British Biotechnology Journal
9(1): 1-9, 2015, Article no.BBJ.20046
ISSN: 2231–2927

SCIENCEDOMAIN *international*
www.sciencedomain.org

Relationship of Genomic G+C Content between Phages/Plasmids and Their Hosts

Hiroshi Nakashima[1*], Keiko Homma[1] and Kazuhiro Mawatari[1]

[1]Department of Clinical Laboratory Science, Graduate Course of Medical Science and Technology, School of Health Sciences, Kanazawa University, 5-11-80 Kodatsuno, Kanazawa 920-0942, Japan.

Authors' contributions

This study was carried out in collaboration between all authors. Authors HN and KM designed the study. Author HN performed the data analysis and wrote the first draft of the manuscript. Authors KH and KM managed the analyses of the study. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BBJ/2015/20046
Editor(s):
(1) Qiang Ge, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA.
Reviewers:
(1) Anonymous, Bangladesh Council of Scientific and Industrial Research, Bangladesh.
(2) Ilham Zahir, University Sidi Mohamed Ben Abdellah, Morocco.
Complete Peer review History: http://sciencedomain.org/review-history/10615

Original Research Article

Received 8th July 2015
Accepted 5th August 2015
Published 19th August 2015

ABSTRACT

Aims: We compared genomic G+C content between bacteriophages/plasmids and their host genomes of 46 species to reveal their relation. To examine the large deviation in the G+C content between bacteriophages and their host genomes, the ancestral bacteriophage which infected early was estimated using homologous genes of bacteriophages based on G+C content at the third codon positions.
Place and Duration of Study: Department of Clinical Laboratory Science, Graduate Course of Medical Science and Technology, School of Health Sciences, Kanazawa University, Japan.
Methodology: The bacteria employed in this study consist of 6 species from Archaea and 40 species from Eubacteria. Orthologs were identified by the two-directional best hit approach of homology search. A phylogenetic tree was obtained by multiple sequence alignment of homologous genes. The ancestral bacteriophage which infected early was estimated based on G+C content at the third codon positions. We assumed that the two bacteriophages have evolved from a common ancestor, and their identical codons were thought to represent their ancestor type.
Results: The relationship of G+C content between bacteriophages/plasmids and host genomes was almost linear. Three bacteriophages were largely deviated from the linear relation. A phylogenetic

**Corresponding author: Email: hinaka@staff.kanazawa-u.ac.jp;*

tree obtained using the orthologs of *Mycobacterium smegmatis* indicated which bacteriophage branched early. Assuming that the G+C content of identical codons represents their common ancestor, the ancestor was estimated that it had similar G+C content with its host.

Keywords: Genomic G+C content; bacteriophage and plasmid; G+C content deviation; G+C content at the third codon positions.

1. INTRODUCTION

It is reported that there are a large number of bacteriophages [1] and they influence microbial world. Phages and plasmids can mediate DNA into bacteria, so they are widely used as vectors to transfer foreign DNA. The mediation of DNA by phages or plasmids is called horizontal gene transfer [2-5]. Plasmids carry genes for inactivation of antibiotics, metabolism of natural products, and production of toxins. A bacterial cell may have several copies of a plasmid or it may have no plasmids.

Comparison of the character of genes between bacteriophages and hosts may provide the clue of phage infection. The simple character of genes is G+C content. The G+C content of bacterial genomes varies among species from 25% to 75%, but is relatively constant within a bacterial genome [6,7]. The nucleotide sequences of genes of bacterial genomes have species-specific dinucleotide compositions [8-10]. Comparative genomic analysis between bacteriophage and host genomes revealed similarity in G+C content in phages of *Staphylococcus aureus* [11] and mycobacteriophages [12]. However, it is reported that mycobacteriophages [13] and phages in *Pseudomonas aeruginosa* [14] have largely deviated G+C content compared with their hosts. To answer the issue of this discrepancy of mycobacteriophages, we examined the G+C content at the third codon positions within the orthologous genes [15]. We inferred the G+C content of the common ancestor and estimated the phages that had undergone major G+C content change.

