frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 October 2019
doi: 10.3389/fmicb.2019.02254

Prokaryotic Genome Expansion Is Facilitated by Phages and Plasmids but Impaired by CRISPR

Na L. Gao[1,2†], Jingchao Chen[3†], Teng Wang[1], Martin J. Lercher[2] and Wei-Hua Chen[1,3,4*]*

[1] Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular-Imaging, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China, [2] Institute for Computer Science and Department of Biology, Heinrich Heine University, Duesseldorf, Germany, [3] College of Life Science, Henan Normal University, Xinxiang, China, [4] Huazhong University of Science and Technology Ezhou Industrial Technology Research Institute, Ezhou, China

OPEN ACCESS

Edited by:
Yasir Muhammad,
King Abdulaziz University,
Saudi Arabia

Reviewed by:
Rodolphe Barrangou,
North Carolina State University,
United States
Giorgio Giraffa,
Research Centre for Animal Production and Aquaculture (CREA),
Italy

***Correspondence:**
Martin J. Lercher
martin.lercher@hhu.de
Wei-Hua Chen
weihuachen@hust.edu.cn

† These authors have contributed equally to this work as first authors

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 28 April 2019
Accepted: 17 September 2019
Published: 16 October 2019

Citation:
Gao NL, Chen J, Wang T, Lercher MJ and Chen W-H (2019) Prokaryotic Genome Expansion Is Facilitated by Phages and Plasmids but Impaired by CRISPR. Front. Microbiol. 10:2254. doi: 10.3389/fmicb.2019.02254

Viruses and plasmids can introduce novel DNA into bacterial cells, thereby creating an opportunity for genome expansion; conversely, CRISPR, the prokaryotic adaptive immune system, which targets and eliminates foreign DNAs, may impair genome expansions. Recent studies presented conflicting results over the impact of CRISPR on genome expansion. In this study, we constructed a comprehensive dataset of prokaryotic genomes and identified their associations with viruses and plasmids. We found that genomes associated with viruses and/or plasmids were significantly larger than those without, indicating that both viruses and plasmids contribute to genome expansion. Genomes were increasingly larger with increasing numbers of associated viruses or plasmids. Conversely, genomes with CRISPR systems were significantly smaller than those without, indicating that CRISPR has a negative impact on genome size. These results confirmed that on evolutionary timescales, viruses and plasmids facilitate genome expansion, while CRISPR impairs such a process in prokaryotes. Furthermore, our results also revealed that CRISPR systems show a preference for targeting viruses over plasmids.

Keywords: prokaryotic genome expansion, viruses, plasmids, CRISPR, horizontal gene transfer

INTRODUCTION

Gene duplication and/or horizontal gene transfer (HGT) play important roles in functional innovation and species adaptation, and are the main sources of genome expansions (Isambert and Stein, 2009; Schonknecht et al., 2013; Nyvltova et al., 2015; Smith et al., 2016; Tsai et al., 2018). In prokaryotes, it has been shown that the importance of HGT for genome expansions can even outweigh that of gene duplication (Pal et al., 2005; Treangen and Rocha, 2011).

Mobile DNA elements such as viruses and plasmids can introduce novel DNAs into the host genomes (Yamaguchi et al., 2001; Jensen and Lyon, 2009; Lindsay, 2010; Malachowa and Deleo, 2010). They often have a very narrow range of hosts; but under certain conditions, such as antibiotic stress, viruses and plasmids can expand their host ranges (Modi et al., 2013). Therefore, viruses and plasmids are important sources of HGT and of prokaryotic innovations, and consequently drive bacterial evolution and adaptation (Koonin and Wolf, 2008; Nogueira et al., 2009; Argov et al., 2017).

Viruses and plasmids are widely distributed in prokaryotes. Unlike plasmids, viruses are parasites that often lead to lysis of their hosts (Deresinski, 2009; Wernicki et al., 2017). Over the course of prokaryotic evolution, bacteria and archaea developed various defense systems against viruses, plasmids, and other invading genetic elements (Luk et al., 2014). CRISPR (clustered regularly interspaced short palindromic repeats), the adaptive immune system of prokaryotes, is a recently recognized player in the ongoing arms race between prokaryotic viruses and hosts, and plays an important role in the dynamic process by which the genomes of prokaryotes and mobile elements coevolve. CRISPR systems are widespread in prokaryotes, exists in about 40% of bacteria and 90% of archaea (Godde and Bickerton, 2006; Makarova et al., 2011; Seed et al., 2013; Huang et al., 2016), or ∼10% of bacteria as revealed by a recent study (Burstein et al., 2016). CRISPR systems can also target plasmids (Marraffini and Sontheimer, 2008), although plasmids are not necessarily detrimental to their host's fitness but instead often carry a diverse range of antimicrobial and biocide resistance genes that may help their hosts to survive under certain conditions (Mccarthy and Lindsay, 2012; Shabbir et al., 2016).

