different, and, thus, we first examined HPTs located within intragenic regions. The mean number of occurrences of intragenic HPTs was 1.93 (min = 1; max = 44.63) (Figure S3), which were identified in 119 genes in all *Escherichia* strains (Figure S4). The mean, standard deviation, minimum, and maximum values for the number of HPT repeats observed in these genes were 6.14, 0.28, 6.00, and 8.00, respectively. In contrast, for genes well conserved across strains, intragenic HPTs were mostly six-base repeats (Figure S4), indicating that the HPT repeats in non-coding regions without open reading frames (ORFs) are longer compared with those in protein/RNA-coding regions. Therefore, variation in length is small for intragenic HPTs and extension of repeat length is restricted to maintain protein functions. Among the identified genes, those with significant variation in the number of HPT repeats across the strains coded for transcriptional regulators, dihydrolipoyl dehydrogenase, permease, dTDP-glucose 4,6-dehydratase, colicin V production protein, phosphoenolpyruvate carboxylase, and *N*-acetylglucosamine-6-phosphate deacetylase (Figure 2; $|Z| > 1.96$, two-sided tests).

We next examined HPTs located within intergenic regions. As shown in Figure 1, the percentage of intergenic HPTs increased to more than 60 % when the number of repeats increased to 10 or more (see “Others”). A total of 9272 genes were found to reside in the nearest neighbor sequences of HPTs (Table S3). Of these, genes encoding hypothetical proteins having unknown functions showed the highest percentage, accounting for 8.91% of the total genes flanking intergenic HPTs (Figure S5). Except for genes encoding hypothetical proteins, the top 30 genes with the highest percentage are shown in Figure S6. In particular, 12 strains (taxid: 1343836, 386585, 155864, 1045010, 1330457, 544404, 444450, 1328859, 741093, 83334, 1048689, and 701177) showed a high frequency of HPTs with transcriptional regulatory genes as neighbors. Notably, most strains that showed this pattern belonged to enterohemorrhagic *Escherichia coli* strains, such as O-55 and O-157 (Figure S6). It has been reported that some HPTs with significant repeat length variation among strains are characteristically located in or near transcriptional regulatory genes, thereby affecting their

function [5,7,27,28]. An alternative theory is that the differences in the number of repeats are caused by exogenous DNA from bacteriophages. In fact, pathogenic *E. coli* strains, such as O-157 Sakai, have a higher proportion of exogenous DNA than other non-pathogenic strains [29]. In addition, our analysis showed that in the O-157 Sakai strain, approximately 20% of observed HPTs were located in genomic regions derived from bacteriophages (Table S4). This result suggests that their HPTs may have been contributed, at least in part, by exogenous-derived DNA.

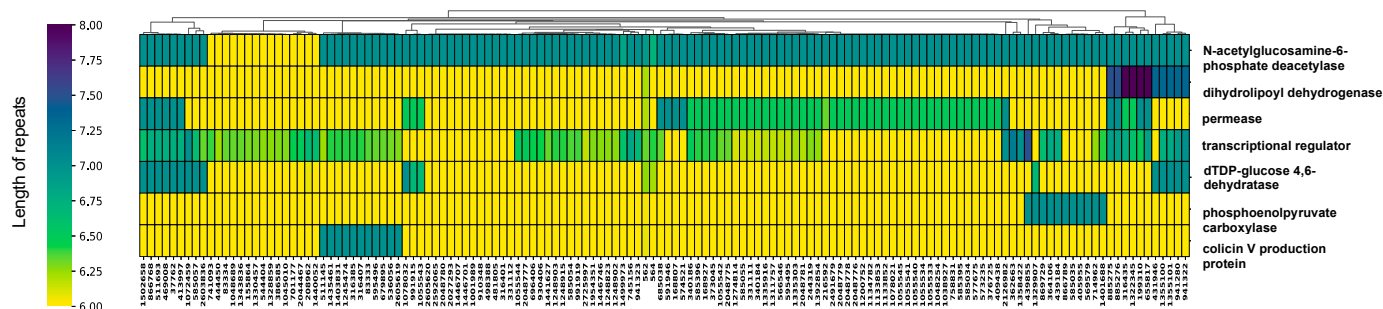


Figure 2. Variation in repeat length in intragenic homopolymeric tracts (HPTs) across *Escherichia* strains. The cluster map shows genes with significantly-variable intragenic HPTs in the genus *Escherichia* clustered by their repeat lengths. The horizontal axis shows the gene (protein) name, and the vertical axis shows the corresponding NCBI Taxonomy ID. The cluster map of all observed intragenic HPTs was shown in Figure S4. The color of the heatmap corresponds to the length of the repeats, with yellow and dark blue colors indicating the shortest and longest repeat lengths, respectively.

