comment
reviews
reports
deposited research
refereed research
interactions
information

Deposited research article

A Protein Similarity Approach For Detecting Prophage Regions In Bacterial Genomes

Geeta V Rao, Preeti Mehta, Srividhya KV and Krishnaswamy S[1][§]

Address: [1]Bioinformatics Center, School of Biotechnology, Madurai Kamaraj University, Madurai - 625021, India.

Correspondence: [§]S Krishnaswamy. Email: krishna@mrna.tn.nic.in

A Protein Similarity Approach For Detecting Prophage Regions In Bacterial Genomes

Geeta V Rao, Preeti Mehta, Srividhya KV and Krishnaswamy S[1][§]

[1] Bioinformatics Center, School of Biotechnology, Madurai Kamaraj University, Madurai - 625021

[§]Corresponding author

Email addresses:
GVR: g_v5@rediffmail.com
PM: mehta_p74@yahoo.com
KVS: vidhya@mkustrbioinfo.com
SK: krishna@mrna.tn.nic.in

Abstract

Background

Numerous completely sequenced bacterial genomes harbor prophage elements. These elements have been implicated in increasing the virulence of the host and in phage immunity. The e14 element is a defective lambdoid prophage element present at 25 min in the *Escherichia coli* K-12 genome. e14 is a well-characterized prophage element and has been subjected to in-depth bioinformatic analysis.

Results

A protein-based comparative approach using BLAST helped identify lambdoid-like prophage elements in a representative set of completely sequenced bacterial genomes. Twelve putative prophage regions were identified in six different bacterial genomes. Examination of the known and newly identified prophage regions suggests that on an average, the prophage elements do not seem to occur either randomly or in a uniform manner along the genome amongst genomes of the selected pathogenic organisms.

Conclusion

The protein based comparative approach can be effectively used to detect lambdoid-like prophage elements in bacterial genomes. It is possible that this method can be extended to all prophage elements and can be made automated.

Background

Bacterial genome nucleotide sequences are being completed at a rapid and increasing rate, thanks to faster and better sequencing techniques. Many completely sequenced bacterial genomes harbor temperate bacteriophages, both functional and defective. The gene products encoded by prophages can have very important effects on the host bacterium, ranging from protection against further phage infection to increasing the virulence of a pathogenic host. Numerous virulence factors from bacterial pathogens are phage encoded [1,2,3] for example, the food poisoning botulinus toxin and *Vibrio cholerae*. The latter is a fascinating case of how multiple phages contribute to bacterial pathogenicity. It is postulated that some adaptations of nonpathogenic bacterial strains to their ecological niche might also be mediated by prophage genomes [4]. As mobile DNA elements, phage DNA is a vector for lateral gene transfer between bacteria [5]. As reviewed by Canchaya *et al* [6] technically difficulty relies in defining prophage sequences in bacterial genomes as mostly they are cryptic or in the state of mutational decay.

Prophages account for a substantial amount of interstrain genetic variability in several bacterial species, for example *Staphylococcus aureus* [7] and *Streptococcus pyogenes* [8]. When genomes from closely related bacteria were compared in a dot-plot analysis, prophage sequences accounted for a major proportion of the differences between the genomes , for example, *Listeria monocytogenes* and *Listeria innocua* [9] and *Escherichia coli* O157 and K-12 [10]. When mRNA expression patterns were studied using microarrays in lysogenic bacteria that underwent physiologically relevant changes in growth conditions, prophage genes figured prominently in the mRNA species changing their expression pattern [11,12]. These data demonstrate that prophages are not a passive genetic cargo of the bacterial chromosome, but are active participants in cell physiology. The medical and evolutionary importance of prophages makes it important that one is able to recognize and understand prophages when they are present.