Based on the above observations, it is reasonable to speculate that over the course of evolution, viruses and plasmids may contribute to the expansion of prokaryotic genomes, while CRISPR systems may impair such a process. These speculations are consistent with recent observations that CRISPR limits HGT by targeting foreign DNAs (Marraffini and Sontheimer, 2008; Bikard et al., 2012). However, controversial observations have also been reported recently. For example, Gophna et al. (2015) did not observe the expected negative correlation between CRISPR activity in microbes with three independent measures of recent HGT, leading them to conclude that the inhibitory effect of CRISPR against HGT is undetectable. Furthermore, a recent study revealed that CRISPR-mediated phage resistance can even enhance HGT by increasing the resistance of transductants against subsequent phage infections (Watson et al., 2018). These observations appear surprising, as the restricted acquisition of foreign genetic material is believed to be one of the sources of the maintenance fitness cost of CRISPR systems and may be one of the reasons for the patchy distribution of CRISPR among bacteria (Frost et al., 2005; Baltrus, 2013). Thus, it is currently unclear what long-term effects CRISPR, viruses, and plasmids have on genome expansion.

In this study, we first collected a comprehensive dataset of prokaryotes and their associations with viruses, plasmids, and CRISPR systems. We then evaluated the contributions of viruses, plasmids, and CRISPR to genome size. After controlling for genome GC (guanine+cytosine) content, which is known to correlate significantly with genome size (Chen et al., 2016a,b), small genome size typically exhibits low GC content, and this bias in base composition has been explained as consequences of genome recoding and selection on efficient resource usage. However, one example is thermophiles, preferentially grow in high heat conditions, which have much more G/C pairs in the coding regions to enhance the stability of mRNA secondary structure (Basak et al., 2010), and decreased genome size to limit their cost of living (Sabath et al., 2013). The evolutionary forces constraining genome size and GC-content have been attributed to a variety of factors, such as environmental energetic constraints. We found that both viruses and plasmids are associated with larger genomes, while the presence of a CRISPR system is associated with small genome size. Genome sizes increase with increasing numbers of associated viruses and plasmids. Our results clearly indicate that in the long run, viruses and plasmids facilitate genome expansions, while CRISPR impairs such a process in prokaryotes. Furthermore, our results also reveal a striking preference of CRISPR systems for targeting viruses rather than plasmids, consistent with the typical consequences of phage and plasmid infections to the hosts and the roles of CRISPR as a defense system.

MATERIALS AND METHODS

Data

We obtained data from three sources. Microbe-phage interaction data was collected from the MVP database, which we described in a previous publication (Gao et al., 2018). MVP is one of the latest and largest databases about microbe-phage interactions, which supplied 26,572 interactions between 9,245 prokaryotes and 18,608 viral clusters based on 30,321 evidence entries (Gao et al., 2018).

The basic genome information from complete archaeal and bacterial genomes, including the number of associated plasmids, was downloaded from the NCBI Genome database[1] (N.R. Coordinators, 2018). In order to remove redundancy and avoid incomplete annotation, we only used the complete closed genomes in this study, which represented only a small part of all genome drafts (mostly incomplete) available from NCBI. We obtained in total 14,575 complete prokaryotic genomes (340 archaeal and 14,286 bacterial genomes) and belonging to 7,151 species. We selected a represented genome for each of species with the highest GC-contents among the strains. Among which, 2,287 prokaryotes were identified associating with plasmids.

The CRISPRs data was obtained from the CRISPRCasDb database[2] (Grissa et al., 2007; Couvin et al., 2018) including 340 archaeal and 16,650 bacterial strains. 2,927 complete prokaryotic genomes (231 archaeal and 2,696 bacterial genomes) were associated with CRISPR systems, while 66 encode CRISPR exclusively on plasmids. The 66 genomes which only contained plasmid-encoded CRISPR systems were removed from all analyses.

In total, 7,085 prokaryotes were found in both of the first two datasets; among these, 2,221 contained plasmids, 2,682 contained viruses, and 2,861 contained CRISPRs on their chromosomes. Detailed information on the dataset can be found in **Supplementary Table 2**.

Statistical Analysis

All data were analyzed using R v3.4 (R Core Team, 2017). All pairwise comparisons between two groups of numeric data (genome

[1]https://www.ncbi.nlm.nih.gov/genome/; accessed on June 16, 2019.

[2]https://crisprcas.i2bc.paris-saclay.fr; last update June 18, 2019.

sizes or genomic GC-contents) were performed by Wilcoxon rank-sum tests. Linear model (LM) analysis was performed with the R function glm. Relative importance analysis was performed with the calc.relimp function available from the R package "relatimpo" (Groemping, 2006).

RESULTS

Prokaryotic Genomes and Their Associations With Viruses, Plasmids and CRISPRs

To systematically investigate the impacts of viruses, plasmids, and CRISPRs on genome expansion, we constructed a list of 7,085 completely sequenced prokaryotic genomes and obtained their associations with viruses, plasmids, and CRISPRs; for details please consult the section "Materials and Methods" and **Supplementary Table 2**.

As shown in **Figure 1A**, we found that 62.15% of prokaryotes had no known associations with infecting viruses. 12.24, 13.62, and 12% of prokaryotes were associated with one, two to three, and more than three viruses, respectively. In addition, we found that 68.02% of prokaryotes did not associate with plasmids, while 15.13, 11.12, and 5.73% of the genomes associated with one, two to three, and more than three plasmids, respectively (**Figure 1B**). Previous studies suggested that the genomic GC-contents as well as nucleotide frequencies of phages and plasmids often closely resembles that of their hosts (Nakashima et al., 2015; Ahlgren et al., 2017; Ren et al., 2017); consistent with these previous observations, we obtained correlation coefficient values of 0.969 and 0.968 between the GC-contents of the host

