

INTERPLAY BETWEEN BACTERIOPHAGES AND RESTRICTION-MODIFICATION SYSTEMS IN ENTEROCOCCI

PETER PRISTAS, ANNA VANDZUROVA, PETER JAVORSKY

Institute of Animal Physiology, Slovak Academy of Sciences, Soltesovej 4-6, 040 01 Kosice, Slovakia, (pristas@saske.sk)

Abstract: The complete genomes of *Enterococcus faecalis* bacteriophages were analyzed for tetranucleotide words avoidance. Very similar tetranucleotide composition was found in all tested genomes with strong underrepresentation of palindromic GATC and GGCC words. This avoidance could be explained as a protection mechanism against host restriction-modification systems as a clear correlation was found between avoidance of palindromic words and the specificity of *E. faecalis* restriction and modification systems. No similar avoidance of tetranucleotide words was observed for non-palindromic words. A weak correlation was observed between avoidance of tetranucleotide palindromes in bacteriophage genomes and the possession of phage encoded DNA methyltransferases confirming the interrelation between bacteriophage genomes composition and restriction and modification systems in enterococci.

Key words: Enterococcus, bacteriophage, restriction modification systems, palindrome avoidance

1. Introduction

Type II restriction-modification systems entail a DNA methyltransferase and an endonuclease of the same recognition sequence specificity. The endonuclease digests foreign DNA that enters the cell, thereby protecting the bacteria from genetic subversion. The methylase modifies the cell's DNA, thereby protecting it from similar digestion. It is generally accepted that restriction systems in bacteria primarily act to protect the organism from foreign DNA, particularly from infection by bacteriophages (BICKLE and KRUGER, 1993). Bacteriophages on other hand have evolved antirestriction mechanisms, encode for their own methyltransferases, and are frequently deficient in recognition sites for restriction endonucleases (SAMSON *et al.*, 2013). The aim of our study was to analyze the tetranucleotide composition of enterococcal bacteriophage genomes, to analyze the effect of phage encoded modification methyltransferases on avoidance of tetranucleotides, and to compare the frequency and variability of restriction endonucleases encoded by enterococci.

Enterococci are ubiquitous bacteria present in the environment, in the gastrointestinal tract of healthy animals and humans, and in foods, especially those of animal origin such as dairy products (GIRAFFA, 2003). Enterococci entry into milk and milk products through the water supply, equipment, and unsanitary and unhygienic conditions during production and handling. In milk products, they are used as probiotics resulting in positive effects on human digestibility. Thanks to the efficient utilisation of organic acids, enterococci contribute to the development of unique sensory characteristics in fermented dairy products. In contrast to these positive roles, some enterococcal strains were suspected to have pathogenic properties for

humans (HUNT, 1998). Enterococcal bacteriophages with very effective bactericidal activity are potentially used as novel antibacterial agents to control enterococcal infections (DERESINSKI, 2009), however in dairy industry the bacteriophage infections result in unacceptably low production of lactic acid and flavour compounds along with reduced proteolysis and led to the major losses in fermentation processes (MC GRATH *et al.*, 2007). Better understanding of bacteriophage genetics and biology could improve this dual role of enterococcal bacteriophages.

We have used Markov chain analysis to evaluate which nucleotide DNA words are over- or underrepresented in the genomes of *E. faecalis* bacteriophages. Markov analysis seems to be the most widely used to evaluate palindrome distribution (PANINA *et al.*, 2000). Briefly, using this model it is possible to calculate how often a word should appear in a sequence on the basis of knowledge of the distribution of the word's fragments. Under the maximum applicable order of the Markov model, the expected count of a particular tetranucleotide (K), e.g. GATC sequence, is defined as:

$$K_{(GATC)} = N_{(GAT)} \times N_{(ATC)} / N_{(AT)} \quad (1)$$

where $N_{(GAT)}$, $N_{(ATC)}$ and $N_{(AT)}$ are the observed counts of oligonucleotides. Similar formulas have been widely used (FUGLSANG, 2003). This number can then be compared with the actual count of the word in the sequence. However, the difference itself does not tell to what extent a word is over- or underrepresented. For this purpose, the normalized statistic Z is used, as proposed by SCHBATH (1997), and used by others for similar purposes. Z (contrast value) is positive for overrepresented words and negative for underrepresented words, and is calculated as:

$$Z = \frac{[N_{(GATC)} - K_{(GATC)}] / \sqrt{n}}{\sqrt{V_{(GATC)}}} \quad (2)$$

where n is number of nucleotides in analysed sequence and the variance $V_{(GATC)}$ is calculated as:

$$V_{(GATC)} = K_{(GATC)} \frac{[N_{(AT)} - N_{(ATC)}] \times [N_{(AT)} - N_{(GAT)}]}{N^2_{(AT)}} \quad (3)$$

Based on the normalized nature of Z , the probability of random observation of $|Z| > 3.29$ is < 0.001 .