Recognizing prophages in bacterial genome sequences is not a straightforward task. Even if the search for prophage elements is restricted to tailed temperate phages (there

are other kinds of temperate DNA phages [13,14]) none of the phage genes are sufficiently conserved to serve as a single marker for prophages, and in any given case, any particular gene could have been deleted from a defective prophage [15,16]. Therefore, using a single gene like integrase or terminase might not be complete for prophage identification. Some prophages have different G+C contents, oligonucleotide frequencies or codon usage from their host genome, but this type of analysis has not progressed to the point that it can unequivocally identify prophage sequences [17]. One must therefore identify prophages in bacterial genome sequences by the similarity of their gene sequences and gene organization to known prophage genes.

E. coli and other enterobacterial genomes are recognized to contain a number of lambda-like cryptic prophages. For example, the very well characterized *E. coli* K-12 genome carries eight convincingly identified prophages and six of these, DLP-12, e14, Rac, QIN, CPS-53, and Eut are lambdoid in nature. A comprehensive bioinformatic analysis has been carried out on the e14 sequence [18]. This analysis showed the modular nature of the e14 element, and that it shares a large part of its sequence with the *Shigella flexneri* phage SfV. Based on this similarity, the regulatory region including the repressor and Cro proteins and their binding sites were identified.

The e14 element is 15.4 kbp long and lies between 1195432 bp and 1210646 bp on the K-12 chromosome. The element uses a homologous region of 216 bases in the *icd* gene as the integration site, though the actual crossover for integration occurs within the first 11 bases at one end of the homology [19]. The integration event caused only two amino acid changes in the isocitrate dehydrogenase protein. The element is capable of excision if the SOS response is triggered. Both excision and re-integration occur in a site-specific manner [20,21]. The e14 element was mapped on the *E. coli* K12 chromosome and cloned by van de Putte *et al* [22]. The element is known to encode several important functions including the *lit* gene involved in T4 exclusion [23,24], the *rglA* (*mcrA*) gene involved in restriction of hydroxymethylated nonglucosylated T4 phages [25,26] and the *pin* gene involved in inversion of an adjacent 1800-basepair segment [22,27]. The element also encodes a Kil function and the concomitant repressor protein [28] and an SOS induced cell division inhibition function attributed to the *sfiC* gene [29].

A protein based COG approach helped detect lambdoid-like prophage elements in a set of eight completely sequenced bacterial genomes [18]. This approach is different from the other approaches in that it does not rely on a single gene like integrase or terminase for prophage detection, but has the potential to use the entire known pool of temperate tailed phage-encoded genes for detection against the COG data [30]. Such a comparative protein level approach can be effectively used to detect defective lambdoid-like prophage elements in bacterial genomes.

Results and Discussion

The e14 element is a very well characterized prophage element [18], which contains all the highly conserved prophage genes like the phage portal and terminase genes. This analysis [18] also involved a protein based COG approach for identifying similar prophages. This takes into consideration the modular nature of prophage genomes and looks for homologs of the genes of the prophage e14 that exist in proximity to each other. The same idea was utilized in this study. The choice of e14 proteins as template for similarity searches for prophage elements was retained as in the earlier analysis. However the search procedure (BLAST instead of COG) was modified in view of possible automation and flexibility. A larger set of genomes from 40 pathogenic organisms were scanned in this analysis.

Identifying prophage elements in bacterial genomes

A set of forty bacterial genomes was chosen for prophage detection, and only the ones that yielded significant BLAST hits (e < = 0.01) are listed in Tables 1 and 2. The BLAST searches were carried out organism-wise and then the hits were sorted based on the locus of occurrence in the genome. Lone hits were analyzed to check whether they form part of prophages reported in literature, and if so, they are included in Table 1.