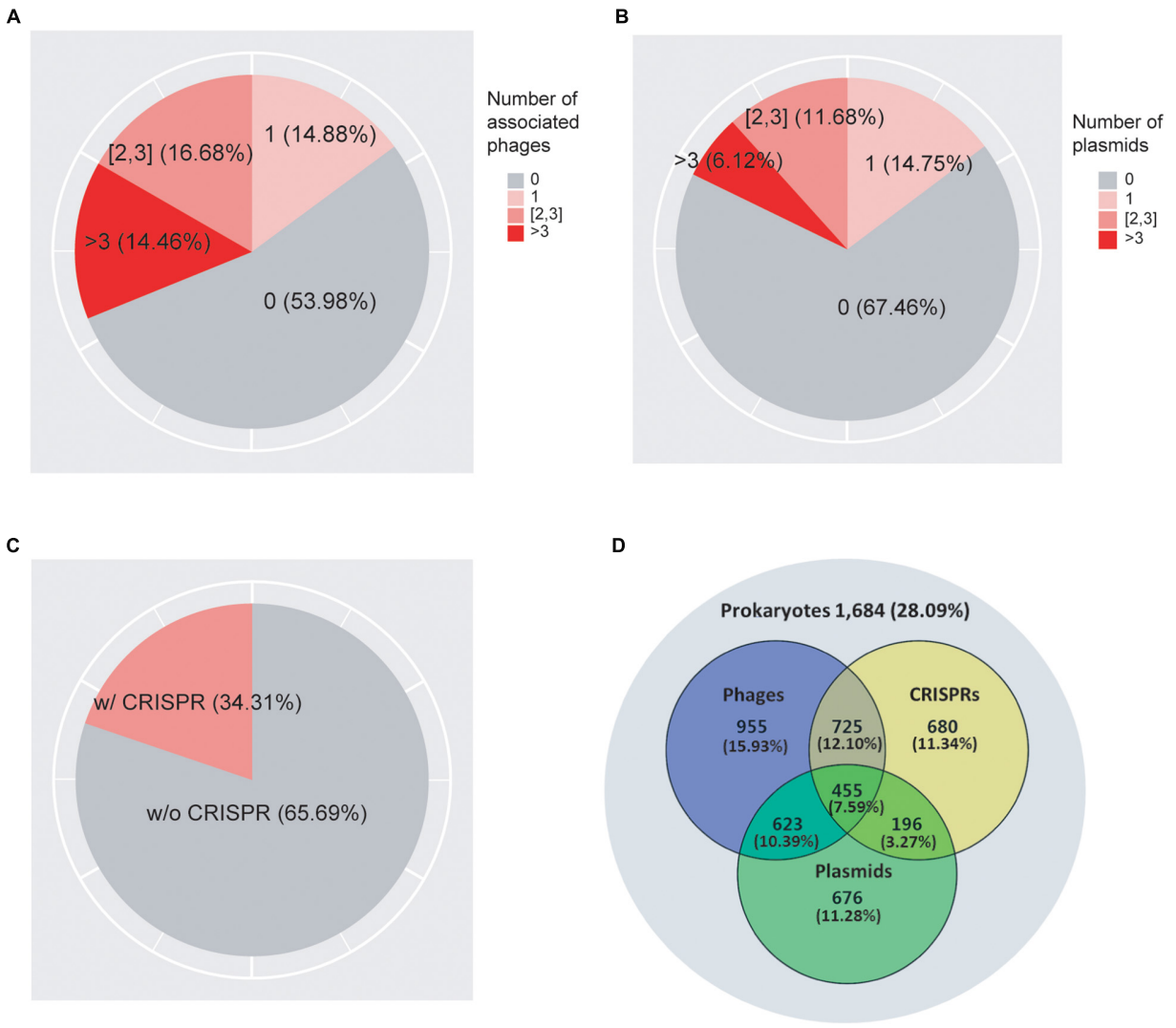

FIGURE 1 | A total of 7,085 prokaryotic genomes and their associations with viruses **(A)**, plasmids **(B)**, and CRISPRs **(C)**. The Venn diagram **(D)** shows the overlap of their distributions in prokaryotes. 1,962 genomes (27.69%) were not found to be associated with viruses, plasmids, or CRISPRs; 439 (6.2%) genomes were associated with all three elements.

genomes and their associated viruses and plasmids, respectively (**Supplementary Figures 1A,B**), confirming the high quality of our association data. We found that in total 40.44% of genomes collected in this study contained either viruses or plasmids but not both, while 14.39% of genomes contained both viruses and plasmids (**Figure 1D**).

As shown in **Figure 1C**, we found CRISPR systems in 40.38% of the prokaryotic genomes; this percentage is within the range of previously reported numbers (Godde and Bickerton, 2006; Makarova et al., 2011; Seed et al., 2013; Burstein et al., 2016; Huang et al., 2016). We found that CRISPRs were significantly enriched in virus-associated compared to non-virus-associated genomes (odds ratio OR = 1.18, $P = 1.07 \times 10^{-3}$ from Fisher's exact test) but not in plasmid-associated compared to non-plasmid-associated genomes (OR = 1.04, $P = 0.43$). In addition, we found that CRISPRs were enriched in virus-associated compared to plasmid-associated genomes, although the significance was only marginal (OR = 1.15, $P = 0.08$, excluding genomes containing both viruses and plasmids), suggesting a strong target preferences of CRISPRs toward viruses (**Table 1**).

