HAL
open science

The use of genomic signature distance between bacteriophages and their hosts displays evolutionary relationships and phage growth cycle determination.

Patrick Deschavanne, Michael Dubow, Christophe Regeard

To cite this version:

Patrick Deschavanne, Michael Dubow, Christophe Regeard. The use of genomic signature distance between bacteriophages and their hosts displays evolutionary relationships and phage growth cycle determination.. Virology Journal, 2010, 7 (1), pp.163. 10.1186/1743-422X-7-163 . inserm-00668213

HAL Id: inserm-00668213

https://inserm.hal.science/inserm-00668213

Submitted on 9 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deschavanne et al. Virology Journal 2010, 7:163
http://www.virologyj.com/content/7/1/163

VIROLOGY JOURNAL

RESEARCH Open Access

The use of genomic signature distance between bacteriophages and their hosts displays evolutionary relationships and phage growth cycle determination

Patrick Deschavanne[1*], Michael S DuBow[2], Christophe Regeard[2]

Abstract

Background: Bacteriophage classification is mainly based on morphological traits and genome characteristics combined with host information and in some cases on phage growth lifestyle. A lack of molecular tools can impede more precise studies on phylogenetic relationships or even a taxonomic classification. The use of methods to analyze genome sequences without the requirement for homology has allowed advances in classification.

Results: Here, we proposed to use genome sequence signature to characterize bacteriophages and to compare them to their host genome signature in order to obtain host-phage relationships and information on their lifestyle. We analyze the host-phage relationships in the four most representative groups of Caudoviridae, the dsDNA group of phages. We demonstrate that the use of phage genomic signature and its comparison with that of the host allows a grouping of phages and is also able to predict the host-phage relationships (lytic *vs*. temperate).

Conclusions: We can thus condense, in relatively simple figures, this phage information dispersed over many publications.

Background

Bacteriophages are the most abundant biological entities on Earth and their total population is estimated at approximately 10^{31} particles on earth [1]. In comparison with the estimated 10^{30} bacterial cells in the biosphere [2], there are thus 10 virus particles for each putative host [3,4]. In aquatic or terrestrial samples, 10^6 to 10^7 viral particles per milliliter of water or gram of soil are regularly reported. Moreover, these viruses are highly dynamic, leading to approximately 10^{23} infections per second [5]. The study of phage diversity is crucial for understanding an ecosystem. For instance, the concept of "killing the winner" has been proposed to explain how phage propagation can control host diversity and abundance [6].

Bacteriophages also participate in the evolution of their bacterial hosts. Horizontal transfer of genes from phage to host and vice versa has been well documented [7,8]. Temperate bacteriophages have the capacity to integrate their DNA into that of the host and can also lead to lysogenic conversion in pathogenic bacteria such as *Vibrio cholerae* [9]. Prophages have been shown to contribute to genome diversification [10] and in some environments, the majority of bacteria contains at least one prophage [4,11]. Lawrence et al, (2002) [12] calculated an average of 2.6 prophages per free living bacterial cell, some genomes can contain up to 10% of prophage DNA [13].

Since the sequencing of the first complete genome of bacteriophage Φ × 174 in 1977 [14], several characteristics have been established concerning phage genomes. The size of completely sequenced genomes varies between 2435 bp (Leuconostoc phage L5) and 497 513 bp (*Bacillus* phage G). However the size distribution of phage genomes is not homogenous, possibly because of a bias linked to isolation techniques [15]. Phages genomes ranging in size from 30 to 60 kb, the majority belonging to the Siphoviridae, have been the most

* Correspondence: patrick.deschavanne@univ-paris-diderot.fr
[1]Molécules thérapeutiques *in silico* (MTI), INSERM UMR-M 973, Université Paris Diderot - Paris 7, Bât Lamarck, 35 rue Hélène Brion, 75205, Paris Cedex 13, France

BioMed Central
© 2010 Deschavanne et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

sequenced (approximately 55% of the total), and small phage genomes (5 to 20 kb) are the second most abundant size range (approximately 27% of the total). An intriguing gap is observed between 80 and 100 kb, with very few complete genome sequences available followed by large genome sequences. The distribution of morphotypes corresponding to the genomes available in the Genbank phage database reflects what has been observed by electron microscopy [16], with a predominance of tailed phages containing double strand genomic DNA. The extraordinary diversity of phages in nature, the dynamism of phage populations and the lack of homology among most phage genes is a recurrent theme. However it is also common to observe an absence of homology among phage genes belonging to phages infecting the same host and therefore likely closely related. As the number of available sequenced phage genomes increases, their mosaic structure is becoming more evident [17-19]. Phage genome mutation rates, combined with recombination leading to genetic mosaicism as well as the lack of an universal gene, analogous to the 16 S rRNA gene, explain why phage classification is based on the nature of the phage nucleic-acid and virion morphology. Family-specific genes such as viral capsid structural genes have been used as taxonomic tools [20]. However, these methods are limited and do not reveal other phage characteristics such as virus-host relationships. Homology-free methods based on the usage of oligonucleotides (sequence signatures) are potentially interesting to try in phage classification. Numerous studies have shown the utility of genomic signatures for different purposes. Dinucleotide frequencies have been used to compare genomic signatures of prokaryotic genomes [21-25] or phage genomes [26]. Methods based on longer oligonucleotides were further developed for the characterization and classification of bacterial species [27]. Local variations of the genomic signature along the sequence of a genome allow the detection of horizontal transfers and pathogenicity islands [28-33] or prophages remnants [34]. More recently genomic signatures were used in an approach to classify virus genomes, and it was observed that, in general, viral genome signatures are close to that of their hosts [35]. Another use of genome sequence signatures, applied to viruses, was to assign environmental genomic fragments either to a known species or to regroup them in new ones [36].

Recent phages metagenomic studies [11,37,38] have reinforced the view that phage genome diversity is extraordinarily high, that phages with dsDNA are predominant in the environment, and that they constitute an enormous "source" of uncharacterized genes. One of the principal questions that remains to be answered is the nature of phage-host relationships in the context of genomic and metagenomic data, such as the phage life cycle (lytic or temperate), morphotype or host range.

In this report, we have used the genomic signatures of phages and their hosts to aid in the understanding of these relationships. Host signatures from four bacterial species infected by a large number of phages has been compared with phage signature. We calculated a "distance" between each phage and its host. We demonstrate that this distance can be used to group the phages and gives indications of the phages growth cycle.

Results and discussions

Choice of the phage genomes used in this study

As of January 2009, there were 521 bacterial and archaeal virus genomes available in the Genbank phage database. Among these genomes, 459 are composed of dsDNA and are mainly distributed among the Orders of the Caudoviridae. The 62 remaining genomes contain ssDNA or RNA and correspond to Microviridae, Leviviridae and Inoviridae members.

We examined the 459 dsDNA phage genomes of the database and, where possible we grouped the different phage genomes by host and collected data concerning their morphotype, whether temperate or lytic, and the genome length.