Genes encoding the BLAST hits for the different e14 proteins, which were within a particular distance (this distance varies from one organism to another; it is the size of the longest prophage in the organism's genome) were then clubbed together. Any

region with two or more genes in this cluster were considered as putative prophage elements and further analyzed. Most of these clusters belong to pre-annotated prophage elements, but twelve putative prophage elements were identified in six organisms- *S.flexneri* 2457T, *S. enterica* LT2 (serovar Typhimurium), *S. pyogenes* M18 MGAS8232, *S. pyogenes* M3 MGAS315, *Vibrio cholerae* N16961 and *P. luminescens subsp. laumondii*TTO1. For the former, prophage regions were delimited using data from the prophage database [31] and from literature [32]. As for the putative prophage regions, the prophage limits are reported from the first hit to the last hit in each cluster (data taken from .ptt files from ftp://ftp.ncbi.nih.gov/genomes/). Prophage loci given in parentheses represent possible outer limits for the prophage regions (Table 2). The genes forming part of these outer limits were not picked up in the similarity searches, but are reported here because they are prophage-related proteins or have strong similarity to prophage proteins.

Of the twelve putative prophage regions identified, five are located near dehydrogenase genes (Table 3). *A priori* there seems to be no attributable reason to this tendency for the putative lambdoid phages to get integrated near a dehydrogenase gene in the bacterial genome. However, it must be noted that the search template e14 is also integrated at the isocitrate dehydrogenase gene in the *E. coli* K12 genome.

Prophage distribution

In order to address the question whether the prophage elements integrate in a random and isotropic manner into bacterial genomes, these genomes were brought into a common reference frame to facilitate comparison. All genome lengths were normalized to 1000 units and prophage coordinates (both known and newly identified ones) were re-calculated in terms of these normalized units. The distribution of prophage elements (Figure 1) is found to be uni-modal with a maximum frequency of occurrence in the range of 400-600 genome units. On an average, the prophage elements do not seem to occur either in a random or in a uniform manner along the genome amongst genomes of the selected pathogenic organisms.

Conclusion

We could identify several lambdoid prophage elements in a representative set of bacterial genomes using a protein similarity approach. It has been observed that lambdoid phages have a strong tendency to get integrated near a dehydrogenase gene in the bacterial genome. A prophage distribution study shows that most of the prophages are found in comparable regions in the bacterial genomes. This exercise was knowingly limited by only taking genes similar to that of e14 into consideration. A similar approach using the entire pool of known lambdoid prophage (or even all temperate prophage) genes with appropriate weighting for the frequency of occurrence of the prophage proteins, should make a much more sensitive and robust technique for detecting prophage elements.

Materials and Methods

The local version of the WWW-BLAST [33,34] was installed and used for sequence analysis. In order to identify e14 homologs, similarity searches at the protein level were done taking the twenty-three e14 proteins as query and the bacterial proteomes as target. The bacterial proteomes were downloaded from NCBI's FTP site (ftp://ftp.ncbi.nih.gov/genomes/). Similarity searches were done using BLASTP with default values. Only the significant hits (e <= 0.01) were used for the analysis.
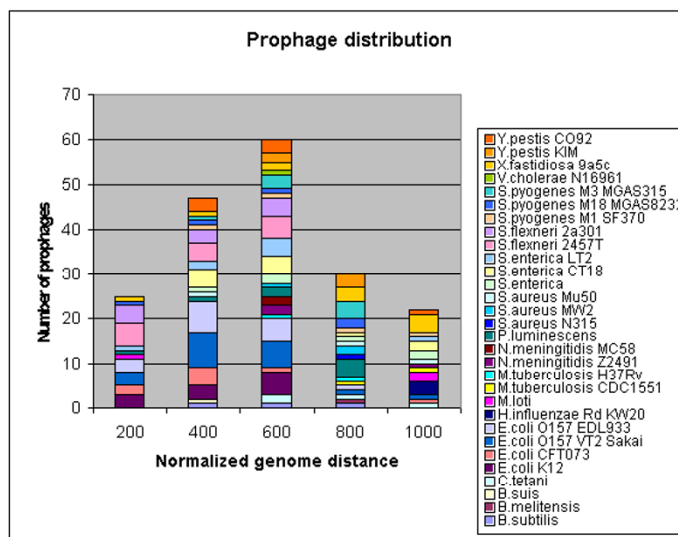
Figures

Figure 1

Figure Legends.