2. MATERIALS AND METHODS

2.1 Genomic G+C Content and Size of Phages and Plasmids

The relation between bacteriophages and their hosts was obtained from the genomes to protein structure and function (GTOP) database [16] web site (http://spock.genes.nig.ac.jp/~genome/gtop.html) and UCSC archaeal genome browser [17] (http://archaea.ucsc.edu/). The G+C contents of bacteriophages, plasmids, and host genomes were retrieved from the National Center for Biotechnology Information (NCBI) database web site (http://www.ncbi.nlm.nih.gov/). The bacteria employed in this study consist of 6 species from Archaea and 40 species from Eubacteria. The 46 species bacteria are listed in Table 1 with phage name, plasmid name and their G+C content. The plasmid expressed by hyphens indicates that the plasmid is not available for the species. The minimum G+C content of genomic sequence was 26.6% in *Mycoplasma pulmonis* and the maximum G+C content was 72.4% in *Streptomyces venezuelae*. The phage genome size was in the range from 5.6 kbp of phage MAV1 in *Mycoplasma arthritidis* to 244.8 kbp of phage KVP40 in *Vibrio parahaemolyticus.* All the phages employed were double stranded DNA (dsDNA) viruses. The plasmid genome size was from 4.7 kbp of plasmid pTLC in *Vibrio cholerae* to 1354.2 kbp of plasmid pSymA in *Sinorhizobium meliloti*.

2.2 Orthologous Proteins

The genomic sequence data of seven bacteriophages of *Mycobacterium smegmatis* and eight bacteriophages of *Pseudomonas aeruginosa* were retrieved from the NCBI database. The amino acid sequences obtained from the genomic data were compared. For example, all the protein sequences from the phage Patience of *Mycobacterium smegmatis* were compared against the all the protein sequences from the phage Fruitloop of *Mycobacterium smegmatis* using the BLASTP program [18]. Orthologs were identified by the two-directional best hit approach using BLASTP. The CLUSTALW program [19] in the GenomeNet [20] was used for multiple sequence alignment of orthologs to obtain a phylogenetic tree.

2.3 G+C Content at the Third Codon Positions

Nucleotide sequence alignments corresponding to the aligned protein sequences were employed

in the analysis. The codons in the nucleotide alignment were classified into three groups according to the definition of Bellgard and Gojobori [21]. Group 1 codons, referred to as IA, are different but code for the identical amino acid. Group 2 codons, DA, are different and code for a different amino acid. Group 3 codons, IC, are identical codons. IA codons having the identical nucleotide at the first position were considered in this study.

3. RESULTS AND DISCUSSION

3.1 Relation of G+C Content between Phages/Plasmids and Host Genomes

The G+C content of phages/plasmids against their host genomes is plotted in Fig. 1. The relationship of G+C content between phages/plasmids and host genomes was almost linear. Three phages, RM378 in *Rhodothermus marinus,* Aeh1 in *Aeromonas hydrophia,* and 44RR2.8t in *Aeromonas salmonicida* were largely deviated from the linear relation (Fig. 1). The G+C content between phages/plasmids and host genomes was within 95% confidence interval except three phages mentioned above. The genomic G+C content was lower in the three phages than in their hosts. The direction of deviation in G+C content is consistent with the report of Rocha and Dandrin [21]. The three phages mentioned above have relatively large genome size; longer than 100 kbp. The relationship between relatively large genome size and the deviation in G+C content to their hosts is not clearly understood. Concerning the plasmids, the G+C content was similar to that of their hosts (Fig. 1). This is consistent with the report that G+C content in plasmids and hosts is highly correlated [22]. The genome of *Vibrio cholerae* [23] consists of two chromosomes, large chromosome approximate 3M bp and small one approximate 1 M bp with G+C content of 47.7% and 46.9%, respectively. It is considered that the smaller chromosome was originally a megaplasmid. This result indicated that even the megaplasmid has the similar G+C content to that of the host.

In a dsDNA, the amount of adenine is equal to the amount of thymine and the amount of guanine is equal to the amount of cytosine. This is known as Chargaff's first parity rule [24,25]. This rule also applies to single stranded DNA and is called Chargaff's second parity rule [26,27]. Mitchell and Bridge [28] tested Chargaff's second parity rule over 3400 genomic sequences and the validity of this rule has been confirmed for genome sequences from archaea, eubacteria, eukaryotes and viruses. Therefore, the mononucleotide composition is represented simply as G+C content. The species were selected taking into consideration the coverage of a wide range of genomic G+C content.