The Caudoviridae corresponded to 84% of available genomes, composed of 57% Siphoviridae, 23% Myoviridae and 20% Podoviridae families (Figure 1A). This distribution is nearly the same as that published in 2007 concerning the phages examined in the electron microscope [16], although 9% of the available genomes have not been characterized or completely annotated. Approximately one third of the genomes contain an indication of the capacity to lysogenize their hosts. Only 21% have been described as exclusively lytic, whereas for 43% of the phages this information is not mentioned (Figure 1B). The majority (60%) of the genomes available in the database infect only 13 species, with a clear dominance of phages infecting *Mycobacterium smegmatis, Staphylococcus aureus, Pseudomonas aeruginosa and Escherichia coli* (Figure 1C). We thus examined the Caudoviridae members infecting these four bacterial species.

Escherichia coli Caudoviridae

Forty-six genomes of the order Caudoviridae infecting *E. coli* can be gathered in Genbank phage database. The genomic signature of each phage was generated, as detailed in Methods, compared with the genomic signature of *E. coli* W3110, and the distance between phages and host was calculated. Other *E. coli* strains were tested but the distances were not significantly different (data not shown). The genomic signature distances, morphotypes, genome lengths and life styles are shown in Figure 2.

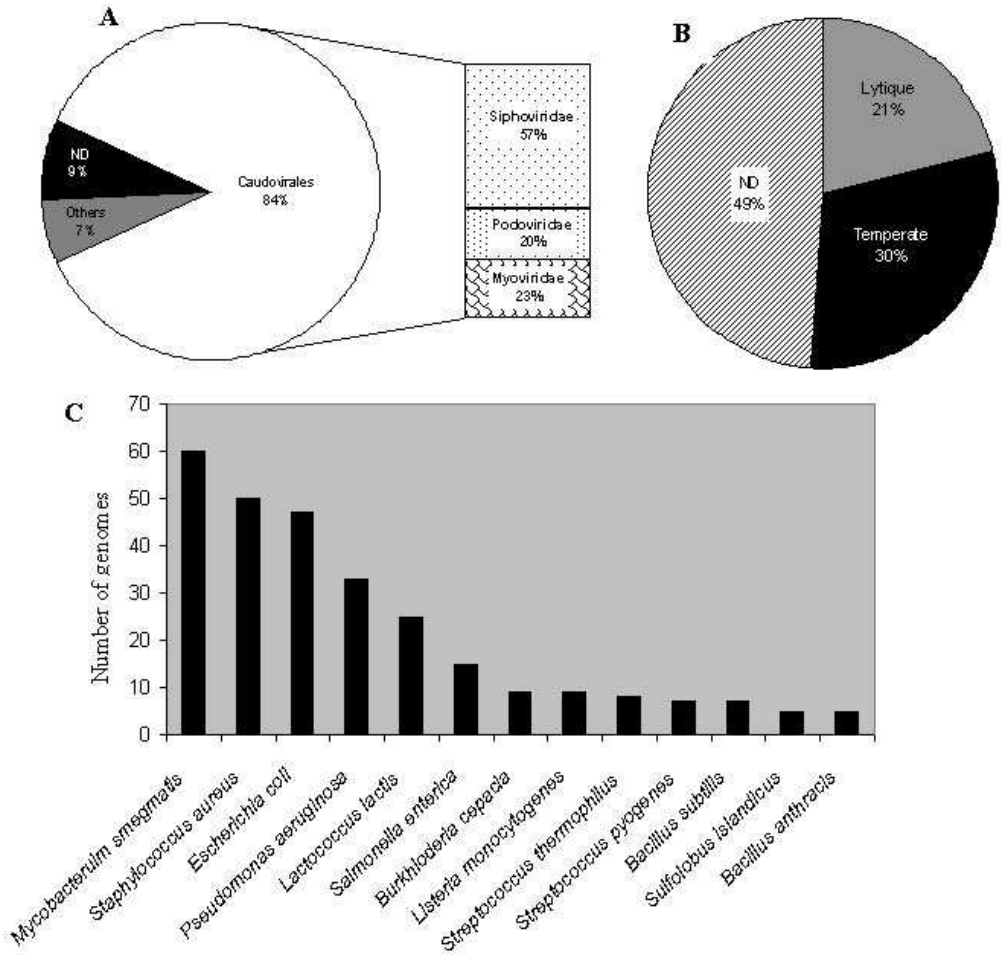
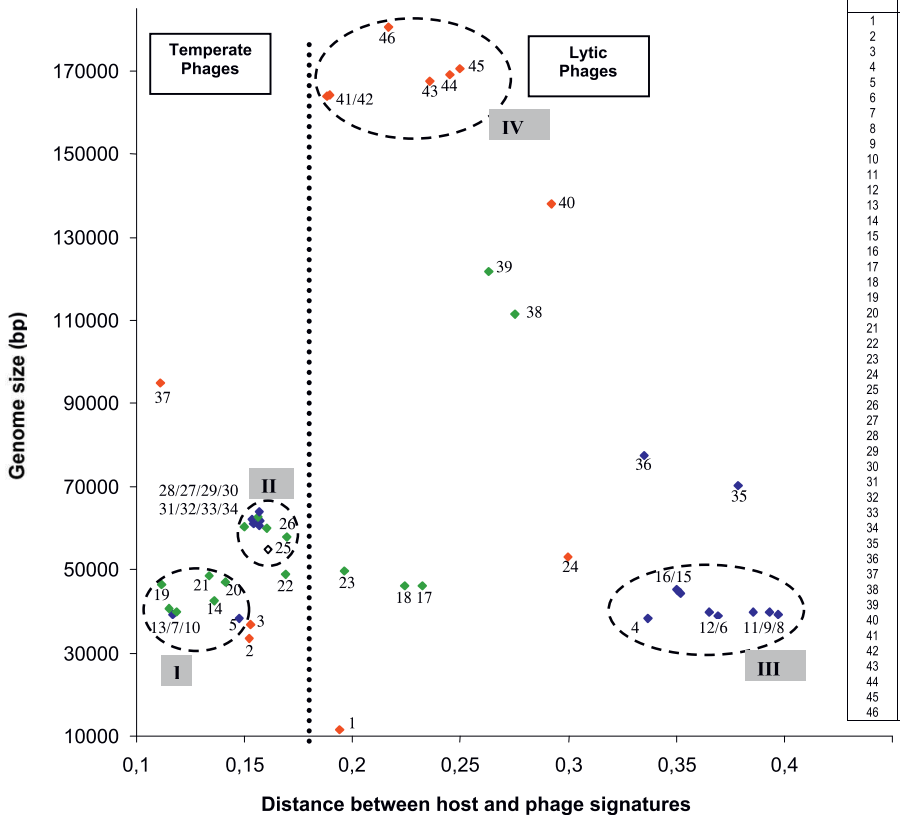


Figure 1 Distribution of completely sequenced bacteriophage genomes retrieved from Genbank-phage Database. A: Proportion of genomes belonging to the different phage families. B: Proportion of genomes from temperate or lytic phages. C: Number of completely sequenced genome of phages infecting the same host. Only host with at least 5 different phages are shown. ND: Not indicated in the database.