These changes in the number of HPT repeats would also cause a frameshift in the gene sequence, and the stop codon could cause loss of transcriptional regulation or protein function. Indeed, we found that the number of HPT repeats and percentage of pseudogenes were positively correlated ($r = 0.707$; Figure S7), and the number of HPT repeats tended to be higher in pseudogenes than in other genomic features (Figure 1). Mutations, including insertions and deletions of HPTs, cause frameshifts that result in multiple open reading frames that cannot be assigned a function in gene prediction.

Our analysis revealed HPTs with long repeat lengths in several genes (Figure S8). In particular, the toxin B gene in some strains has long HPTs (Figure 3). These overrepresented repeats might have been derived from exogenous DNA from bacteriophages, such as the Shiga toxin gene *stx* in the O-157 strain [30,31]. In addition, we also examined the number of potentially pathogenic ORFs and HPTs among strains that differed in their pathogenicity (e.g., O-157 and K12 strains); we observed a significant positive correlation between the number of pathogenic ORFs and HPTs (Figure S9; $R = 0.99$, $p = 1.6 \times 10^{-12}$). This result suggests that the accumulation of HPTs could be associated with their pathogenic factors. Interestingly, HPTs were also observed at high frequency in O antigen genes, which have been used to classify serotypes. This suggests that the localization of HPTs contributes to O antigen variation in *Escherichia*.

In conclusion, we identified several features of HPT localization in *Escherichia* genomes, which could provide useful insights into HPT genomic diversity and plasticity.