Viruses and Plasmids Are Associated With Larger Genomes, While CRISPR Is Associated With Smaller Ones

We next investigated which factors contribute significantly to genome size. Previous results have shown a strong correlation between genomic GC content and genome size (Chen et al., 2016a); GC content may even play a causal role in shaping genome size (Chen et al., 2016b). Applying a (LM, see section "Materials and Methods" for details), we found that GC content was indeed the strongest predictor of genome size (**Table 2**). The LM analysis also revealed that the presence/absence of viruses, plasmids, and CRISPR all significantly influenced genome size; the presences of viruses and of plasmids were associated with increased genome sizes, while CRISPR was associated with decreased genome sizes (**Table 2**). We estimated that the relative importance of these factors for genome size were 89% for GC-content, 6.11% for virus presence, 3.22% for plasmid presence, and 0.04% for CRISPR presence. This revealed that GC-content was indeed the most significant predictor of genome size; the presence of plasmids and viruses also had a significant influence on genome size; as compared with other factors, the presence/absence of CRISPR had relative small influence on genome size. Interestingly, we found that the presence of both viruses and plasmids in the same genome was associated with a smaller genome size than expected (i.e., the interaction term viruses*plasmids was negative, **Table 2**). We hypothesized that there are fitness costs inherent to expanding or limiting the genome size, when a given prokaryote is in a highly diverse and competitive environments. In addition to the CRISPR systems, there are other known and novel anti-phage defense systems in the microbial pan-genome (Doron et al., 2018). Unless stated otherwise, we thus limit our further analyses to prokaryotes that contained either viruses or plasmids but not both. Note that our conclusions on the influence of viruses, plasmids, and CRISPR systems on genome size remain unchanged if we perform separate analyses on genomes containing no viruses and on genomes containing no plasmids (**Table 2**).

TABLE 1 | Estimated enrichment of CRISPR in virus-associated and plasmid-associated genomes compared to other genomes, and enrichment of CRISPR in virus-associated compared to plasmid-associated genomes.

Comparison	Odds ratio[b]	*P*-value[c]
Virus-associated vs. others	1.18	1.07×10^{-3}
Plasmid-associated vs. others	1.04	0.43
Virus- vs. plasmid-associated	1.08	0.21
Virus-associated vs. others[a]	1.17	7.48×10^{-3}
Plasmid-associated vs. others[a]	0.97	0.67
Virus- vs. Plasmid-associated[a]	1.15	0.08

[a]excluding genomes contained both viruses and plasmids. [b]odds ratio OR > 1 indicates enrichment of CRISPR in the first group, while OR < 1 indicates depletion. [c]P-values indicate whether CRISPR is significantly enriched or depleted in the first group as compared with the second according to Fisher' exact test.

TABLE 2 | Relative importance of various factors for genome size in a linear model (LM).

Dataset	Factor	Coefficient	*P*-value	Relative importance
All	GC%	0.086	$<2 \times 10^{-16}$	91.84%
	Plasmid	0.714	$<2 \times 10^{-16}$	5.91%
	Virus	0.454	$<2 \times 10^{-16}$	2.22%
	CRISPR	−0.043	0.248	0.03%
	Virus*plasmid	−0.130	0.104	–
No plasmids	GC%	0.087	$<2 \times 10^{-16}$	96.62%
	Virus	0.454	$<2 \times 10^{-16}$	3.18%
	CRISPR	−0.108	0.017	0.20%
No viruses	GC%	0.087	$<2 \times 10^{-16}$	93.16%
	Plasmid	0.713	$<2 \times 10^{-16}$	6.77%
	CRISPR	0.066	0.168	0.07%

*The equation of "All" dataset used in the LM is size ∼ GC% + plasmid + virus + CRISPR + virus*plasmid. Here, size represents the genome size; GC% represents the genomic GC-content of the host genome; plasmid, virus, and CRISPR represent whether the host genomes are associated with plasmids, viruses, and CRISPR, respectively. The "Coefficient" column contains estimated regression coefficients calculated by ordinary least squares. Relative importance was calculated using the "relaimpo" package (Groemping, 2006); the equation of "No plasmids" dataset is size ∼ GC% + virus + CRISPR; and the equation of "No viruses" dataset is size ∼ GC% + plasmid + CRISPR.*

Increasing Numbers of Viruses and Plasmids Are Associated With Increased Genome Sizes

We next investigated the impact of the numbers of viruses and plasmids on genome size. Viruses and plasmids often have very narrow host ranges (Suzuki et al., 2014; Gao et al., 2018); the number of known associations with viruses may indicate the ability of the prokaryotic host to acquire external novel DNA. Consistent with our expectation, we found that genomes associated with more viruses had larger overall genomes (**Figure 2A**; **Supplementary Figure 2A**). We observed similar results with plasmids (**Figure 2B**; **Supplementary Figure 2B**).