The data obtained indicate strong underrepresentation of palindromic tetranucleotides in *E. faecalis* bacteriophage sequences which could be probably explained as a protection mechanism against host restriction-modification systems.

2. Materials and methods

Complete genome sequences of *E. faecalis* bacteriophages and *E. faecalis* genome (RefSeq accession number NC_004668) were taken from the RefSeq NCBI Reference

Sequence Database available at <http://www.ncbi.nlm.nih.gov/refseq/>. The presence (marked as *Y* in Table 1) or absence (marked as *N* in Table 1) of modification methyltransferase in the genomes of bacteriophages was taken from annotations available in RefSeq database. For the list of bacteriophage sequences used through the study see Table 1. The frequency of restriction endonucleases in *E. faecalis* was taken from Rebase database (ROBERTS *et al.*, 2010).

Table 1. The list and corresponding Z score of bacteriophage sequences used in the study.

Bacteriophage	Accession number	Genome size (bp)	MTase ^a	Tetranucleotide average z-score		The most underrepresented word (score)	Reference
				all	palindrome		
phiFL4A	NC_013644	37856	Y	0	-3.62	GATC (-9.55)	YASMIN <i>et al.</i> 2010
phiFL3A	NC_013648	39576	Y	0	-3.39	GATC (-9.34)	YASMIN <i>et al.</i> 2010
phiFL2A	NC_013643	36270	Y	0	-3.19	GATC (-9.52)	YASMIN <i>et al.</i> 2010
phiFL1A	NC_013646	38764	Y	0	-3.14	GATC (-8.77)	YASMIN <i>et al.</i> 2010
phiEF24C	NC_009904	142072	N	0	-5.85	GATC (-27.68)	UCHIYAMA <i>et al.</i> 2008
phiEfl1	NC_013696	42822	Y	0	-3.21	GATC (-9.09)	STEVENS <i>et al.</i> 2011
EFRM31	NC_015270	16945	N	0	-2.95	GATC (-8.85)	FARD <i>et al.</i> 2010
EFAP-1	NC_012419	21115	N	0	-2.77	GATC (-12.76)	SON <i>et al.</i> 2010
EF62phi	NC_017732	30505	N	0	-2.82	CGCG (-6.07)	BREDE <i>et al.</i> 2011
BC-611	NC_018086	53996	N	0	-5.55	GATC (-19.79)	HORIUCHI <i>et al.</i> 2012

^aY in the Mtase column indicate the presence, N the absence of modification methyltransferase in the genome of bacteriophage

Tetranucleotide counts, Z score values, and the Pearson's correlation coefficients were calculated using Tetra software (TEELING *et al.*, 2004). The matrix of correlation coefficients was converted into distance matrix using DAMBE software version 5 and similarity dendrogram was constructed using Neighbor-Joining algorithm implemented in the software (XIA, 2013). Statistica® software package (StatSoft, Tulsa, Oklahoma) was used to compare the Z (contrast value) distribution between datasets using Pearson's correlation coefficient.

3. Results and discussion

Bacteriophages, although simple in organization, are the most diverse life forms in the biosphere (ACKERMANN and KROPINSKI, 2007). Bacteriophage life cycle completely relies on bacterial host. During the lytic cycle, bacteriophage infection redirects host metabolism towards the replication of the phage nucleic acid and assembly of new phage particles, which are then released upon cell death and lysis. The bacteriophages thus have an important role in bacterial evolution and have led to a great variety of defense mechanisms. Bacteriophages, for their part, have developed

counter defense mechanisms to evade the bacterial defense mechanisms (SAMSON *et al.*, 2013). One of the best studied systems to protect bacteria from invading bacteriophages is the possession of restriction-modification systems. These systems entail a DNA methyltransferase and an endonuclease of the same recognition sequence specificity. The endonuclease recognizes short, usually palindromic oligonucleotides and digests bacteriophage DNA that enters the cell, thereby protecting the bacteria from genetic subversion. The methylase modifies the cell's DNA, thereby protecting it from similar digestion (WILSON, 1991). Bacteriophages have developed anti-restriction systems and are frequently deficient in oligonucleotides recognized by restriction endonucleases. Bacteriophages changed their genomes by adopting point mutations which reduce the number of restriction sites (TOCK and DRYDEN, 2005).