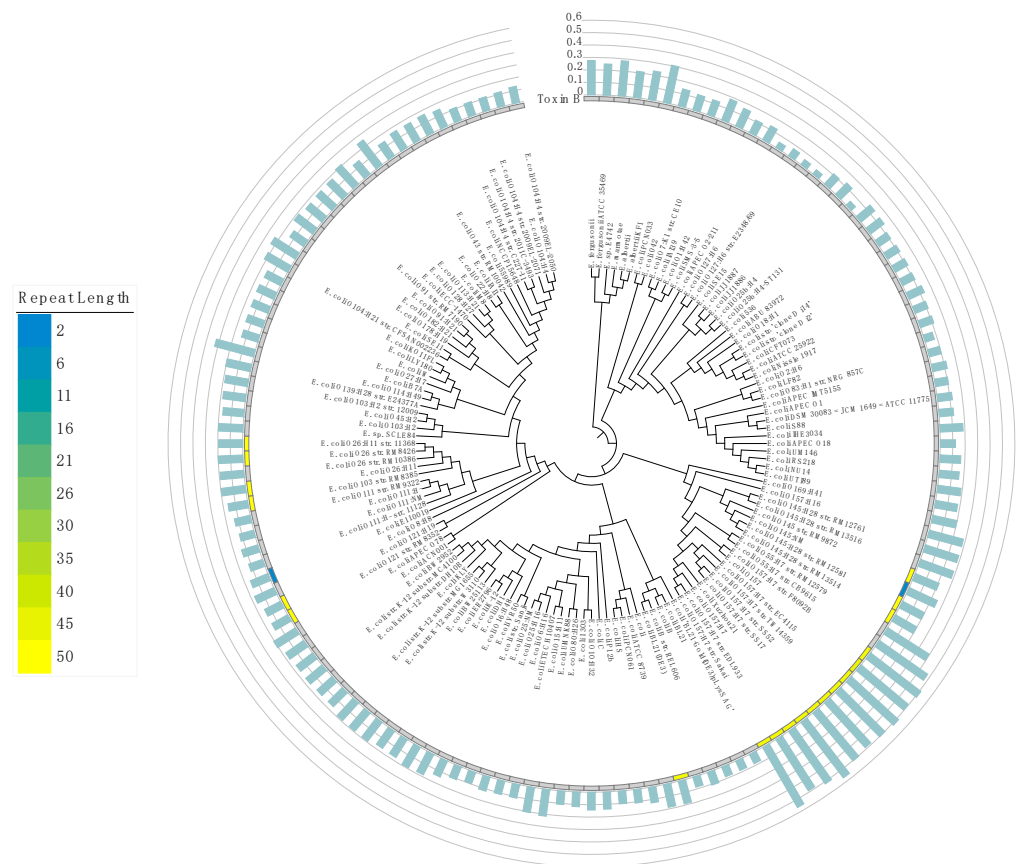


Figure 3. Maximum likelihood phylogenetic tree of the genus *Escherichia*. The phylogenetic tree was constructed based on the core genes in all 140 strains of the genus *Escherichia*; the NCBI Taxonomy ID was replaced by the name of each strain. The distribution of the number of HPT repeats in the toxin B gene is shown in a heatmap. Blue and yellow colors denote the smallest and largest number of repeats, respectively. The gray color indicates the loss of toxin B. The bar plot in light green shows the frequency of occurrence of transcriptional regulatory genes in intergenic HPTs (%). The axis of the bar plot is 0 to 0.6% with 0.1% intervals.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/cimb44020034/s1>.

Author Contributions: K.I. conceived of and designed the study, performed the research, analyzed data, and wrote the paper. N.N. conceived of and designed the study, performed the research, and wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the KAKENHI program of the Japan Society for the Promotion of Science, grant number 19K16246, 20H05822, and 21H04358 to K.I.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: We are grateful to members of our research group for their help and valuable discussions.

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

References

1. Stibitz, S.; Aaronson, W.; Monack, D.; Falkow, S. Phase variation in *Bordetella pertussis* by frameshift mutation in a gene for a novel two-component system. *Nature* **1989**, *338*, 266–269. [[CrossRef](#)]