E. coli phage groups

A first feature that is revealed by an analysis of Figure 2 is the coherent grouping of phages. This grouping is in agreement with a 6 groups K-means classification based on phage signatures. The number of groups is greater than those described in Figure 2 to take into account the isolated phages. The groups based on signature distance correspond to the different known and identified groups of coliphages. For example, all the phages belonging to the lytic T7 super-group (group III) have a relatively homogenous distance signature. For temperate phages, two groups can be observed. The first group (group I), containing the lambda-like phages, is characterized by a short distance signature, perhaps reflecting a more ancient prophage life style. The second lambdoid group (II) is very homogenous and contains phages characterized by their ability to carry shiga toxin-like encoding genes. Our representation appears to be compatible with the "classification scheme" suggested by Casjens [39]. The last group (IV) corresponds to the T4 super-group that contains phages with genomes ranging from 164 to 180 kb in length. These genomes have the peculiarity of having a low GC%, necessitating the normalization of genomic signature of the host and phages (see Methods). In spite of the fact that genomes are larger and then likely least host dependent, the overall observed distance is less than that of the T7 super-group. In the *E. coli* phage landscape, several phages remain isolated. Phage ΦEcoM-Gj1 has been recently described and its genome reveals a unique pattern of different origins. It is the first phage with a Myoviridae morphotype but with a T7-like RNA polymerase and a large subunit terminase related to that of phage T1 [40]. Phage EPS7 has been isolated and its genome recently analyzed [41]. This phage belongs to the T5 family and its close genomic signature distance is not surprising. The addition to this group of the phage rv5 is tempting, although rv5 is a Myoviridae. Moreover, the proximity



Number	name	Genome size	Accession	References
1	P4	11624	NC_001609	[72]
2	P2	33593	NC_001895	Not published
3	Mu	36717	NC_000929	[44]
4	T3	38208	NC_003298	[73]
5	HK620	38297	NC_002730	[74]
6	13a	38841	NC_011045	Not published
7	phiV10	39104	NC_007804	Not published
8	EcoDS1	39252	NC_011042	Not published
9	K1F	39704	NC_007456	[75]
10	HK97	39732	NC_002167	[76]
11	BA14	39816	NC_011040	[77]
12	T7	39937	NC_001604	[78]
13	HK022	40751	NC_002166	[76]
14	phiP27	42575	NC_003356	[79]
15	K1-5	44385	NC_008152	[80]
16	K1E	45251	NC_007637	[81]
17	JK06	46072	NC_007291	Not published
18	RTP	46219	NC_007603	[82]
19	N15	46375	NC_001901	Not published
20	cdtI	47021	NC_009514	Not published
21	lambda	48502	NC_001416	[83]
22	T1	48836	NC_005833	[49]
23	TLS	49902	NC_009540	[84]
24	phiEcoM-GJ1	52975	NC_010106	[40]
25	YYZ-2008	54896	NC_011356	Not published
26	BP-4795	57930	NC_004813	[85]
27	Stx1 converting phage	59866	NC_004913	[86]
28	Stx2-converting phage 86	60238	NC_008464	Not published
29	VT2-Sakai	60942	NC_000902	[87]
30	933W	61670	NC_000924	[88]
31	Stx2 converting phage I	61765	NC_003525	[89]
32	Stx2-converting phage 1717	62147	NC_011357	Not published
33	Stx2 converting phage II	62706	NC_004914	[86]
34	Min27	63395	NC_010237	Not published
35	N4	70153	NC_008720	[90]
36	phiEco32	77554	NC_010324	[43]
37	P1	94800	NC_005856	[91]
38	EPS7	111382	NC_010583	[41]
39	T5	121750	NC_005859	Not published
40	rv5	137947	NC_011041	Not published
41	RB49	164018	NC_005066	[92]
42	Phi1	164270	NC_009821	Not published
43	RB69	167560	NC_004928	Not published
44	T4	168903	NC_000866	[93]
45	JS98	170523	NC_010105	[94]
46	RB43	180500	NC_007023	Not published

Figure 2 Distribution of the genomic signature distances of *E. coli* phages as a function of size of phage genomes [72-94]. Red symbol: Myoviridae, green symbol: Siphoviridae, blue symbol Podoviridae, white symbol: family not indicated. The numbers correspond to the phages listed in the Table.

of the T4 super group and the putative T5 group is coherent. Analysis of the T5 sequence by Wang et al (2005) [42] revealed that in the "top 10" homologous phages and genes, RB49, RB69 and T4 are first on the list. Like ΦEcoM-Gj1, phage ΦEco32 has been described as a genome with a large degree of mosaicism [43]. The genomic signature distance of phage N4 seems to allow it to be grouped with ΦEco32, but no genetic relationship can be retrieved from the literature. Finally, phages Mu and P2 show very close distances, whereas Mu is able to integrate as a prophage by a transposition mechanism, while P2 has a site-specific mechanism of genome integration. It is noticeable that significant homology between phage Mu and P2 have been observed for the tail fiber encoding genes [44]. Phage P4, the satellite phage of P2, is a defective phage that exists as a plasmid, shows a more divergent distance signature. Figure 2 confirms that there is no correlation between morphotypes and groups or subtype of phages, although several groups appear to be more homogenous than others. For example, the temperate phage group represented by phage 933W (II) appears more susceptible to exchange modules encoding tail fibers. There is also no significant correlation between genome length and the distance between the host and phage signature. However, our representation, using a combination of the distance signatures, genome length and phage characteristics (life style and morphotype), allows us, independently of sequences comparison, to obtain a coherent picture of the "relationship landscape" of the bacteriophages of *E. coli*.

E. coli phage life styles

The second striking observation is the apparent separation between temperate and lytic phages. All the temperate phages are characterized by a host-phage distance ≤0.2. The genomic signature distance seems therefore be sufficiently robust, without any direct sequence comparison, to distinguish these two different life styles. The short genomic distance for temperate bacteriophages is likely due to the long timescale of the "prophage" state. This hypothesis was first suggested by Lawrence and Ochman (1997) [45] to explain that horizontally acquired genes will, over time, adopt the molecular characteristics of the host genome, and has been