Using TETRA software palindrome deficiency in genomes of all available enterococcal bacteriophages was analyzed. A very similar pattern of palindrome avoidance was observed in genomes of enterococcal bacteriophages as well as in *E. faecalis* genome. While the frequencies of all tetranucleotide words were found to be normally distributed around Z value 0, frequencies of palindromes were found to be strongly underrepresented in both bacteriophage and *E. faecalis* genomes (Fig. 1). For bacteriophages Z values of tetranucleotide palindromes were in range from 7.51 to -27.68. Average Z value of tetranucleotide palindromes for all bacteriophages were in range from -2.77 to -12.09, indicating strong underrepresentation of tetranucleotide palindrome words in genomes of *E. faecalis* bacteriophages. The most underrepresented word in all but one tested bacteriophage genome was GATC word (average Z value -12.09, see Table 1).

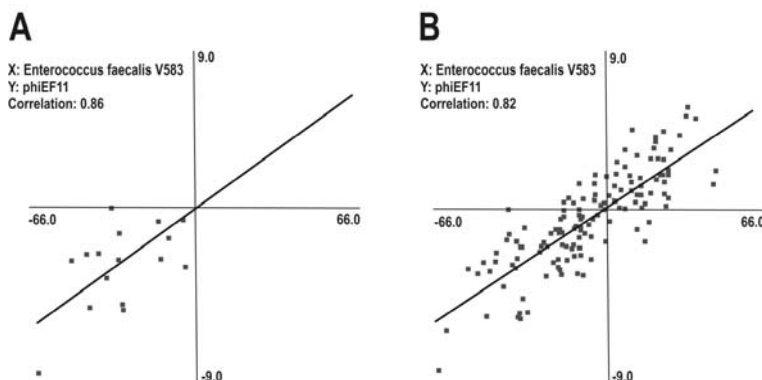


Fig. 1. Correlation of Z values of *Enterococcus faecalis* V583 complete genome (RefSeq accession number NC_004668) and genome of phiEfl1 bacteriophage (RefSeq accession number NC_013696). Correlation of tetranucleotide palindromes (part A) compared to all palindromes (part B) is shown.

Several authors showed that the lowest contrast values show palindromes which serve as a recognition sites of most frequently occurring endonucleases. ROCHA *et al.* (1998), reported that GATC and GGCC palindromes are between 6 the most underrepresented tetranucleotide palindromes in complete genome of *Bacillus subtilis*. Surprisingly, much higher degree of tetranucleotide palindrome avoidance was

observed in the genome of *E. faecalis* V583 compared to bacteriophage genomes (Table 2). While average Z score observed for bacteriophage genomes was -3.70 much lower Z score (-30.55) was observed for *E. faecalis* V583 chromosome. However a correlation was observed between Z values of bacteriophages and chromosome (Pearson Correlation Coefficient R=0.68). GATC and GGCC tetranucleotide palindromes are among 4 the most underrepresented words in both genomes. Clear correlation was observed between the word avoidance and the existence of restriction-modification systems with given specificity (Table 2). From the *E. faecalis* bacterium restriction endonucleases recognizing GATC, GGCC, CCGG, and CGCG tetranucleotide are described (VANAT *et al.*, 1993; RADLINSKA *et al.*, 2005; anonymous at rebase.neb.com, and our unpublished data) and these 4 tetranucleotide words are the most underrepresented in the genome of *E. faecalis*. These 4 words show the lowest Z score compared to other words as well. Significant underrepresentation of other tetranucleotide words was observed in the genome of *E. faecalis* V583 indicating that much higher number of restriction-modification systems than currently known has occurred in the genome of the *E. faecalis*, each of them leaving a trace of underrepresentation of a short palindrome.

Table 2. Correlation between the frequency of tetranucleotide palindromes in the genomes of *E. faecalis* bacteriophages and *E. faecalis* V583 complete genome and frequency of restriction-modification systems in *E. faecalis*. ^a - denotes lack of RMS, + the presence of RMS

Tetranucleotide	Bacteriophages		Chromosome		Known RMS ^a
	z-score	rank	z-score	rank	
AATT	-6.21	3	-29.54	11	-
AGCT	-2.16	12	-34.78	7	-
ACGT	-3.66	8	-10.78	14	-
ATAT	1.03	15	-5.00	15	-
GATC	-12.69	1	-66.48	1	+
GGCC	-4.81	4	-51.21	2	+
GCGC	-2.81	11	-3.93	16	-
GTAC	1.09	16	-14.21	13	-
CATG	-6.27	2	-30.05	10	-
CGCG	-4.79	5	-43.59	4	+
CCGG	-4.34	6	-44.99	3	+
CTAG	-3.27	9	-15.07	12	-
TATA	-3.98	7	-36.53	6	-
TGCA	-1.82	12	-31.68	8	-
TCGA	-2.93	10	-31.25	9	+
TTAA	-1.54	14	-39.77	5	-

Bacteriophages frequently employ additional strategies to overcome host restriction and modification systems e.g. site specific modification of the phage genome by bacteriophage encoded modification methyltransferases (SAMSON *et al.*, 2013). Among *E. faecalis* bacteriophages modification methyltransferase was found in the genomes of 5 from 10 tested bacteriophages (Table 1).