Figure 1

Comparative prophage distribution across genomes

All genome lengths were normalized to 1000 units and prophage loci for both known and newly identified ones were calculated in terms of these normalized units. The graph was drawn taking normalized genome distance along X-axis and the number of prophages along Y-axis.

Table 1: **Prophage elements identified but already known.** Prophage elements detected in other genomes using similarity to e14 proteins as a criterion. BLAST hits for the e14 proteins in different organisms were examined, and only the significant hits (e < = 0.01) are listed. The boundaries of the prophage elements as reported [31,32] are provided. Entries marked * are based on Mehta *et al* [18].

Organism	Proteins in e14 element	Related genes identified	Locus as reported [30,31]	Prophage name
B. subtilis *	b1152	Bsu1274	1316849-1347491	PBSX
	b1152, b1158	Bsu2593, Bsu2572	2652219-2700977	SKIN
B. melitensis **M**	b1151	BMET1349	1394344-1404607	Bruc1
B. suis	b1151	BR0586	578083-584877	Brs1
C. tetani E88	b1140, b1158	CTC01567, CTC01557	1663821-1696302	Cpt2
	b1149, b1151, b1152	CTC02132, CTC02131, CTC02115, CTC02134	2242455-2281387	Cpt3
E. coli K12*	b1156, b1158	b0561, b0544	564025-585326	DLP12
	b1156, b1157, b1158	b1546, b1547, b1545	1630450-1646830	QIN
	b1156, b1157, b1158	b1373, b1372, b1374	1409966-1433025	Rac
	b1154, b1156	b2353, b2355	2464404-2474619	KpLE1
E. coli	b1140, b1145	c1519, c1546	1397370-1452231	CP073-4
	b1140, b1145, b1155	c1400, c1410, c1475	1327053-1372820	CP073-2
	b1142, b1145, b1147, b1149, b1158	c3200, c3197, c3195, c3192, c3146	3019963-3065315	CP073-5
	b1155	c0969	909332-942273	CP073-1
	B1155	c0649	627155-630053	CP073-6
E. coli O157 VT-2 Sakai	b1140, b1141, b1155, b1156, b1157	ECs1609, ECs1610, ECs1651, ECs1650	1618153-1665049	Sp8
	b1140, b1140, b1141, b1149	ECs1757, ECs1813, ECs1758, ECs1792	1757506-1815680	Sp9
	b1140, b1149	ECs1501, ECs1542	1541470-1589892	Sp6
	b1140	ECs1055	1161091-1210740	Sp4
	b1140	ECs2773	2668007-2712035	Sp14
	b1141	ECs0801	891123-929708	Sp3

	b1145	ECs2990	2895926-2943804	Sp15
	b1145, b1154, b1155, b1155, b1156, b1156, b1157, b1157, b1158	ECs0274, ECs0280, ECs0282, ECs0281, ECs0281, ECs0282, ECs0280, ECs0283, ECs0284	300041-310626	Sp1
	b1145	ECs1185	1246012-1308719	Sp5
	b1145, b1149, b1149	ECs2279, ECs2276, ECs2251	2203952-2250093	Sp12
	b1145	ECs2209	2158174-2203951	Sp11
	b1149	ECs1598, ECs1592	1594570-1610032	Sp7
	b1149	ECs1971	1921414-1972525	Sp10
	b1152, b1153, b1158	ECs4987, ECs4988, ECs4992	5040843-5079601	Sp18
	b1158	ECs3240	3192983-3201533	Sp16
	b1158	ECs3516	3475965-3500163	Sp17
E. coli	b1149, b1151, b1140	z1359, z1362, z1323	1250521‒1295458	CP-933M
	b1149, b1151, b1140	z1803, z1806, z1764	1626722-1673485	CP-933N
	b1149, b1149, b1151, b1145	z6045, z6070, z6042, z6073	2285976-2329446	CP-933P
	b1145, b1154, b1155, b1157, b1158	z0309, z0314, z0315, z0317, z0318	300070-310251	CP-933H
	b1140, b1141, b1155, b1157	z1866, z1867, z1920, z1918	1702185‒1756006	CP-933X
	b1140, b1155	z2966, z2983	2668339-2688870	CP-933T
	b1149, b1151	z1854, z1849	1678706-1693737	CP-933C
	b1140	z3130	2743223-2788348	CP-933U
	b1140, b1145	z2036, z2090	1849488-1930250	CP-933O
	b1145, b1149	z3358, z3332	2966382-3015014	CP-933V
*H. influenzae*Rd KW20	b1152, b1153	HI1520, HI1521	1559962-1594275	FluMu
M. loti MAFF303099 *	b1149, b1151	Mlr8521, Mlr8522	6975633-7011594	Meso2
M. tuberculosis CDC1551	b1158	MT3573	3870821-3879383	Mt2