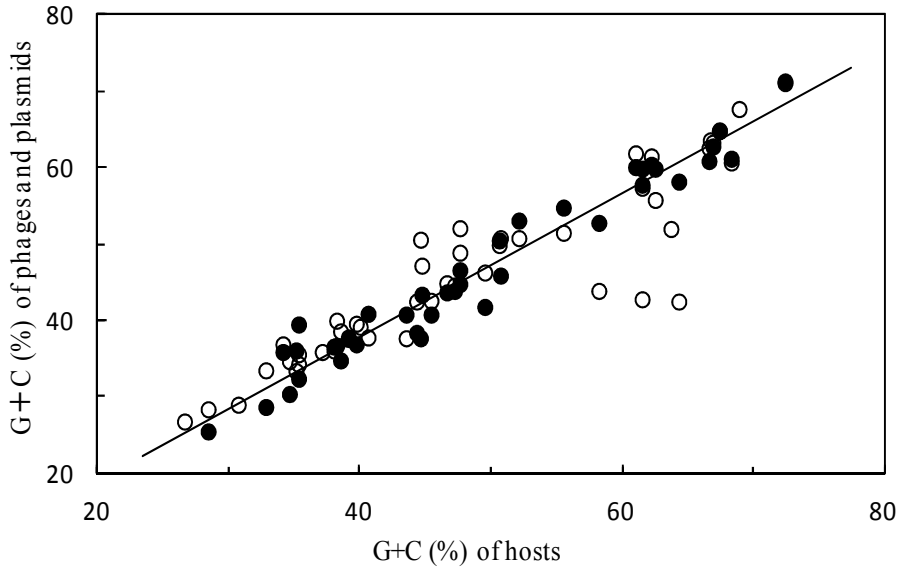

Fig. 1. Plot of the genomic G+C content (%) of phages (open circles) and plasmids (filled circles) against hosts

Table 1. List of species used in this study

No	Domain	Phylum	Section	Species (host)	GC (%)	Phage	GC (%)	Plasmid	GC (%)
1	Archaea	Crenarchaeota	Thermoprotei	*Acidianus hospitalis*	34.1	*Acidianus filamentous virus 1*	36.9	pAH1	35.9
2			Thermoprotei	*Sulfolobus islandicus*	35.1	S. islandicus filamentous virus	33.4	pYN01	36.1
3	Archaea	Euryarchaeota	Halobacteria	*Natrialba magadii*	61.0	Virus PhiCh1	61.9	pNMAG01	60.1
4			Halobacteria	*Haloarcula hispanica*	62.5	pleomorphic virus 1	55.8	pHH400	59.9
5			Methanobacteria	*Methanobacterium thermoautotrophicum*	49.5	Bacteriophage psiM2	46.3	pFV1	41.8
6			Thermococci	*Pyrococcus abyssi*	44.7	*Pyrococcus abyssi virus 1*	47.2	pGT5	43.4
7	Bacteria	Actinobacteria	Actinobacteridae	*Mycobacterium smegmatis*	67.4	Bacteriophage Bxz1	64.8	pMYCSM01	64.9
8			Actinobacteridae	*Streptomyces venezuelae*	72.4	Bacteriophage VWB	71.1	pSVH1	71.3
9		Bacteroidetes	Bacteroidia	*Rhodothermus marinus*	64.3	*Bacteriophage RM378	42.5	pRMAR01	58.2
10		Cyanobacteria	Chroococcales	*Synechococcus sp. PCC 7502*	40.6	Phage S-PM2	37.8	pSYN7502.01	40.9
11		Firmicutes	Bacilli	*Bacillus subtilis*	43.5	Bacteriophage B103	37.7	pBEST195S	40.8
12			Bacilli	*Lactobacillus casei*	46.6	Bacteriophage A2	44.9	pBD-II	43.7
13			Bacilli	*Lactobacillus gasseri*	35.3	Bacteriophage phi adh	35.6	pLgLA39	39.5
14			Bacilli	*Lactobacillus johnsonii*	34.6	Prophage Lj928	34.7	p9785L	30.4
15			Bacilli	*Lactococcus lactis*	35.3	Bacteriophage bIL170	34.3	pKF147A	32.4
16			Bacilli	*Listeria ivanovii*	37.1	Bacreriophage A511	35.9	-----	
17			Bacilli	*Listeria monocytogenes*	38.0	Bacteriophage A118	36.1	pLM5578	36.6
18			Bacilli	*Staphylococcus aureus*	32.8	Prophage phiPV83	33.5	pN315	28.7
19			Bacilli	*Streptococcus mitis*	40.0	Phage SM1	39.2	----	
20			Bacilli	*Streptococcus pneumoniae*	39.7	Bacteriophage EJ-1	39.6	pSpnP1	36.9
21			Bacilli	*Streptococcus pyrogenes*	38.5	Temperate phage phiNIH1.1	38.6	pSM19035	34.8
22			Bacilli	*Streptococcus thermophilus*	39.1	Phage Sfi21	37.6	pSMQ308	37.8
23			Clostridia	*Clostridium perfringeus*	28.4	Bacteriophage phi3626	28.4	pCP13	25.5
24		Proteobacteria	Alphaproteobacteria	*Sinorhizobium meliloti*	62.2	Phage PBC5	61.5	pSymA	60.4
25			Betaproteobacteria	*Burkholderia cenocepacia*	66.9	Phage Bcep781	63.3	pBCJ2315	62.8
26			Betaproteobacteria	*Burkholderia cepacia*	66.7	Phage Bcep1	63.6	----	
27			Betaproteobacteria	*Burkholderia pseudomallei*	68.3	Bacteriophage phi1026b	60.7	pPHB194	61.2
28			Deltaproteobacteria	*Myxococcus xanthus*	68.9	Bacteriophage Mx8	67.7	-----	
29			Gammaproteobacteria	*Actinobacillus actinomycetemcomitans*	44.3	Bacteriophage Aaphi23	42.5	S57	38.4