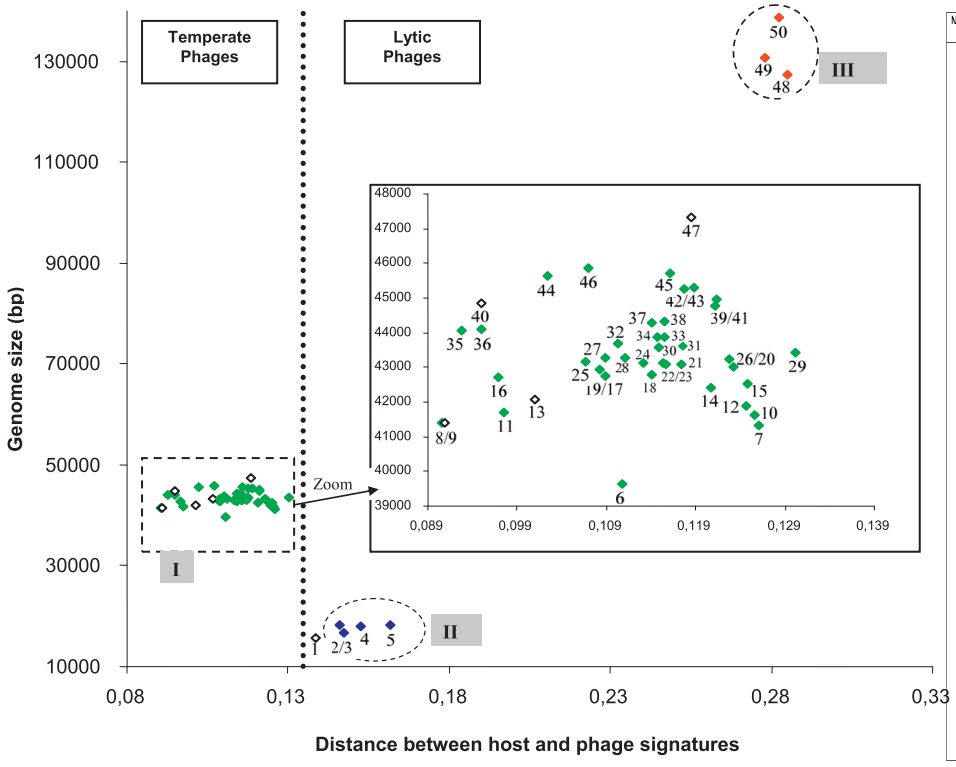
recently confirmed in a study comparing the sequenced genomes of different strains of the same species [46]. Thus, for temperate phages, the more time a genome remains in a prophage state the smaller should be the genomic signature distance. The difference between temperate and lytic phages of *E. coli* is intriguing because it should not be difficult for a temperate phage to lose its ability to lysogenize its host [39]. The high rate of horizontal transfers in phage genomes is also an argument for the possible acquisition of a functional module involved in lysogeny. The use of genomic signature distances may allow the detection of a temperate phage that has recently lost its lysogenic capacity. In *E. coli*, such examples have not yet been identified, whereas several examples in other species have been reported [47,48]. A lytic phage for which the distance resembles temperate distances is represented by phage T1. In the genome of T1, a homolog of the phage N15 *cor* gene, involved in lysogenic conversion, can be found. When phylogenetic trees are constructed, several lines of descent, including temperate phages such as N15, HK022 and HK97 have been suggested [49]. The largest temperate phage genome P1 shares with N15 the shortest distance. However, the only thing in common between these two phages is a plasmid prophage form, suggesting that the homogenization process between phage and host genomic signatures may be more efficient for plasmids.

Staphylococcus aureus Caudoviridae

Fifty phages of the order Caudoviridae with completely sequenced genomes and infecting *Staphylococcus aureus* were analyzed using the same procedures as described for the *E. coli* bacteriophages.

S. aureus phage groups

Only 8 phages outside of the 39-47 kb genome length range and an average of distances of 0.12 were observed (Figure 3). *S. aureus* strains are often involved in pathogenesis, and represent an important cause of nosocomial infections. Thus, temperate phages with the capacity of lysogenic conversion, such as those containing Panton-Valentine Leukocidin toxins [50-53] are frequently examined. The genomic comparison of 27 phages reported by Kwan et al (2005) [19], based on genome size, nucleotide sequence and proteome comparisons, leads to the description of three separate groups. These 3 groups are retrieved by a K-means classification based on phage signatures and are also clearly evidenced using



Number	name	Genome size	Accession	References
1	PT1028	15603	NC_007045	[95]
2	44AHJD	16784	NC_004678	[54]
3	SAP-2	17938	NC_009875	Not published
4	66	18199	NC_007046	[95]
5	P68	18227	NC_004679	[54]
6	187	39620	NC_007047	[95]
7	2638A	41318	NC_007051	[95]
8	PVL	41401	NC_002321	[96]
9	tp310-1	41407	NC_009761	Not published
10	52A	41620	NC_007062	[95]
11	77	41708	NC_005356	[95]
12	55	41902	NC_007060	[95]
13	tp310-3	41966	NC_009763	Not published
14	92	42431	NC_007064	[95]
15	phiSauS IPLA88	42526	NC_011614	Not published
16	phi13	42722	NC_004617	[97]
17	69	42732	NC_007048	[95]
18	29	42802	NC_007061	[95]
19	phiSLT	42942	NC_002661	[53]
20	phiMR11	43011	NC_010147	[98]
21	phiETA	43081	NC_003288	[99]
22	3A	43095	NC_007053	[95]
23	71	43114	NC_007059	[95]
24	phiNM	43128	NC_008583	[52]
25	ROSA	43155	NC_007058	[95]
26	80	43231	NC_007063	[95]
27	phiETA2	43265	NC_008798	Not published
28	phiETA3	43282	NC_008799	Not published
29	X2	43440	NC_007065	[95]
30	96	43576	NC_007057	[95]
31	11	43604	NC_004615	[97]
32	37	43681	NC_007055	[95]
33	80alpha	43864	NC_009526	[100]
34	53	43883	NC_007049	[95]
35	phiNM3	44061	NC_009617	[53]
36	phiN315	44082	NC_004740	[101]
37	85	44283	NC_007050	[95]
38	phiMR25	44342	NC_010808	Not published
39	47	44777	NC_007054	[95]
40	phiPVL108	44857	NC_008689	[50]
41	phi 12	44970	NC_004616	[97]
42	phiSauS-IPLA35	45244	NC_011612	Not published
43	EW	45286	NC_007056	[95]
44	phiPV83	45636	NC_002486	[55]
45	tp310-2	45710	NC_009762	Not published
46	42E	45861	NC_007052	[95]
47	phi2958PVL	47342	NC_011344	[51]
48	K	127395	NC_005880	[102]
49	Twort	130706	NC_007021	[95]
50	G1	138715	NC_007066	[95]

Figure 3 Distribution of the genomic signature distances of *S. aureus* phages as a function of size of phage genomes[95-102].Red symbol: Myoviridae, green symbol: Siphoviridae, blue symbol Podoviridae, white symbol: family not indicated. The numbers correspond to the phages listed in the Table.

genomic distances signatures. Group II is composed of lytic Podoviridae with genome sizes inferior to 20 Kb and genomic signature distances around 0.15. Phages 44AHJD and P68 have been classified as Φ29-like by the ICTV, and the presence of a terminal protein at the genome extremities has been confirmed [54]. Phage PT1028 could be assigned to the same group because of its genome size, but no significant homologies can be observed with phages 44AHJD, P68 and 66 [19]. Our results allows us to add phage SAP-2 to group I. Phages K, Twort and G1 (group III) have genomes of approximately 130 Kb, belong to the Myoviridae family, are lytic phages, and have a clearly different signature distance (≅ 0.3) in comparison with the other *S. aureus* phages. The remaining 42 phages were classified in the same group (group I) and contain all the phages of class II, as defined by Kwan et al, 2005 [19]. The highest distance value is observed for phage X2 (0.13) and the smallest value was observed for phage PVL (0.09).