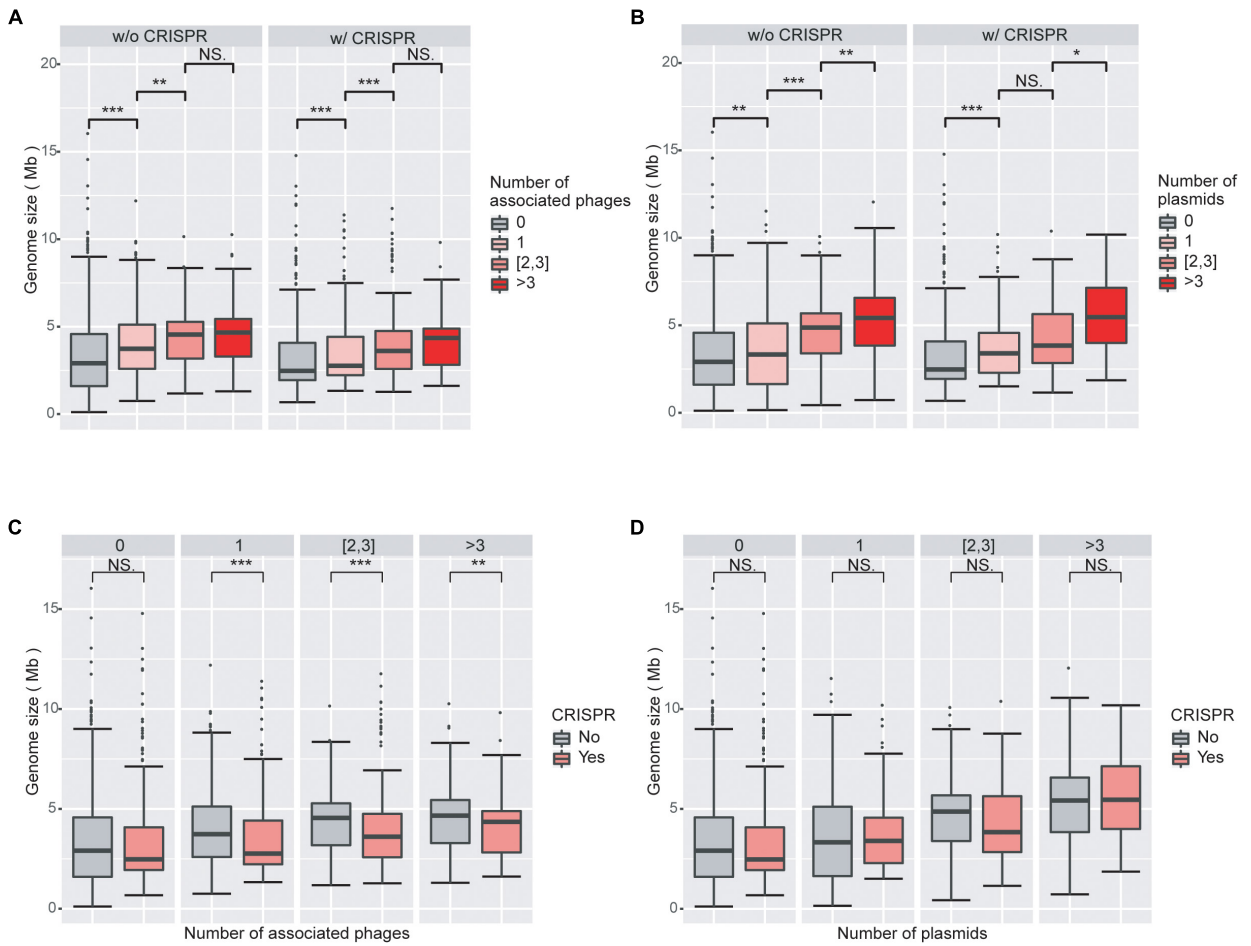

FIGURE 2 | Increasing numbers of viruses and plasmids are associated with increased genome sizes, while virus-associated genomes with CRISPR systems are smaller than those without CRISPR systems. **(A)** Boxplot of genomes size as a function of the number of associated viruses. Genome sizes are larger with increasing numbers of associated viruses, when genomes without CRISPR systems. **(B)** Boxplot of genomes size as a function of the number of associated plasmids. The impact of plasmids on genome size is similar to that of viruses. **(C)** Boxplot of genome size as a function of the presence/absence of CRISPRs in genomes associated with viruses. Virus-associated genomes with CRISPR systems are significantly smaller in size than those without CRISPR, regardless of the number of viruses they are associated with. **(D)** Boxplots of genome sizes in genomes associated with plasmids as a function of the presence/absence of CRISPRs. CRISPRs have no significant impact on genome sizes in genomes associated with plasmids. Wilcoxon rank sum tests were used to compare between groups. Level of significance: $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; NS. $P \geq 0.05$.

Consistent with the results from the LM analysis, we found that virus-associated genomes are statistically significantly smaller when they encode a CRISPR system compared to when they do not (**Figure 2C**). However, we did not find a corresponding trend in plasmid-associated genomes (**Figure 2D**). These results are consistent with the different fitness consequences of virus and plasmid invasions to the prokaryotic hosts. Both viruses and plasmids can bring exogenous DNA to prokaryotes and decrease the fitness of their hosts, for example by increasing the burden on the host's transcription and translation apparatus. However, viruses typically cause substantial additional fitness decreases through virion production and assembly and eventually host lysis, while plasmids often carry genes that are beneficial to the survival of their hosts under certain circumstances (Dionisio et al., 2005; Jiang et al., 2013). It is thus likely that the CRISPR systems in prokaryotes are more sensitive to viruses than to plasmids. This line of argument is also consistent with our results that the presence of CRISPRs is more enriched in virus-associated than in plasmid-associated genomes.

The Influence of Associated Viruses, Plasmids, and CRISPR on Genome GC-Content

We then investigated which factors contribute significantly to genome GC-content. Consistent with our previous results (LM analysis, **Table 2**), we found that genome size was indeed the most significant predictor of GC-content, with a relative importance of almost 99% (LM analysis, **Table 3**). The presence of plasmids also had a significant influence on GC-content, with a relative importance of 1% (**Table 3**). The presence/absence of viruses and CRISPR had no significant influence on GC-content

by themselves; surprisingly, however, the presence of phages reduced the influence of plasmid presence on GC content.