No	Domain	Phylum	Section	Species (host)	GC (%)	Phage	GC (%)	Plasmid	GC (%)
30			Gammaproteobacteria	*Aeromonas hydrophia*	61.5	*Bacteriophage Aeh1	42.8	pAHH04	59.9
31			Gammaproteobacteria	*Aeromonas salmonicida*	58.2	*Bacteriophage 44RR2.8t	43.9	4	52.8
32			Gammaproteobacteria	*Eschirichia coli*	50.6	Phage T3	49.9	p1ESCUM	50.5
33			Gammaproteobacteria	*Klebsiella oxytoca*	55.5	Bacteriophage phiKO2	51.5	pKOX_NDM1	54.8
34			Gammaproteobacteria	*Haemophilus influenzae*	38.2	Phage HP1	40.0	pF3031	36.7
35			Gammaproteobacteria	*Pseudomonas aeruginosa*	66.6	Phage phi CTX	62.6	pKLC102	60.9
36			Gammaproteobacteria	*Pseudomonas putida*	61.5	Phage gh-1	57.4	pND6-2	57.8
37			Gammaproteobacteria	*Salmonella enterica*	52.1	Phage epsilon15	50.8	pSLT	53.1
38			Gammaproteobacteria	*Shigella flexneri*	50.7	Bacgteriophage V	50.8	pSFxv_1	45.9
39			Gammaproteobacteria	*Vibrio cholerae*	47.6	Bacteriophage K139	48.9	pTLC	46.6
40			Gammaproteobacteria	*Vibrio harveyi*	44.6	Bacteriophage VHML	50.6	pVCR1	37.7
41			Gammaproteobacteria	*Vibrio parahaemolyticus*	45.4	Bacteriophage KVP40	42.6	pVPUCMV	40.8
42			Gammaproteobacteria	*Xanthomonas oryzae*	63.7	Bacteriophage Xp10	52.0	-----	
43			Gammaproteobacteria	*Yersinia enterocolitica*	47.2	Bacteriophage PY54	44.6	pYVe8081	43.9
44			Gammaproteobacteria	*Yersinia pestis*	47.6	Bacteriophage L-413C	52.1	pCD1	44.8
45		Tenericutes	Mollicutes	*Mycoplasma arthritidis*	30.7	Phage MAV1	29.0	-----	
46			Mollicutes	*Mycoplasma pulmonis*	26.6	Phage P1	26.8	-----	

*The three phages indicated by * have largely deviated G+C content compared to their hosts*

3.2 Phylogenetic Tree of Orthologs

Usually, the genomic G+C content of phages and plasmids in a same host species is rather constant as seen in *Staphylococcus aureus* [11]. The phages both in Mycobacterium genus [13] and in *Pseudomonas aeruginosa* [14] have wide range of G+C content. The hosts also have wide range of G+C content in *Mycobacterium* (58% in *M. laprae* - 69% in *M. avium*) and in *Pseudomonas* (57% in *P.* sp S9 - 67% in *P. aeruginosa*). It is reported that the phages have the ability to infect different species within the same genus *(e.g*. between *M. smegmatis* and *M. tuberculosis*) [13].