S. aureus phage life styles

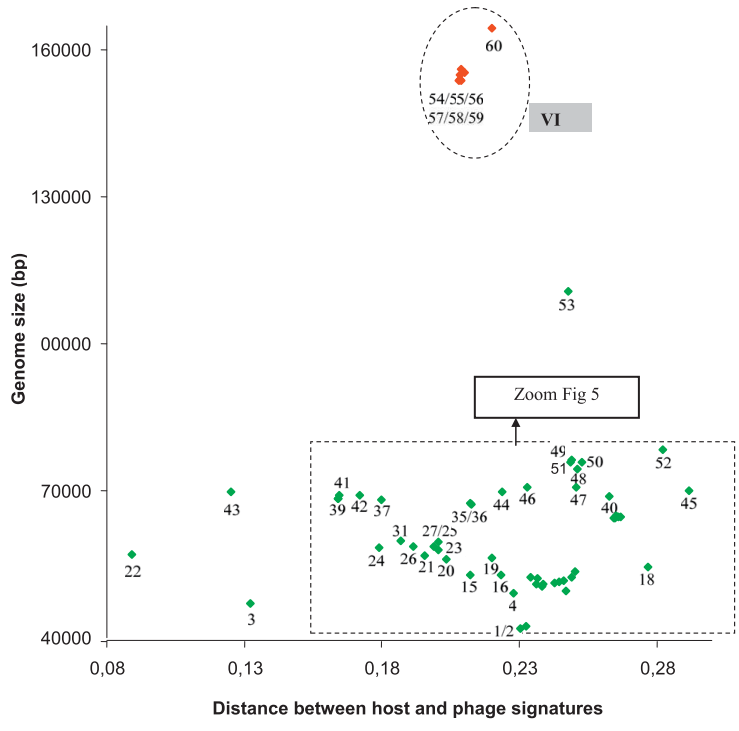
When information concerning morphotype and life style is available, group I phages belong to the Siphoviridae family and are temperate. It is interesting to note that, as for *E. coli* phages, the temperate phage genomes of *S. aureus* display a tendency to have undergone an amelioration process. The phages that show the smallest distances, PVL, PVL108 and phiPV83 have mutations or insertions that prevent their induction by Mitomycin C [50,51,55]. Phages SLT and 2958PVL possess significant homologies and genome organization with the three "inactive" phages cited above, but their genomic signatures have less resemblance to the host signature.

Mycobacterium smegmatis Caudoviridae

Sixty completely sequenced genomes of bacteriophages infecting *M. smegmatis* are available in the Genbank phage database. The overall landscape of the Mycobacteriophages obtained with the genomic signature distance (Figure 4) represents the high degree of genetic diversity described using sequence homologies and genome organization methods [18,56]. The distances vary between 0.008 (Che9c) and 0.29 (Predator), with an average (0.22) comparable to that observed in *E. coli* phages.

M. smegmatis phage groups

Six clusters have been described on the basis of nucleotide similarity [18]. A k-means classification is difficult to



Number	name	Genome size	Accession	References
1	BPs	41901	NC_010762	not published
2	Halo	42289	NC_008202	[103]
3	Brujita	47057	NC_011291	not published
4	D29	49136	NC_001900	[48]
5	Solon	49487	NC_011267	not published
6	Bxb1	50550	NC_002656	[58]
7	Bxz2	50913	NC_004682	[57]
8	Jasper	50968	NC_011020	not published
9	U2	51277	NC_009877	[103]
10	Lockley	51478	NC_011021	not published
11	DD5	51621	NC_011022	not published
12	Che12	52047	NC_008203	[103]
13	Bethlehem	52250	NC_009878	[103]
14	L5	52297	NC_001335	[104]
15	TM4	52797	NC_003387	[59]
16	Pukovnik	52892	NC_011023	not published
17	KBG	53572	NC_011019	[57]
18	Giles	54512	NC_009993	[60]
19	Che9d	56276	NC_004686	[57]
20	PMC	56692	NC_008205	[103]
21	Llij	56852	NC_008196	[103]
22	Che9c	57050	NC_004683	[57]
23	Boomer	58037	NC_011054	not published
24	Fruitloop	58471	NC_011288	not published
25	Pacc40	58554	NC_011287	not published
26	Ramsey	58578	NC_011289	not published
27	Che8	59471	NC_004680	[57]
28	PBI1	64494	NC_008198	[103]
29	Adjutor	64511	NC_010763	not published
30	Butterscotch	64562	NC_011286	not published
31	Tweety	58695	NC_011285	[61]
32	Troll4	64618	NC_011285	not published
33	PLot	64787	NC_008200	[103]
34	Gumball	64807	NC_011290	not published
35	Qyrzula	67188	NC_008204	[103]
36	Rosebush	67480	NC_004684	[57]
37	Phaedrus	68090	NC_011057	not published
38	Orion	68427	NC_008197	[103]
39	Chah	68450	NC_011284	not published
40	Konstantine	68952	NC_011292	not published
41	PG1	68999	NC_005259	not published
42	Pipefish	69059	NC_008199	[103]
43	Corndog	69777	NC_004685	[57]
44	Nigel	69904	NC_011044	not published
45	Predator	70110	NC_011039	not published
46	Cooper	70654	NC_008195	[103]
47	Barnyard	70797	NC_004689	[57]
48	244	74483	NC_008194	[103]
49	Kostya	75811	NC_011056	not published
50	Cjw1	75931	NC_004681	[57]
51	Porky	76312	NC_011055	not published
52	Wildcat	78441	NC_008206	[103]
53	Omega	110865	NC_004688	[57]
54	Catera	153766	NC_008207	[103]
55	Rizal	153894	NC_011272	not published
56	ScottMcG	154017	NC_011269	not published
57	Spud	154906	NC_011270	not published
58	Cali	155372	NC_011271	not published
59	Bxz1	156102	NC_004687	[57]
60	Myrna	164602	NC_011273	not published

Figure 4 Distribution of the genomic signature distances of *M. smegmatis* phages as a function of size of phage genomes[103,104]. Red symbol: Myoviridae, green symbol: Siphoviridae, blue symbol Podoviridae, white symbol: family not indicated. The numbers correspond to the phages listed in the Table.

perform due to the proximity of small groups of phage (as seen in Figure 4 and 5) that impedes a proper classification. By extrapolation, we have encircled the different clusters, taking as a limit the smallest and the longest genomes. Many phages not yet studied by genome sequence comparison can be added to the different groups. Others, like Omega, Gilles, Predator, Konstantine etc, seem to be more isolated. Group VI, composed only of Myoviridae, is the easiest to discern, whereas to the other group a zoom of the picture is necessary (Figure 5). As observed in *E. coli* and *S. aureus* phages, phage genomes that display significant similarities tend to have similar genomic distance signatures and similar genome size range. For example, cluster V contains phages with significant sequences similarity. However two subgroups are also possible to construct on the basis of genomic signature distances: subgroup A phages number 35, 36, 44 and 46; subgroup B phages number 37, 38, 39, 41, 42 (see table in Figure 4). Indeed, phages of subgroup B show a genomic signature that more closely resembles that of the host. Group II is a very homogenous group for both the genomic signature distance as well as for genome length. Phage Fruitloop appears outside of cluster III but shares a comparable genomic distance signature. The same observation is likely valid for the phages TM4 and Pukovnik that probably belong to group I. In contrast to *E. coli, S. aureus* and *P. aeruginosa* phages, no genomes of Podoviridae infecting *M. smegmatis* have been sequenced, probably because this morphotype (short tail) is not adapted to the complex cell wall of this bacterium

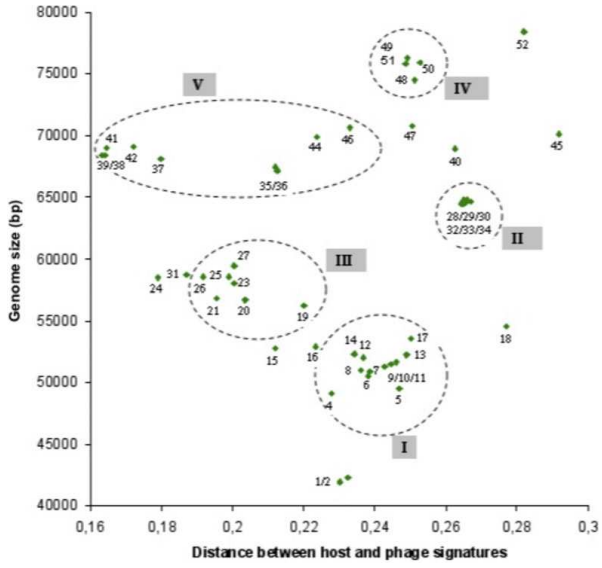


Figure 5 Zoom of Figure 4 allowing to visualize groups of genomes between 40 and 80 kb. Red symbol: Myoviridae, green symbol: Siphoviridae, blue symbol Podoviridae, white symbol: family not indicated. The numbers refer to the Table in Figure 4.