M. tuberculosis	b1158	Rv1586c	1780641-1788503	□Rv1
*N. meningitidisZ 2491 **	b1152, b1153, b1157	NMA1323, NMA1324, NMA1325	1207416-1236260	Pnm2
	b1152, b1153	NMA1826, NMA1825	1768546-1807515	Pnm1
N. meningitidis MC58	b1153, b1155, b1157	NMB1114, NMB1119, NMB1115	1099901-1133957	NeisMu1
S. aureus N315	b1149	SA1777	2005924-2049520	φN315
S. aureus MW2	b1149, b1152, b1159	MW1401, MW1392, MW1403	1529381-1573005	φSa2mw
	b1149	MW1908	2046605-2088749	φSa3mw
S. aureus	b1149, b1145	SAV1966, SAV1998	2083583-2126179	φMu50B
S. enterica (serovar Typhi Ty2)	b1155, b1156, b1157, b1158	t3435, t3434, t3434, t3435, t3433, t3437	3501128-3538076	Stt4
	b1155, b1156, b1157, b1158	t1349, t1349, t1351, t1346	1314607-1441766	Stt1
	b1155, b1156	t1867, t1867	1928058-1972330	Stt2
	b1158	t2667	2735202-2754628	Stt3
S. enterica CT18 (serovar Typhi)	b1140, b1141, b1143, b1144, b1155, b1156	STY2077, STY2076, STY2069, STY2068, STY2013, STY2013	1889471-1933558	Sti4b
	b1155, b1155, b1156, b1156, b1157, b1158	STY3693, STY3692, STY3692, STY3693, STY3691, STY3695	3515470-3548975	Sti8
	b1155, b1156, b1157, b1158	STY1639, STY1639, STY1638, STY1637, STY1640, STY1641, STY1642, STY1643	1538899-1572919	Sti3

	b1155, b1156, b1158	STY1073, STY1073, STY1075	1008747-1052755	Sti1
	b1158	STY2889	2760475-2768771	Sti7
S. enterica LT2 (serovar Typhimurium)	b1145, b1156, b1157	STM0898, STM0927, STM0926	962612-1006517	Fels-1
	b1154, b1155	STM2235, STM2233	2330961-2345217	Stm6
	b1155, b1158	STM2704, STM2705, STM2702	2844427-2879233	Fels-2
	b1156, b1157	STM2586, STM2588	2728976-2776816	Gifsy-1
	b1156, b1157	STM1050, STM1049	1098228-1143714	Gifsy-2
	b1140, b1144	S0941, S0921	897790-930670	T5
	b1140, b1155, b1156	S2146, S2118, S2118	2021895-2044342	T11
	b1149, b1151	S1228, S1223	1177837-1191596	T7
	b1154	S0319	313843-327223	T2
	b1155, b1156	S2329, S2329	2208707-2214367	T12
	b1158	S2781	692891-709118	T3
S. flexneri 2a301	b1140, b1159	SF2044, SF2041	2049694-2066397	Flex9
	b1149, b1151	SF1146, SF1140	1175319-1188408	Flex5
	b1154, b1155	SF0311, SF0310	311291-328079	Flex2
S. pyogenes	b1140	SPy1488	1192854-1222549	370.2
	b1157, b1158	SPy0671, SPy0655	527569-571887	370.1
S. pyogenes	b1145	SpyM18_1306	1041280-1087739	ϕspeL/M
	b1145	SpyM18_1504	1206360-1241416	ϕ370.3-like
	b1149, b1158	SpyM18_0751, SpyM18_0716	578093-618765	ϕspeC
	b1149	SpyM18_0369	293882-332714	ϕspeA
S.	b1145	SpyM3_1143	1137743-1171867	ϕ315.3
	b1149	SpyM3_0946	977738-1018193	ϕ315.2
	b1149	SpyM3_0710	749213-788176	ϕ315.1
X. fastidiosa 9a5c*	b1140, b1149	XF1642, XF1645	1585980-1631056	XfP4
Y. pestis KIM	b1145, b1152, b1153, b1154, b1157	Y2954, Y2937, Y2935, Y2936, Y2935, Y2934	3237524-3255252	Yers3
	b1155, b1156	Y2185, Y2185	2417129-2456467	Yers1