The orthologs of bacteriophages of *Mycobacterium smegmatis* were examined. Gene products such as gp33 of Boomer, gp30 of Fruitloop, gp238 of Bxz1, gp11 of Perseus, gp11 of Jasper, gp26 of Lebrou, and gp42 of Patience were obtained as orthologs. A phylogenetic tree of the orthologs indicated that Patience branched first then Lebrou branched (Fig. 2).

The genomic G+C content of Patience 50.3% was the lowest and that of Lebrou 58.8% was second lowest. Similarly, eight bacteriophages of *Pseudomonas aeruginosa* were examined to obtain orthologs. However, no orthologs were obtained due to the no sequence match each other as reported by Kwan et al. [14]. The reason of no sequence match among proteins from phages of *P. aeruginosa* is unclear.

3.3 G+C Content at the Third Codon Positions

The G+C content at the third codon positions against the total G+C content of seven orthologous genes of bacteriophages of *Mycobacterium smegmatis* is plotted in Fig. 3. It showed a linear relationship as reported by Muto and Osawa [6]. The ortholog, gp42 of Patience showed the maximum 33% sequence match with the gp30 of Fruitloop. The nucleotide sequence alignment corresponding to the protein sequence alignment of Patience and Fruitloop was prepared. There were 67 IA codons, 217 DA codons and 45 IC codons, and their G+C contents at the third codon positions were plotted (Fig. 3). The G+C content at the third positions of IC codons of Patience and Fruitloop was the same 82.2% according to the definition. The two phages have evolved from a common ancestor, and the IC codons are thought to represent their ancestor type. The difference in G+C content at the third position between the IA and IC codons was -32.8% for Patience and 1.2% for Fruitloop. Assuming that the G+C content of IC codons represents their common ancestor, the G+C content of Patience had changed significantly toward reduction in the G+C content, whereas the G+C content of Fruitloop has remained unchanged. This result indicated that Fruitloop is more conservative than Patience and the ancestor of bacteriophage of *Mycobacterium smegmatis* have higher G+C content than

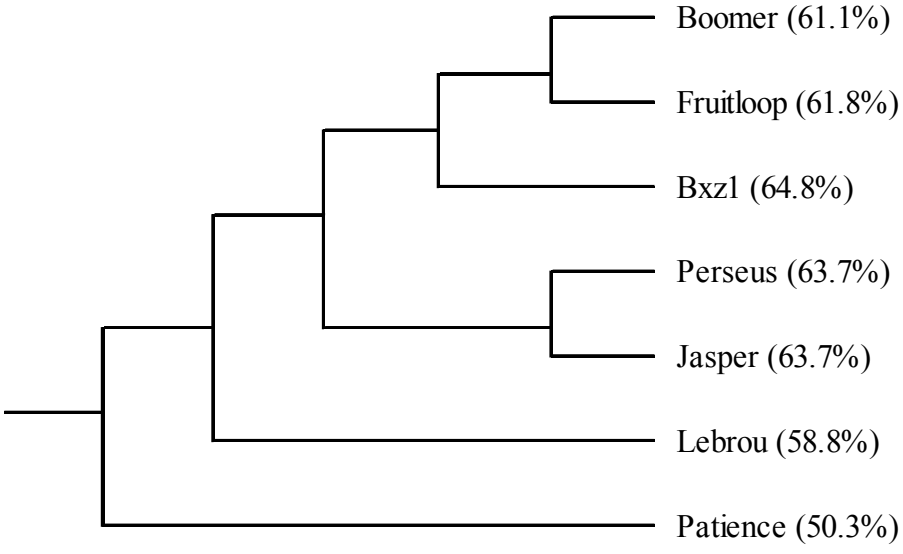

Fig. 2. A phylogenetic tree obtained from the orthologs of seven phages. The genomic G+C content (%) of the phages is indicated in the parentheses