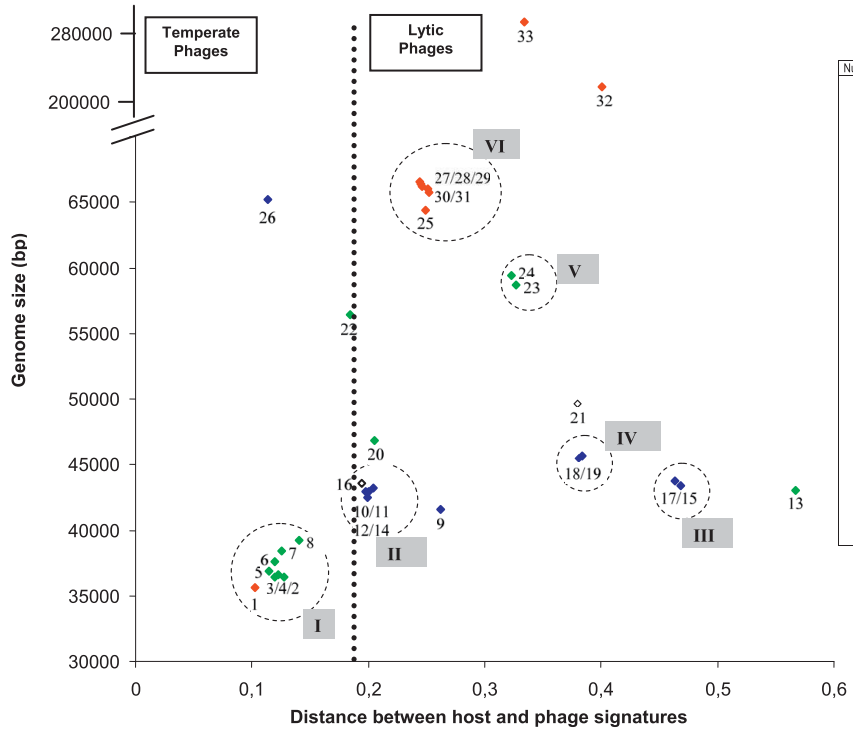
[18]. It is clear that, like all other clustering attempts, our representation is unlikely to completely reflect reality, and as more phages genomes infecting the same host become known, better clustering will likely occur.

M. smegmatis phage life styles

Contrary to *E. coli, S. aureus* and *P. aeruginosa* phages, the distinction between lytic and temperate life style seems less easy to establish for the different mycobacteriophages. As explained in [57],"most of the phages form plaques with hazy appearance, not obviously either clear or turbid". However stable lysogens can be isolated from these hazy plaques. D29 is a lytic phage very similar to temperate phage L5 [48]. Its status as a lytic phage is due to a 3,6 Kb deletion that removes the repressor. Bxb1 is a temperate phage that forms turbid plaques with a halo, probably due to an enzymatic activity associate with tail particles [58]. TM4 is not considered a temperate phage, thought it was isolated after Mitomicyn C treatment, because no integrase or repressor homolog are present in its genome [59]. Giles and Tweety forms lightly turbid plaques, reflecting a low frequency of lysogeny, but can be considered as temperate because they possess integrases [60,61]. The picture of the genomic distance signatures shows that nearly all the phage genomes are distributed around the average distance. It is interesting to note that Brujita, Che9c and Corndog are more close to their host. Therefore, here we can't propose a "frontier" between lytic and temperate mycobacteriophages. Several hypothesis could explained this fact: (1) all the mycobacteriophages isolated until now are temperate (or are derivatives of temperate like D29); (2) the determination of life style on the basis of plaques morphologies (or the laboratory conditions) is not adapted to the mycobacteriophages; (3) finally we can imagine that these mycobacteriophages have only recently been able to infect *Mycobacterium smegmatis*, or have a different life style as chronic infection, and therefore the amelioration process can't be yet detected by the genomic signature distances.

Pseudomonas aeruginosa caudoviridae

Thirty-three completely sequenced genomes of phages belonging to the order Caudoviridae and infecting *Pseudomonas aeruginosa* are available in the Genbank phage database. It should be noted that, a significant number of these phages have a %GC significantly lower than that of the host (65%). As seen in Figure 6, although several phages genomes with a GC% that resembles that of the host (e.g. MP22, D3112, B3) show short distances, some others (e.g. YuA, M6) have a similar %GC and a greater distance. In addition, phiKZ, (like T4) has a very low GC% (33%), but the calculated distance is less than that of phage 73 that has a 20% greater GC%. Different hypotheses have been proposed to explain this



Number	name	Genome size	Accession	Reference
1	phiCTX	35580	NC_003278	[105]
2	MP22	36409	NC_009818	[64]
3	DMS3	36415	NC_008717	[106]
4	MP29	36632	NC_011613	not published
5	MP38	36885	NC_011611	not published
6	D3112	37611	NC_005178	[107]
7	B3	38439	NC_006548	[65]
8	F10	39199	NC_007805	[62]
9	LKA1	41593	NC_009936	[66]
10	phiKMV	42519	NC_005045	[108]
11	PT5	42954	NC_011105	not published
12	PT2	42961	NC_011107	not published
13	73	42999	NC_007806	[62]
14	LKD16	43200	NC_009935	[66]
15	119X	43365	NC_007807	[62]
16	LUZ19	43548	NC_010326	not published
17	PaP2	43783	NC_005884	not published
18	PaP3	45503	NC_004466	[69]
19	LUZ24	45625	NC_010325	[70]
20	PAJU2	46872	NC_011373	[67]
21	PA11	49639	NC_007808	[62]
22	D3	56425	NC_002484	[71]
23	YuA	58663	NC_010116	[68]
24	M6	59446	NC_007809	[62]
25	LBL3	64427	NC_011165	not published
26	F116	65195	NC_006552	[109]
27	PB1	65764	NC_011810	not published
28	F8	66015	NC_007810	[62]
29	14-1	66235	NC_011703	not published
30	SN	66390	NC_011756	not published
31	LMA2	66530	NC_011166	not published
32	EL	211215	NC_007623	[110]
33	phiKZ	280334	NC_004629	[111]

Figure 6 Distribution of the genomic signature distances of *P. aeruginosa* phages as a function of size of phage genomes [105-111]. Red symbol: Myoviridae, green symbol: Siphoviridae, blue symbol Podoviridae, white symbol: family not indicated. The numbers correspond to the phages listed in the Table. On the Y axis, a discontinuity was added to accommodate phages 32 and 33.

phenomenon such as the fact that recent horizontal gene transfers in phages infecting hosts with a lower % GC may allow these phages to interact. It is also possible that this large range of %GC is a characteristic of these phages [62]. However, this variation in %GC may also reflect the known high phylogenetic versatility of the *Pseudomonas* genus [63].