We also investigated whether these factors contribute significantly to GC-content when genomes contain no viruses/plasmids. As expected, genome size remained the most significant factor for the prediction of genome GC-content, as shown in **Table 3**, with a relative importance of around 99%.

As shown in **Supplementary Table 3**, we find that the number of associated viruses and plasmids contribute significantly to GC-content, but we don't find clear and consistent trends in GC-content as a function of the number of associated viruses or plasmids (**Supplementary Figures 3A–F**).

DISCUSSION

We expected that viruses and plasmids could facilitate genome expansions because they can bring novel DNAs (genes or fragments) into prokaryotic cells that can be integrated into the host genome, while CRISPR immune systems could impair such a process by targeting and eliminating foreign DNAs. However, recent studies presented inconsistent results regarding this topic (Marraffini and Sontheimer, 2008; Makarova et al., 2011; Bikard et al., 2012; Gophna et al., 2015; Watson et al., 2018).

To address this issue, we constructed a comprehensive dataset of prokaryotic genomes and their associations with viruses and plasmids. By dividing genomes into distinct groups according to whether they associated with viruses and/or plasmids and/or contained CRISPRs, we revealed that genomes with viruses or with plasmids were significantly larger than those without, and genome sizes increased with increasing numbers of associated viruses/plasmids. Conversely, virus-associated (but not plasmid-associate) genomes with CRISPRs were significantly smaller in size than those without, regardless of the number of associated viruses. These results confirm that in the long run, viruses and plasmids facilitate genome expansions while CRISPR impairs virus-driven genome expansions. In some cases, prokaryotes could utilize foreign DNAs to expand their metabolic capacities and/or enhance their physiological properties (e.g., antibiotic resistance), leading to genome expansion. Conversely, foreign DNAs that did not have immediate benefits would be unlikely to be incorporated, the genomes tend to stay "small(er)." The "Refusal" process is achieved by defense systems including CRISPR. In addition to the CRISPR systems, there are other known and novel anti-phage defense systems, such as Abi, R-M, toxin/anti-toxin and so on (Doron et al., 2018). There are fitness costs inherent to expanding and limiting the genome size (requires more time and energy), which could have major competitiveness impacts when a given prokaryote is in a highly diverse and competitive environments.

It is worth noting that the CRISPR systems themselves could lead to "genome expansion" through incorporating new spacer sequences into CRISPR arrays. On average a genome can contain ∼40 CRISPR spacers, with total length of ∼1.1 k for all the CRISPR array regions. Despite these modest additions to genome size, we still found that CRISPR-containing genomes were smaller, suggesting that the CRISPR arrays had limited impact on the total genome size.

Genome size evolution has previously been reported to be associated with that of genomic GC-content (Gao et al., 2017). Thus, it appeared possible that virus- and/or plasmid-association has a direct effect not only on genome size but also on GC-content. However, in this study, we found only minor influences of viruses and plasmids on genomic GC-content (**Table 3** and **Supplementary Table 1**). We also split our data into archaea and bacteria, and found similar results in bacteria subgroup not in archaea. This is likely due to the less samples of archaea (**Supplementary Tables 4–7**).

Our results also imply that CRISPR immune systems might be more sensitive toward invading viruses than plasmids, consistent with the differential fitness burdens brought by the two types of foreign invaders to the hosts (Canchaya et al., 2004; Weinberger et al., 2012; Jiang et al., 2013; Pleska and Guet, 2017).

Our results differ significantly from several previous studies (Gophna et al., 2015; Watson et al., 2018). For example, Gophna et al. (2015) reported that the inhibitory effect of CRISPR against HGT is undetectable using three independent measures of recent HGT. However, it is known that CRISPR spacers – which were used by Gophna et al. (2015) to assess CRISPR activity – have very high turnover rates, on the time-scale of days (Deveau et al., 2008; Horvath et al., 2008; Tyson and Banfield, 2008), while HGT genes may take a very long time to be incorporated into existing gene networks (Lercher and Pal, 2008), suggesting that it is only possible to look at the impacts of CRISPRs on HGTs at evolutionary scales. Interestingly, Gophna et al. (2015) also studied spacer acquisition and concluded there was a bias toward frequently encountered invasive exogenous genetic elements, especially infecting viruses; this is consistent with our conclusion that CRISPRs tend to be more sensitive toward invading viruses than plasmids. Recently, Watson et al. (2018) reported that the CRISPR system of the bacterium *Pectobacterium atrosepticum* enabled the host to resist phage infection, but that this enhanced rather than impeded HGT by transduction. However, it is yet

TABLE 3 | Relative importance of various factors for GC-content (GC%) in a linear model (LM).