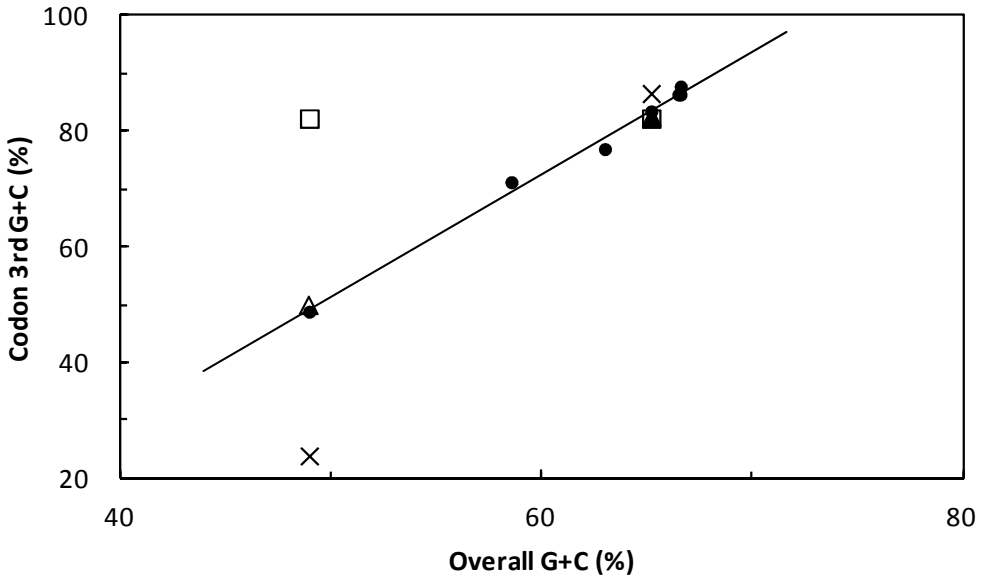

Fig. 3. G+C content at the third codon position versus the total G+C content for the orthologous DNA sequences of seven phages (filled circles)

The G+C content at the third position for the three groups of codons of aligned DNA sequence of Patience and that of Fruitloop is plotted: crosses for IA codons, open triangles for DA codons, and open squares for IC codons

Patience. The analysis of orthologs of Lebrou and Bxz1 showed that the ancestor of bacteriophage of *Mycobacterium smegmatis* had higher G+C content than Lebrou. These results suggested that Patience and Lebrou had branched early (Fig. 2) and their genomes changed toward reduction in G+C content. This result indicated that after the infection established, the G+C content of bacteriophage had changed.

The bacteriophage of *Mycobacterium smegmatis* and that of *Pseudomonas aeruginosa* showed similar G+C content to their host, even though they have potential of having large deviation of G+C content compared to their host. In this study, two phages in Aeromonas genus and one in Rhodothermus showed large deviation of G+C content compared to that of the hosts. Actually, the phages in Aeromonas genus have wide range of G+C content (37% in phage 65 – 62% in phage phiO18P). In Rhodothermus, only one phage genome sequence is determined, so we do not know whether the phages in Rhodothermus have wide range of G+C content or not. It is interesting that the genus of Mycobacterium, Pseudomonas, Aeromonas, and Rhodothermus have genomic G+C content around 60%. It is not clear that this character is related to the large deviation in G+C content between phage and host genomes.

Infection of bacteriophage to its host need many steps such as bacteriophage attachment to the host, release its DNA into the host cell, replication of bacteriophage DNA, release new bacteriophages. Relation of the similar G+C content between bacteriophage and host genome is not included in the above steps directly, but it might be involved in replication step. This hypothesis need to be validated.