***P. aeruginosa* phage groups**

As observed for the phages infecting the three other hosts used in this study, it was possible to group the phages as a function of the distance and the genome length (Figure 6). An 8 groups K-means classification is in agreement with this classification. The higher number of groups is due to the isolated phages. Group I is composed of "Mu-like" genomes, but are all Siphoviridae, with the exception of phiCTX. Phage MP22 for example, has been recently sequenced [64] and is highly similar to D3112 except in the gene c and in the late genes of virion morphogenesis. DMS3 has been described has having a high degree of similarity with phage D3112. B3 belongs to the same group of transposable phages and displayed some genetic relationships with D3112 using DNA hybridization [65] and sequence comparisons [62]. MP29, MP38 and F10 are encircled in the same group, whereas F10 presents no significant sequence similarities with B3 and D3112 [62]. The second group is composed of T7 super-group phages, such as phiKMV and LKD16. The unpublished genomes of phages PT5, PT2 and Luz19 are present in the same group. Phages LKD16 and phiKMV present 83% DNA homology with significant differences localized in their early regions [66]. In contrast, LKA1 only show homology at the protein level (48% of the predicted proteins) with phiKMV. Phages 119X, LUZ19 and PaP2 have genomes with very similar length, but only 119x and PaP2 show very similar distances, and the presence of group III is supported by the positive nucleotide comparison between these two phages [62]. Phages LUZ24 and PaP3 are Podoviridae that share 71% nucleotide identities, are grouped, and also share the same genomic signature distance. 24% of the PAJU2 genome is similar to that of phage D3 and 46% of the PAJU2 predicted proteins show similarity with D3 proteins [67], but the nearly 10 kb genome length difference appears to separate them. The last group reported in the literature is the one containing phages M6 and YuA (group V). These two phages share 91% nucleotide identities [68] and have very similar genomic signature distances. Phages LBL3, PB1, F8, 14-1, SN and LMA2 probably constitute another coherent group (group VI). They all show a very homogenous

genomic signature distance, the same morphotype and a genome length of 64-66 KB. A Dot-plot genome comparison shows significant nucleotide similarities between these genomes (data not shown).

P. aeruginosa phage life styles

The overall landscape of phages infecting *P. aeruginosa* seems less easy to differentiate between lytic and temperate. Indeed the distance observed for the phiKMV group, although higher than the distance observed for the D3112 temperate group, is less different than what is observed for the T7 group of *E. coli*. However, in contrast with the *M. smegmatis* phages, it seems possible to propose a demarcation point separating the lytic and temperate phages, although several atypical cases remain. PaP3, for example, has been described as a temperate phage and LUZ24 as a lytic phage, but their behavior is not totally clear. Indeed, the integration of PaP3 in the host genome has only been demonstrated by restriction enzyme analysis [69], and no indication of immunity or reactivation of an integrated PaP3 prophage is available. On the other hand, LUZ24 forms clear plaques on 36 different strains of *P. aeruginosa*, but small and turbid plaques on strain PAO [70]. The distance (0.38) observed for these two phages is more compatible with a distance characteristic of other lytic phages with genome length of the same order (such as the group of *E. coli* phage containing K1E, K1-5). D3 is a temperate phage with a lambdoid organization, and homologies with HK022/HK97 have been established [71]. A putative integrase has also been detected in the genome of PAJU2, and a lysogenic strain has been isolated [67]. Phage YuA has been described as a temperate phage and it possesses a putative repressor and integrase. But, like phage phiJL001 with which significant similarity is observed, isolation of a stable lysogenic strain was not possible [68]. The YuA distance is more characteristic of the other lytic phages, however it is always possible that the capacity of YuA to infect *P. aeruginosa* is recent and that its genome has not yet evolved through an amelioration process. Finally, like *E. coli* phages P1 and N15, F116 shows a very short distance confirming the hypothesis that the amelioration process is more efficient for phages able to lysogenize their hosts in a plasmid form.

Conclusions

Bacteriophage genome comparisons, without the need to use tools based on sequence homology is possible using genomic signatures. Our analysis and results, present in one picture per host, allow us to group the phages infecting *E. coli, S. aureus, M. smegmatis* and *P. aeruginosa* and to determine their life cycle (temperate *vs*. lytic).

The hypothesis of the "amelioration" process for the genomes of temperate phages is reinforced by our results. Indeed, the majority of the temperate phages display a shorter genomic signature distance between their genome and that of their host than that of the lytic phages. The genomic distance signature can therefore be a useful tool to predict phage life style.

Finally, putative evolutionary groups, for which available data is often dispatched over a fragmented scientific literature, have been identified on the basis of a conserved genomic signature distance for a coherent genome size range. The genomic signature distance could therefore be a useful tool to assign, without homology sequence comparison, a new phage sequence DNA to a known phage group.

Methods

1/DNA sequences

Caudoviridae (dsDNA) viral genomes infecting four bacterial species and their corresponding host sequences were retrieved from GenBank phage database: *Escherichia coli* (46 phage genomes), *Pseudomonas aeruginosa* (33 phage genomes), *Staphylococcus aureus* (50 phage genomes) and *Mycobacterium smegmatis* (60 phage genomes).

2/Sequence signatures

The signature of each sequence is defined as the frequencies of all possible tetranucleotides in the two strands of a sequence represented by a vector. The four hosts under study and their respective phages display large differences in base composition: *E. coli* (strain w3110) GC% = 50.8, *P. aeruginosa* (strain PA7) GC% = 66.6, *S. aureus* (strain RF122) GC% = 32.8, *M. smegmatis* (strain MC2-155) GC% = 67.4. Genomic signatures depend on the relative nucleotide proportion within a genome [27]. As phages infecting the same hosts can present a broad spectrum of nucleotide base composition differences, in order to compare their signatures to their host, we standardized the signatures [27]. Assuming that the succession of nucleotides along a sequence follows a random model (a zero order Markov chain; i.e. that the probability of a particular nucleotide depends only on the nucleotide concentration), the probability to observe a given word is the product of the probabilities of its constituent letters. Therefore, we constructed mock signatures based on the genome base composition under consideration. These signatures were subtracted from the genomic signature of the genome studied in order to obtain the standardized signature.

In order to compare genome signatures, we computed the Euclidian distance $\sqrt{\sum_i (V_i - H_i)^2}$ between host

and virus signatures: where V corresponds to the virus signature and H to that of the host and i indicates the tetranucleotide under consideration.

Author details
[1]Molécules thérapeutiques *in silico* (MTI), INSERM UMR-M 973, Université Paris Diderot - Paris 7, Bât Lamarck, 35 rue Hélène Brion, 75205, Paris Cedex 13, France. [2]Université Paris-Sud XI, Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiments 409, 91405 Orsay, France.

Authors' contributions
PD designed the study, made the experiments and helped in writing the manuscript. CR analyzed the results, drew the figures and wrote the manuscript. MSD helped in analyzing and presenting the results and helped in writing the manuscript. All authors read and approved the manuscript.
We thank the anonymous referee for his comments and suggestions.

Competing interests
The authors declare that they have no competing interests.