Y. pestis CO92	b1145, b1152, b1153, b1154, b1157	YP01233, YP01250, YP01251, YP01250a, YP01252, YP01252	1392489-1416524	YP3
	b1155, b1156	YP02134, YP02134	2364324-2413098	YP5

Table 2: Putative prophage elements newly identified in six organisms. Prophage elements that were newly identified in the selected genomes using similarity to e14 proteins as a criterion. BLAST hits for the e14 proteins in different organisms were examined, and only the significant hits (e < = 0.01) are listed. Estimates of the prophage region are provided with the outer limits given in parentheses.

Organism	Proteins ine14 element	Related genes	Prophage region (outer limit)	Location (outer limit)	Prophage name
S. enterica LT2 (serovar Typhimurium)	b1140, b1158, b1156, b1140	STM1861, STM1865, STM1868, STM1871	STM1861– STM1871 (STM1860– STM1882)	1957835- 1967922 (1956854- 1975533)	St1
S. flexneri **T**	b1158, b1140	S2707, S2723	S2707 — S2723 (S2705 — S2723)	2602155- 2613694 (2600230- 2613694)	Sf1
S. pyogenes M18 MGAS8232	b1146, b1151, b1159	SpyM18_0 636, SpyM18_0 620, SpyM18_0 615	spyM18_0615 - spyM18_0636 (spyM18_060 9- spyM18_0640)	495793-506387 (492411- 511356)	Sp1
S. pyogenes M3 MGAS315	b1146, b1159	SpyM3_03 99, SpyM3_03 92	SpyM3_0392 - SpyM3_0399 (SpyM3_0386 - SpyM3_0403)	434301-439876 (430946- 444845)	Spy1
V. cholerae N16961	b1159, b1159	VCA0307, VCA0309	VCA0307- VCA0309 (VCA0281- VCA0324)	319123-322036 (300467- 328558)	Vc1

*P. luminescens subsp. laumondii*TTO 1	b1155, b1156, b1157, b1157, b1158, b1155, b1156, b1157	plu0018, plu0019, plu0021, plu0020, plu0029, plu0033, plu0033, plu0034	plu0018-plu0034 (plu0008-plu0034)	17678-29999 (10251-29999)	Pl1
	b1140, b1155, b1155, b1155, b1155, b1156, b1156, b1157	plu2947, plu2956, plu2958, plu2961, plu2873, plu2873, plu2961, plu2784	plu2873-plu2961 (plu2870-plu2961)	3409531-3466582 (3405940-3466582)	Pl2
	b1158, b1155, b1156, b1157	plu3296, plu3332, plu3327, plu3326	plu3s296-plu3332 (plu3296-plu3338)	3912405-3962420 (3912405-3966510)	Pl3
	b1155, b1155, b1155, b1156, b1156, b1157	plu3023, plu3012, plu3024, plu3024, plu3012, plu3013	plu3012-plu3024	3512834-3525239	Pl4
	b1146, b1149, b1156, b1157	plu3476, plu3473, plu3497, plu3421, plu3498	plu3421-plu3498	4037155-4092707	Pl5
	b1157, b1157	plu1460, plu1463	plu1460-plu1463	1753139-1756507	Pl6
	b1155, b1156, b1156, b1157, b1157, b1157	plu2035, plu2035, plu2023, plu2024, plu2022, plu2034	plu2022-plu2035	2390996-2405104	Pl7

Table 3: Prophages found near a dehydrogenase gene.