Dataset	Factors	Coefficient	*P*-value	Relative importance
All	Size	4.081	$<2 \times 10^{-16}$	99.12%
	Plasmid	−1.423	5.5×10^{-5}	0.85%
	Virus	−0.089	0.788	0.02%
	CRISPR	0.115	0.656	0.01%
	Virus*plasmid	−0.434	0.438	–
No plasmids	Size	4.132	$<2 \times 10^{-16}$	99.85%
	Virus	−0.139	0.678	0.01%
	CRISPR	0.618	0.048	0.14%
No viruses	Size	4.107	$<2 \times 10^{-16}$	99.30%
	Plasmid	−1.442	7.7×10^{-5}	0.60%
	CRISPR	0.528	0.109	0.10%

*The equation of "All" dataset used in the LM is GC% ∼ size + plasmid + virus + CRISPR + virus*plasmid; the equation of "No plasmids" dataset is GC% ∼ size + virus + CRISPR; and the equation of "No viruses" dataset is GC% ∼ size + plasmid + CRISPR.*

to be seen whether or not this phenomenon is unique to *P. atrosepticum*. Though our findings are known to hold true globally, there will certainly be some exceptions with fewer reports at present.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

W-HC initialized, conceptualized, and designed the study. NG, JC, and TW analyzed the data, wrote and edited the manuscript. ML contributed to key discussions and methods on findings, and prepared the tables and figures. W-HC and ML edited the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This manuscript has been released as a Pre-Print on BioRxiv (Na et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02254/full#supplementary-material

REFERENCES

Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A., and Sun, F. (2017). Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res.* 45, 39–53. doi: 10.1093/nar/gkw1002

Argov, T., Azulay, G., Pasechnek, A., Stadnyuk, O., Ran-Sapir, S., Borovok, I., et al. (2017). Temperate bacteriophages as regulators of host behavior. *Curr. Opin. Microbiol.* 38, 81–87. doi: 10.1016/j.mib.2017.05.002

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.* 28, 489–495. doi: 10.1016/j.tree.2013.04.002

Basak, S., Mukhopadhyay, P., Gupta, S. K., and Ghosh, T. C. (2010). Genomic adaptation of prokaryotic organisms at high temperature. *Bioinformation* 4, 352–356. doi: 10.6026/97320630004352

Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12, 177–186. doi: 10.1016/j.chom.2012.06.003

Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 7:10613. doi: 10.1038/ncomms10613

Canchaya, C., Fournous, G., and Brussow, H. (2004). The impact of prophages on bacterial chromosomes. *Mol. Microbiol.* 53, 9–18. doi: 10.1111/j.1365-2958.2004.04113.x

Chen, W.-H., Van Noort, V., Lluch-Senar, M., Hennrich, M. L., Wodke, J. A. H., Yus, E., et al. (2016a). Integration of multi-omics data of a genome-reduced bacterium: prevalence of post-transcriptional regulation and its correlation with protein abundances. *Nucleic Acids Res.* 44, 1192–1202. doi: 10.1093/nar/gkw004

Chen, W.-H., Lu, G., Bork, P., Hu, S., and Lercher, M. J. (2016b). Energy efficiency trade-offs drive nucleotide usage in transcribed regions. *Nat. Commun.* 7:11334. doi: 10.1038/ncomms11334

Coordinators, N. R. (2018). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 46, D8–D13. doi: 10.1093/nar/gkx1095

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Neron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

Deresinski, S. (2009). Bacteriophage therapy: exploiting smaller fleas. *Clin. Infect. Dis.* 48, 1096–1101. doi: 10.1086/597405

Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., et al. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1390–1400. doi: 10.1128/jb.01412-07

Dionisio, F., Conceicao, I. C., Marques, A. C., Fernandes, L., and Gordo, I. (2005). The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biol. Lett.* 1, 250–252. doi: 10.1098/rsbl.2004.0275

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., et al. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359:eaar4120. doi: 10.1126/science.aar4120

Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732. doi: 10.1038/nrmicro1235

Gao, N., Lu, G., Lercher, M. J., and Chen, W. H. (2017). Selection for energy efficiency drives strand-biased gene distribution in prokaryotes. *Sci. Rep.* 7:10572. doi: 10.1038/s41598-017-11159-3

Gao, N. L., Zhang, C., Zhang, Z., Hu, S., Lercher, M. J., Zhao, X.-M., et al. (2018). MVP: a microbe–phage interaction database. *Nucleic Acids Res.* 46, D700–D707. doi: 10.1093/nar/gkx1124

Godde, J. S., and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* 62, 718–729. doi: 10.1007/s00239-005-0223-z

Gophna, U., Kristensen, D. M., Wolf, Y. I., Popa, O., Drevet, C., and Koonin, E. V. (2015). No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. *ISME J.* 9, 2021–2027. doi: 10.1038/ismej.2015.20

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. doi: 10.1186/1471-2105-8-172

Groemping, U. (2006). Relative importance for linear regression in R: the package relaimpo. *J. Stat. Softw.* 17, 11–27.

Horvath, P., Romero, D. A., Coute-Monvoisin, A. C., Richards, M., Deveau, H., Moineau, S., et al. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1401–1412. doi: 10.1128/jb.01415-07

Huang, Q., Luo, H., Liu, M., Zeng, J., Abdalla, A. E., Duan, X., et al. (2016). The effect of Mycobacterium tuberculosis CRISPR-associated Cas2 (Rv2816c) on stress response genes expression, morphology and macrophage survival of *Mycobacterium smegmatis*. *Infect. Genet. Evol.* 40, 295–301. doi: 10.1016/j.meegid.2015.10.019

Isambert, H., and Stein, R. R. (2009). On the need for widespread horizontal gene transfers under genome size constraint. *Biol. Direct* 4:28. doi: 10.1186/1745-6150-4-28

Jensen, S. O., and Lyon, B. R. (2009). Genetics of antimicrobial resistance in *Staphylococcus aureus*. *Future Microbiol.* 4, 565–582. doi: 10.2217/fmb.09.30

Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B. R., and Marraffini, L. A. (2013). Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet.* 9:e1003844. doi: 10.1371/journal.pgen.1003844

Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719. doi: 10.1093/nar/gkn668

Lercher, M. J., and Pal, C. (2008). Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.* 25, 559–567. doi: 10.1093/molbev/msm283

Lindsay, J. A. (2010). Genomic variation and evolution of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 300, 98–103. doi: 10.1016/j.ijmm.2009.08.013

Luk, A. W., Williams, T. J., Erdmann, S., Papke, R. T., and Cavicchioli, R. (2014). Viruses of haloarchaea. *Life* 4, 681–715. doi: 10.3390/life4040681

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577

Malachowa, N., and Deleo, F. R. (2010). Mobile genetic elements of *Staphylococcus aureus*. *Cell. Mol. Life Sci.* 67, 3057–3071. doi: 10.1007/s00018-010-0389-4

Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843–1845. doi: 10.1126/science.1165771

Mccarthy, A. J., and Lindsay, J. A. (2012). The distribution of plasmids that carry virulence and resistance genes in *Staphylococcus aureus* is lineage associated. *BMC Microbiol.* 12:104. doi: 10.1186/1471-2180-12-104

Modi, S. R., Lee, H. H., Spina, C. S., and Collins, J. J. (2013). Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* 499, 219–222. doi: 10.1038/nature12212

Nakashima, H., Homma, K., and Mawatari, K. (2015). Relationship of genomic G+C content between phages/plasmids and their hosts. *Br. Biotechnol. J.* 9:1–9. doi: 10.9734/bbj/2015/20046

Nogueira, T., Rankin, D. J., Touchon, M., Taddei, F., Brown, S. P., and Rocha, E. P. (2009). Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr. Biol.* 19, 1683–1691. doi: 10.1016/j.cub.2009.08.056

Nyvltova, E., Stairs, C. W., Hrdy, I., Ridl, J., Mach, J., Paces, J., et al. (2015). Lateral gene transfer and gene duplication played a key role in the evolution of *Mastigamoeba* balamuthi hydrogenosomes. *Mol. Biol. Evol.* 32, 1039–1055. doi: 10.1093/molbev/msu408

Pal, C., Papp, B., and Lercher, M. J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* 37, 1372–1375. doi: 10.1038/ng1686

Pleska, M., and Guet, C. C. (2017). Effects of mutations in phage restriction sites during escape from restriction-modification. *Biol. Lett.* 13:20170646. doi: 10.1098/rsbl.2017.0646

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., and Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5:69. doi: 10.1186/s40168-017-0283-5

Sabath, N., Ferrada, E., Barve, A., and Wagner, A. (2013). Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol. Evol.* 5, 966–977. doi: 10.1093/gbe/evt050

Schonknecht, G., Chen, W. H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., et al. (2013). Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* 339, 1207–1210. doi: 10.1126/science.1231707

Seed, K. D., Lazinski, D. W., Calderwood, S. B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* 494, 489–491. doi: 10.1038/nature11927

Shabbir, M. A., Hao, H., Shabbir, M. Z., Wu, Q., Sattar, A., and Yuan, Z. (2016). Bacteria vs. Bacteriophages: parallel evolution of immune arsenals. *Front. Microbiol.* 7:1292. doi: 10.3389/fmicb.2016.01292

Smith, G., Macias-Munoz, A., and Briscoe, A. D. (2016). Gene duplication and gene expression changes play a role in the evolution of candidate pollen feeding genes in *Heliconius* butterflies. *Genome Biol. Evol.* 8, 2581–2596. doi: 10.1093/gbe/evw180

Suzuki, H., Brown, C. J., and Top, E. M. (2014). "Genomic signature analysis to predict plasmid host range," in *Molecular Life Sciences*, eds R. D. Wells, J. S. Bond, J. Klinman, and B. S. S. Masters (New York, NY: Springer).

Treangen, T. J., and Rocha, E. P. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7:e1001284. doi: 10.1371/journal.pgen.1001284

Tsai, Y. M., Chang, A., and Kuo, C. H. (2018). Horizontal gene acquisitions contributed to genome expansion in insect-symbiotic *Spiroplasma* clarkii. *Genome Biol. Evol.* 10, 1526–1532. doi: 10.1093/gbe/evy113

Tyson, G. W., and Banfield, J. F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* 10, 200–207.

Watson, B. N. J., Staals, R. H. J., and Fineran, P. C. (2018). CRISPR-Cas-Mediated phage resistance enhances horizontal gene transfer by transduction. *mBio* 9, e2406–e2417. doi: 10.1128/mBio.02406-17

Weinberger, A. D., Sun, C. L., Plucinski, M. M., Denef, V. J., Thomas, B. C., Horvath, P., et al. (2012). Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* 8:e1002475. doi: 10.1371/journal.pcbi.1002475

Wernicki, A., Nowaczek, A., and Urban-Chmiel, R. (2017). Bacteriophage therapy to combat bacterial infections in poultry. *Virol. J.* 14:179.

Yamaguchi, T., Hayashi, T., Takami, H., Ohnishi, M., Murata, T., Nakayama, K., et al. (2001). Complete nucleotide sequence of a *Staphylococcus aureus* exfoliative toxin B plasmid and identification of a novel ADP-ribosyltransferase, EDIN-C. *Infect. Immun.* 69, 7760–7771. doi: 10.1128/iai.69.12.7760-7771.2001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.