Based on degree of 4 bp palindromes avoidance the matrix of similarity coefficients between all pairs of bacteriophage genomes was constructed and the similarity tree was constructed indicating that bacteriophages encoding

methyltransferase (**phiFL4A**, **phiFL3A**, **phiFL2A**, **phiFL1A**, and **phiEf11**, shown in bold in Fig. 2) have slightly different 4 bp palindrome composition. While average Z score of palindrome avoidance in bacteriophages possessing modification methyltransferase was -3.31, Z score in bacteriophages lacking modification methyltransferase was -3.99. This is probably due to protective effect of methyltransferases and decreased pressure on avoidance of tetranucleotide words.

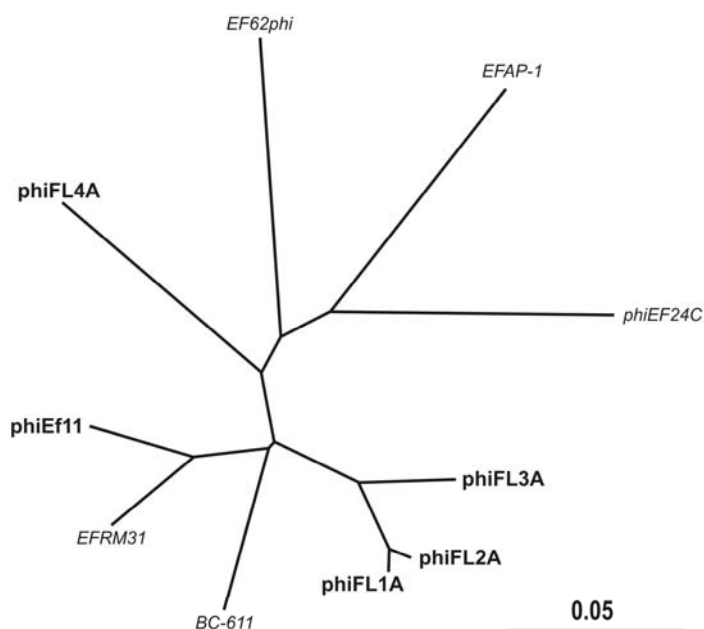


Fig. 2. Neighbor-Joining tree showing the relatedness of palindrome avoidance in bacteriophages possessing modification methyltransferase gene (shown in bold) or not. The bar indicates distance level 0.05.

4. Conclusions

Strong underrepresentation of palindromic tetranucleotide words was observed in genomes of *Enterococcus faecalis* bacteriophages and host bacterium. The most underrepresented are GATC and GGCC words. This avoidance could be explained as a protection mechanism against host restriction-modification systems as a clear correlation was found between avoidance of palindromic words and the specificity of *E. faecalis* restriction-modification systems. No similar avoidance of tetranucleotide words was observed for non-palindromic words. A weak correlation was observed between avoidance of tetranucleotide palindromes in bacteriophage genomes and the possession of phage encoded DNA methyltransferases confirming the interrelation between bacteriophage genomes composition and restriction-modification systems in enterococci.

Acknowledgments: The work was financially supported by the European Regional Development Fund project 26220220065.