2. Park, S.F.; Purdy, D.; Leach, S. Localized reversible frameshift mutation in the *flhA* gene confers phase variability to flagellin gene expression in *Campylobacter coli*. *J. Bacteriol.* **2000**, *182*, 207–210. [[CrossRef](#)]
3. Saunders, N.J.; Jeffries, A.C.; Peden, J.F.; Hood, D.W.; Tettelin, H.; Rappuoli, R.; Moxon, E.R. Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. *Mol. Microbiol.* **2000**, *37*, 207–215. [[CrossRef](#)]
4. Kearns, D.B.; Chu, F.; Rudner, R.; Losick, R. Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Mol. Microbiol.* **2004**, *52*, 357–369. [[CrossRef](#)] [[PubMed](#)]
5. Orsi, R.H.; Bowen, B.M.; Wiedmann, M. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. *BMC Genom.* **2010**, *11*, 102. [[CrossRef](#)] [[PubMed](#)]
6. Orsi, R.H.; Ripoll, D.R.; Yeung, M.; Nightingale, K.K.; Wiedmann, M. Recombination and positive selection contribute to evolution of *Listeria monocytogenes* InlA. *Microbiology* **2007**, *153*, 2666–2678. [[CrossRef](#)] [[PubMed](#)]
7. Pernitzsch, S.R.; Tirier, S.M.; Beier, D.; Sharma, C.M. A variable homopolymeric G-repeat defines small RNA-mediated posttranscriptional regulation of a chemotaxis receptor in *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E501–E510. [[CrossRef](#)]
8. Levinson, G.; Gutman, G.A. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **1987**, *4*, 203–221. [[CrossRef](#)]
9. Treier, M.; Pfeifle, C.; Tautz, D. Comparison of the gap segmentation gene hunchback between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. *EMBO J.* **1989**, *8*, 1517–1525. [[CrossRef](#)]
10. Hancock, J.M. The contribution of slippage-like processes to genome evolution. *J. Mol. Evol.* **1995**, *41*, 1038–1047. [[CrossRef](#)]
11. Scholz, C.F.P.; Brüggemann, H.; Lomholt, H.B.; Tettelin, H.; Kilian, M. Genome stability of *Propionibacterium acnes*: A comprehensive study of indels and homopolymeric tracts. *Sci. Rep.* **2016**, *6*, 20662. [[CrossRef](#)]
12. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **2016**, *11*, e0163962. [[CrossRef](#)]
13. Reback, J.; McKinney, W.; Den Van Bossche, J.; Augspurger, T.; Cloud, P.; Klein, A.; Roeschke, M.; Hawkins, S.; Tratner, J.; She, C.; et al. pandas-dev/pandas: Pandas 1.0.3. *Zenodo* **2020**. [[CrossRef](#)]
14. Tukey, J.W. *The Problem of Multiple Comparisons*; Chapman & Hall, Princeton University: London, UK, 1953.
15. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference (SCIPY 2010), Austin, TX, USA, 28 June–3 July 2010. [[CrossRef](#)]
16. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A Better, Faster Version of the PHAST Phage Search Tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [[CrossRef](#)]
17. Ishiya, K.; Aburatani, S. Outlier detection for minor compositional variations in taxonomic abundance data. *Appl. Sci.* **2019**, *9*, 1355. [[CrossRef](#)]
18. Waskom, M.; Gelbart, M.; Botvinnik, O.; Ostblom, J.; Hobson, P.; Lukauskas, S.; Gemperline, D.C.; Augspurger, T.; Halchenko, Y.; Warmenhoven, J.; et al. mwaskom/seaborn: v0.11.1. *Zenodo* **2020**. [[CrossRef](#)]
19. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
20. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [[CrossRef](#)]
21. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)]
22. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)] [[PubMed](#)]
23. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [[CrossRef](#)] [[PubMed](#)]
24. De Nies, L.; Lopes, S.; Busi, S.B.; Galata, V.; Heintz-Buschart, A.; Laczny, C.C.; May, P.; Wilmes, P. PathoFact: A Pipeline for the Prediction of Virulence Factors and Antimicrobial Resistance Genes in Metagenomic Data. *Microbiome* **2021**, *9*, 49. [[CrossRef](#)]
25. Dechering, K.J.; Cuelenaere, K.; Konings, R.N.; Leunissen, J.A. Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. *Nucleic Acids Res.* **1998**, *26*, 4056–4062. [[CrossRef](#)]
26. Josenhans, C.; Eaton, K.A.; Thevenot, T.; Suerbaum, S. Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in *fljP*, a gene encoding a basal body protein. *Infect. Immun.* **2000**, *68*, 4598–4603. [[CrossRef](#)] [[PubMed](#)]
27. Gogol, E.B.; Cummings, C.A.; Burns, R.C.; Relman, D.A. Phase variation and microevolution at homopolymeric tracts in *Bordetella pertussis*. *BMC Genom.* **2007**, *8*, 122. [[CrossRef](#)]
28. Zhou, Y.N.; Lubkowska, L.; Hui, M.; Court, C.; Chen, S.; Court, D.L.; Strathern, J.; Jin, D.J.; Kashlev, M. Isolation and characterization of RNA polymerase *rpoB* mutations that alter transcription slippage during elongation in *Escherichia coli*. *J. Biol. Chem.* **2013**, *288*, 2700–2710. [[CrossRef](#)] [[PubMed](#)]
29. Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* **2001**, *8*, 11–22. [[CrossRef](#)] [[PubMed](#)]
30. Shaikh, N.; Tarr, P.I. *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: Integrations, excisions, truncations, and evolutionary implications. *J. Bacteriol.* **2003**, *185*, 3596–3605. [[CrossRef](#)] [[PubMed](#)]
31. Yara, D.A.; Greig, D.R.; Gally, D.L.; Dallman, T.J.; Jenkins, C. Comparison of Shiga toxin-encoding bacteriophages in highly pathogenic strains of Shiga toxin-producing *Escherichia coli* O157:H7 in the UK. *Microb. Genom.* **2020**, *6*, e000334. [[CrossRef](#)] [[PubMed](#)]