Organism	Prophage region (outer limit)	Location (outer limit)	Dehydrogenase gene
S. flexneri **T**	S2707 — S2723 (S2705— S2723)	2602155-2613694 (2600230-2613694)	S2726 : IMP dehydrogenase
S. enterica LT2 (serovar Typhimurium)	STM1861— STM1871 (STM1860— STM1882)	1957835-1967922 (1956854-1975533)	STM1886: glucose-6-phosphate dehydrogenase
S. pyogenes M18 MGAS8232	spyM18_0615 - spyM18_0636 (spyM18_0609 - spyM18_0640)	495793-506387 (492411-511356)	spyM18_0608: Putative nucleotide sugar dehydrogenase
*S. pyogenes*M3 MGAS315	SpyM3_0392 - SpyM3_0399 (SpyM3_0386 - SpyM3_0403)	434301-439876 (430946-444845)	SpyM3_0385:Putative nucleotide sugar dehydrogenase
*P. luminescens subsp. laumondii*TTO1	plu0018— plu0034 (plu0008— plu0034)	17678-29999 (10251-29999)	plu0007:Aspartate semialdehyde dehydrogenase

References

1. Boyd EF and Brussow H: **Common themes among bacteriophage-encoded virulence factors and diversity among the Bacteriophages involved.** *Trends Microbiol* 2002, **10:** 521-529

2. Wagner PL and Waldor MK: **Bacteriophage control of bacterial virulence.** *Infect. Immun* 2002, **70:** 3985-3993

3. Waldor MK: **Bacteriophage biology and bacterial virulence.** *Trends Microbiol* 1998, **6:** 295-297

4. Brussow H and Hendrix R: **Phage genomics: Small is beautiful.** *Cell* 2002, **108:** 13-16

5. Bushman F: Lateral DNA transfer: mechanisms and consequences. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor 2002

6. Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H: **Prophage genomics**. *Microbiol Mol Biol Rev*. 2003, **67 :** 238-276

7. Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K, *et al* : **Genome and virulence determinants of high virulence community-acquired MRSA.** *Lancet* 2002, **359:** 1819-1827

8. Smoot JC, Barbian KD, Van Gompel JJ, Smoot LM, Chaussee MS, *et al* : **Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks.** *Proc. Natl. Acad. Sci.* USA 2002, **99:** 4668-4673

9. Glaser P, Frangeul L, Buchreiser C, Rusniok C, Amend A, *et al* : **Comparative**

genomics of *Listeria* species. *Science* 2001, **294:** 849-852

10. Perna NT, Plunkett G, III, Burland V, Mau B, Glasner JD, *et al* : **Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.** *Nature* 2001, **409:** 529-533

11. Smoot LM, Smoot JC, Graham MR, Somerville GA, Sturdevant DE, *et al* : **Global differential gene expression in response to growth temperature alteration in group A *Streptococcus.*** *Proc. Natl. Acad. Sci.* USA 2001, **98:** 10416-10421

12. Whiteley M, Bangera MG, Bumgarner RE, Parsek, MR, Teitzel GM, Lory S and Greenberg EP: **Gene expression in *Pseudomonas aeruginosa* biofilms**. *Nature* 2001, **413:** 860-864

13. Davis BM and Waldor MK: **Filamentous phages linked to virulence of *Vibrio cholerae.*** *Curr Opin Microbiol.* 2000, **36:** 35-42