References

- ACKERMANN, H.W., KROPINSKI, A.M.: Curated list of prokaryote viruses with fully sequenced genomes. *Res. Microbiol.*, 158, 2007, 555-566.
- BICKLE, T.A., KRUGER, D.H.: Biology of DNA restriction. *Microbiol. Rev.*, 57, 1993, 434-450.
- BREDE, D.A., SNIPEN, L.G., USSERY, D.W., NEDERBRAGT, A.J., NES, I.F.: Complete genome sequence of the commensal *Enterococcus faecalis* 62, isolated from a healthy Norwegian infant. *J. Bacteriol.*, 193, 2011, 2377-2378.
- DERESINSKI, S.: Bacteriophage Therapy: Exploiting Smaller Fleas. *Clin. Infect. Dis.*, 48, 2009, 1096-1101.
- FARD, R.M.N., BARTON, B.D., HEUZENROEDER, M.: Novel Bacteriophages in *Enterococcus* spp. *Curr. Microbiol.*, 60, 2010, 400-406.
- FUGLSANG, A.: Distribution of potential type II restriction sites (palindromes) in prokaryotes. *Biochem. Biophys. Res. Commun.*, 310, 2003, 280-285.
- GIRAFFA, G.: Functionality of enterococci in dairy products. *Int. J. Food. Microbiol.*, 88, 2003, 215-222.
- MC GRATH, S., FITZGERALD, G.F., VAN SINDEREN, D.: Bacteriophages in dairy products: pros and cons. *Biotechnol. J.*, 2, 2007, 450-455.
- HORIUCHI, T., SAKKA, M., HAYASHI, A., SHIMADA, T., KIMURA, T., SAKKA, K.: Complete genome sequence of bacteriophage BC-611 specifically infecting *Enterococcus faecalis* strain NP-10011. *J. Virol.*, 86, 2012, 9538-9539.
- HUNT, C.P.: The emergence of enterococci as a cause of nosocomial infection. *Br. J. Biomed. Sci.*, 55, 1998, 149-156.
- PANINA, E.M., MIRONOV, A.A., GELFAND, M.S.: Statistical analysis of complete bacterial genomes: Avoidance of palindromes and restriction-modification systems. *Mol. Biol.*, 34, 2000, 215-221.
- RADLINSKA, M., PIEKAROWICZ, A., GALIMAND, M., BUJNICKI, J.M.: Cloning and preliminary characterization of a GATC-specific beta(2)-class DNA:m(6)A methyltransferase encoded by transposon Tn1549 from *Enterococcus* spp. *Pol. J. Microbiol.*, 54, 2005m249-252.
- ROBERTS, R.J., VINCZE, T., POSFAI, J., MACELIS, D.: REBASE-a database for DNA restriction and modification: enzymes, genes and genomes. *Nucl. Acids Res.*, 38, 2010, D234-D236.
- ROCHA, E.P.C., VIARI, A., DANCHIN, A.: Oligonucleotide bias in *Bacillus subtilis*: general trends and taxonomic comparisons. *Nucl. Acids Res.*, 26, 1998, 2971-2980.
- SAMSON, J.E., MAGADAN, A.H., SABRI, M., MOINEAU, S.: Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.*, 11, 2013, 675-687.
- SCHBATH S: An efficient statistic to detect over- and under-represented words in DNA sequences. *J. Comput. Biol.*, 4, 1997, 189-192.
- SON, J.S., JUN, S.Y., KIM, E.B., PARK, J.E., PAIK, H.R., YOON, S.J., KANG, S.H., CHOI, Y.J.: Complete genome sequence of a newly isolated lytic bacteriophage, EFAP-1 of *Enterococcus faecalis*, and antibacterial activity of its endolysin EFAL-1. *J. Appl. Microbiol.*, 108, 2010, 1769-1779.

- STEVENS, R.H., EKTEFAIE, M.R., FOUTS, D.E.: The annotated complete DNA sequence of *Enterococcus faecalis* bacteriophage ϕ Ef11 and its comparison with all available phage and predicted prophage genomes. *FEMS Microbiol. Lett.*, 317, 2011, 9-26.
- TEELING, H., WALDMANN, J., LOMBARDOT, T., BAUER, M., GLÖCKNER, F.O.: TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics*, 5, 2004, 163.
- TOCK, M.R., DRYDEN, D.T.F.: The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, 8, 2005, 466-472.
- UCHIYAMA, J., RASHEL, M., MAEDA, Y., TAKEMURA, I., SUGIHARA, S., AKECHI, K.: Isolation and characterization of a novel *Enterococcus faecalis* bacteriophage phi EF24C as a therapeutic candidate. *FEMS Microbiol. Lett.*, 278, 2008, 200-206.
- VANAT, I., PRISTAS, P., KUTEJOVA, E., JUDOVA, J., GODANY, A., JAVORSKY, P.: SbvI restriction endonuclease from *Streptococcus bovis*. *Let. Appl. Microbiol.*, 17, 1993, 297-299.
- WILSON, G.G.: Organization of restriction-modification systems. *Nucl. Acids Res.*, 19, 1991, 2539-2566.
- XIA, X.: DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.*, 30, 2013, 1720-1728.
- YASMIN, A., KENNY, J. G., SHANKAR, J., DARBY, A.C., HALL, N., EDWARDS, C., HORSBURGH, M. J.: Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. *J. Bacteriol.*, 192, 2010, 1122-1130.