4. CONCLUSION

The relationship of G+C content between bacteriophages/plasmids and host genomes was almost linear. Three bacteriophages were largely deviated from the linear relation. A phylogenetic tree obtained using the orthologs of *Mycobacterium smegmatis* indicated which bacteriophage branched early. Assuming that the G+C content of identical codons represents their common ancestor, the ancestor was estimated that it had similar G+C content with its host. This result suggested that similar G+C content between bacteriophages and host genomes might be related in phage infection.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Wommack KE, Colwell RR. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev. 2000;64:69-114.
2. Médigue C, Rouxel T, Vigier P, Hénaut A, Danchin A. Evidence for horizontal gene transfer in *Escherichia coli* speciation. J Mol Biol. 1991;222:851-856.
3. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000;405:299-304.
4. Dutta C, Pan A. Horizontal gene transfer and bacterial diversity. J Biosci. suppl 1, 2002;27:27-33.
5. Ochman H, Lerat E, Daubin V. Examining bacterial species under the specter of gene transfer and exchange. Proc Natl Acad Sci USA. 2005;(suppl 1)102:6595-6599. DOI:10.1073/pnas.0502035102.
6. Muto A, Osawa S. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci USA. 1987;84:166-169. DOI:10.1073/pnas.84.1.166.
7. Lawrence JG, Ochman H. Amelioration of bacterial genomes: Rates of change and exchange. J Mol Evol. 1997;44:383-397. DOI:10.1007/PL00006158.
8. Karlin S, Burge C. Dinucleotide relative abundance extremes: A genomic signature. Trends Genet. 1995;11:283-290.
9. Karlin S, Mrázek J, Campbell AM. Compositional biases of bacterial genomes and evolutionary implications. J Bacteriol. 1997;179:3899-3913.
10. Nakashima H, Ota M, Nishikawa K, Ooi T. Genes from nine genomes are separated into their organisms in the dinucleotide composition space. DNA Res. 1998;5:251-259. DOI:10.1093/dnares/5.5.251.
11. Kwan T, Liu J, DuBow M, Gross P, Pelletier J. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc Natl Acad Sci USA. 2005;102:5174-5179. DOI:10.1073/pnas.0501140102.
12. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth, C, Lewis JA, et al. Origins of highly mosaic mycobacteiophage genomes. Cell. 2003;113:171-182.
13. Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Bustamante CG, Boyle MM, et al. On the nature of mycobacteriophage diversity and host preference. Virology. 2012;434:187-201. DOI:10.1016/j.virol.2012.09.026.
14. Kwan T, Liu J, DuBow M, Gross P, Pelletier J. Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. J Bacteriol. 2006;188: 1184-1187. DOI:10.1128/JB.188.3.1184-1187.2006.
15. Bellgard MI, Gojobori T. Significant differences between the G+C content of synonymous codons in orthologous genes and the genomic G+C content. Gene. 1999;238:33-37.
16. Fukuchi S, Homma K, Sakamoto S, *et al*. The GTOP database in 2009: updated content and novel features to expand and deepen insights into protein structures and functions. Nucleic Acids Res. 2009; 37:D333-D337. DOI:10.1093/nar/gkn855.
17. Chan PP, Holmes AD, Smith AM, Tran D, Lowe TM. The UCSC archaeal genome browser: 2012 update. Nucleic Acid Res. 2012;40:D646-D652.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403-410. DOI:10.1016/S0022-2836(05)80360-2.
19. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–4680.
20. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at Genome Net. Nucleic Acids Res. 2002; 30:42-46.
21. Rocha EPC, Danchin A. Base composition bias might result from competition for metabolic resources. Trends Genet. 2002; 18:291-294.
22. Nishida H. Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. Int J Evol Biol. 2012;2012:342482. DOI:10.1155/2012/342482.
23. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature. 2000;406:477-483.
24. Chargaff E, Lipshitz R, Green C, Hodes ME. The composition of the

desoxyribonucleic acid of salmon sperm. J Biol Chem. 1951;192:223-230.
25. Chargaff E, Lipshitz R, Green C. Composition of the desoxypentose nucleic acids of four genera of sea-urchin. J Biol Chem. 1952;195:155-160.
26. Karkas JD, Runder R, Chargaff E. Separation of *B. subtilis* DNA into complementary strands, II. Template functions and composition as determined by transcription with RNA polymerase. Proc Natl Acad Sci USA. 1968;60:915-920.
27. Runder R, Karkas JD, Chargaff E. Separation of *B. subtilis* DNA into complementary strands, III. Direct analysis. Proc Natl Acad Sci USA. 1968;60:921-922.
28. Mitchell D, Bridge R. A test of Chargaff's second rule. Biochem Biophys Res Commun. 2006;340:90-94. DOI:10.1016/j.bbrc.2005,11.160.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/10615