14. Stromsten NJ, Benson SD, Burnett RM, Bamford DH, Bamford JK: **The *Bacillus thuringiensis* linear double-stranded DNA phage Bam35, which is highly similar to the *Bacillus cereus* linear plasmid pBClin15, has a prophage state.** *J Bacteriol.* 2003, **185:** 6985-6989

15. Tang S, Nuttall S, Ngui K, Fisher C, Lopez P, Dyall-Smith M: **HF2: a double stranded DNA tailed haloarcheal virus with a mosaic genome.** *Mol Microbiol.* 2002, **44:** 283-296

16. Bergsland KJ, Kao C, Yu YN, Gulati R, Snyder L : **A site in the T4 bacteriophage major head protein gene that can promote the inhibition of all translation in *Escherichia coli*.** *J Mol Biol.* 1990, **213:** 477-494

17. Blaisdell BE, Campbell AM and Karlin S: **Similarities and dissimilarities of phage genomes.** *Proc. Natl. Acad. Sci.* USA 1996, **93:** 5854-5859

18. Mehta P, Casjens S and Krishnaswamy S: **Analysis of the lambdoid prophage element e14 in the *E.coli* K12 genome.** *BMC Microbiol.* 2004**, 4:**1

19. Blattner, F.R. *et al.* : T**he complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277:** 1453-1474

20. Brody H and Hill CW : **Attachment site of the genetic element e14**. *J Bacteriol.* 1988, **170:** 2040-2044

21. Hill CW, Gray JA and Brody H : **Use of the Isocitrate Dehydrogenase Structural Gene for attachment of e14 in *Escherichia coli* K-12**. *J Bacteriol.*, 1989, **171:** 4083-4084.

22. van de Putte P, Plasterk R and Kuijpers A : **A Mu gin complementing function and an invertible DNA region in the *Escherichia coli* K-12 are situated on the genetic element e14.** *J Bacteriol.* 1984, **158:** 517-522

23. Kao C, Gumbs E and Snyder L: **Cloning and characterization of the *Escherichia coli lit* gene, which blocks bacteriophage T4 late gene expression**. *J Bacteriol.* 1987*,* **169:** 1232-1238

24. Kao C and Snyder L : **The *lit* gene product which blocks bacteriophage T4 late gene expression is a membrane protein encoded by a cryptic DNA element, e14**. *J Bacteriol* 1988, **170:** 2056-2062

25. Ravi RS, Sozhamannan S and Dharmalingam K : **Transposon mutagenesis and genetic mapping of the *rglA* and *rglB* loci of *Escherichia coli.*** *Mol Gen Genet.* 1985, **198:** 390-392

26. Kelleher JE and Raleigh EA : **Response to UV damage by four *Escherichia coli* K-12 Restriction Systems.** *J Bacteriol.* 1994, **176:** 5888-5896

27. Enomoto M Oosawa K and Momata H : **Mapping of the pin locus coding for a site-specific recombinase that causes flagellar-phase variation in** *Escherichia coli* **K-12.** *J Bacteriol.* 1983, **156:** 663-668

28. Plasterk RH and van de Putte P : **The invertible P-DNA segment in the chromosome of Escherichia** *coli.* *The EMBO Journal* 1985, **:** 237-242

29. Maguin E, Brody H, Hill CW and D'Ari R : **SOS-associated division inhibition gene sfiC is part of excisable element e14 in** *Escherichia coli.* . 1986, **168:** 464-466

30. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV : **The COG database: new developments in phylogenetic classification of proteins from complete genome**. *Nucl Acids Res.* 2001, **29:** 22-28

31. Prophage Database http://203.90.127.174:8082/prophagedb

32. Casjens S: **Prophages and bacterial genomics: what have we learned so far?** *Mol. Microbiol* 2003, **49:** 277-300

33. Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ: **Basic Local Alignment Search Tool**. *J. Mol. Biol.* 1990, **215:** 403-410

34. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ : **Gapped BLAST and PSI-BLAST: A new generation of protein database search programs.** *Nucl Acids Res.* 1997, **25:** 3389-3402