Received: 19 April 2010 Accepted: 17 July 2010 Published: 17 July 2010

References
1. Hatfull GF: **Bacteriophage genomics.** *Curr Opin Microbiol* 2008, **11(5)**:447-53.
2. Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: the unseen majority.** *Proc Natl Acad Sci USA* 1998, **95(12)**:6578-83.
3. Wommack KE, Colwell RR: **Virioplankton: viruses in aquatic ecosystems.** *Microbiol Mol Biol Rev* 2000, **64(1)**:69-114.
4. Weinbauer MG: **Ecology of prokaryotic viruses.** *FEMS Microbiol Rev* 2004, **28(2)**:127-81.
5. Suttle CA: **Marine viruses–major players in the global ecosystem.** *Nat Rev Microbiol* 2007, **5(10)**:801-12.
6. Thingstad TF, Lignell R: **Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand.** *Aquat Microb Ecol* 1997, **13**:19-27.
7. Weinbauer MG, Rassoulzadegan F: **Are viruses driving microbial diversification and diversity?** *Environ Microbiol* 2004, **6(1)**:1-11.
8. Canchaya C, Fournous G, Chibani-Chennoufi S, Dilmann ML, Brussow H: **Phage as agents of lateral gene transfer.** *Curr Opin Microbiol* 2003, **6**:417-424.
9. Waldor MK, Mekalanos JJ: **Lysogenic conversion by a filamentous phage encoding cholera toxin.** *Science* 1996, **272(5270)**:1910-4.
10. Ohnishi M, Kurokawa K, Hayashi T: **Diversification of Escherichia coli genomes: are bacteriophages the major contributors?** *Trends Microbiol* 2001, **9(10)**:481-5.
11. Williamson SJ, Rusch DB, Yooseph S, Halpern AL, Heidelberg KB, Glass JI, Fadrosh D, Miller CS, Sutton G, Frazier M, Venter JC: **The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples.** *PLoS One* 2008, **3(1)**:e1456.
12. Lawrence JG, Hatfull GF, Hendrix RW: **Imbroglios of Viral Taxonomy: Genetic Exchange and Failings of Phenetic Approaches.** *J Bact* 2002, **184**:4891-4905.
13. Brüssow H, Hendrix RW: **Phage Genomics: Small Is Beautiful.** *Cell* 2002, **108**:13-16.
14. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M: **Nucleotide sequence of bacteriophage phi X174 DNA.** *Nature* 1977, **265(5596)**:687-95.
15. Serwer P, Hayes SJ, Thomas JA, Hardies SC: **Propagating the missing bacteriophages: a large bacteriophage in a new class.** *Virol J* 2007, **4**:21.
16. Ackermann HW: **5500 Phages examined in the electron microscope.** *Arch Virol* 2007, **152(2)**:227-43.
17. Dorscht J, Klumpp J, Bielmann R, Schmelcher M, Born Y, Zimmer M, Calendar R, Loessner MJ: **Comparative genome analysis of Listeria bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site.** *J Bacteriol* 2009, **191(23)**:7206-15.
18. Hatfull GF, Cresawn SG, Hendrix RW: **Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution.** *Res Microbiol* 2008, **159(5)**:332-9.
19. Kwan T, Liu J, DuBow M, Gros P, Pelletier J: **The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages.** *Proc Natl Acad Sci USA* 2005, **102(14)**:5174-9.
20. Sullivan MB, Coleman ML, Quinlivan V, Rosenkrantz JE, Defrancesco AS, Tan G, Fu R, Lee JA, Waterbury JB, Bielawski JP, Chisholm SW: **Portal protein diversity and phage ecology.** *Environ Microbiol* 2008, **10(10)**:2810-23.
21. Karlin S, Burge C: **Dinucleotide relative abundance extremes: a genomic signature.** *Trends In Genetics* 1995, **11**:283-290.
22. Karlin S, Mràzek J, Campbell AM: **Compositional biases of bacterial genomes and evolutionary implications.** *J Bact* 1997, **179**:3899-3913.
23. Campbell A, Mràzek J, Karlin S: **Genome signature comparisons among prokaryote, plasmid and mitochondrial DNA.** *Proc Nat Acad Sci USA* 1999, **96**:9184-9189.
24. Coenye T, Vandamme P: **Use of the genomic signature in bacterial classification and identification.** *Syst Appl Microbiol* 2004, **27(2)**:175-85.
25. van Passel, Kuramae EE, Luyf AC, Bart A, Boekhout T: **The reach of the genome signature in prokaryotes.** *BMC Evol Biol* 2006, **6**:84.
26. Blaisdell BE, Campbell AM, Karlin S: **Similarities and dissimilarities of phage genomes.** *Proc Natl Acad Sci USA* 1996, **93(12)**:5854-9.
27. Deschavanne PJ, Giron A, Vilain J, Fagot G, Fertil B: **Genomic signature: characterization and classification of species assessed by Chaos Game Representation of sequences.** *Molecular Biology and Evolution* 1999, **16**:1391-1399.
28. Rosas-Magallanes V, Deschavanne P, Quintana-Murci L, Brosch R, Gicquel B, Neyrolles O: **Horizontal Transfer of a Virulence Operon to the Ancestor of Mycobacterium tuberculosis.** *Mol Biol Evol* 2006, **23**:1129-1135.
29. Dufraigne C, Fertil B, Lespinats S, Giron A, Deschavanne P: **Detection and characterization of horizontal transfers in prokaryotes using genomic signature.** *Nucleic Acids Res* 2005, **33(1)**:e6.
30. Karlin S: **Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes.** *Trends in Microbiology* 2001, **9(7)**:335-343.
31. Regeard C, Maillard J, Dufraigne C, Deschavanne P, Holliger C: **Indications for acquisition of reductive dehalogenase genes through horizontal gene transfer by Dehalococcoides ethenogenes strain 195.** *Appl Environ Microbiol* 2005, **71(6)**:2955-61.
32. Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, Deschavanne P: **Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli.** *Mol Biol Evol* 2007, **24(8)**:1861-71.
33. van Passel MW, Bart A, Thygesen HH, Luyf AC, van Kampen AH, van der Ende A: **An acquisition account of genomic islands based on genome signature comparisons.** *BMC Genomics* 2005, **6**:163.
34. Srividhya KV, Alaguraj V, Poornima G, Kumar D, Singh GP, Raghavenderan L, Katta AV, Mehta P, Krishnaswamy S: **Identification of prophages in bacterial genomes by dinucleotide relative abundance difference.** *PLoS One* 2007, **2(11)**:e1193.
35. Pride DT, Wassenaar TM, Ghose C, Blaser MJ: **Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses.** *BMC Genomics* 2006, **7**:8.
36. Teeling H, Meyerdierks A, Bauer M, Amann R, Glockner FO: **Application of tetranucleotide frequencies for the assignment of genomic fragments.** *Environ Microbiol* 2004, **6(9)**:938-47.
37. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F: **The marine viromes of four oceanic regions.** *PLoS Biol* 2006, **4(11)**:e368.
38. Kristensen DM, Mushegian AR, Dolja VV, Koonin EV: **New dimensions of the virus world discovered through metagenomics.** *Trends Microbiol* 2009.
39. Casjens SR: **Diversity among the tailed-bacteriophages that infect the Enterobacteriaceae.** *Res Microbiol* 2008, **159(5)**:340-8.
40. Jamalludeen N, Kropinski AM, Johnson RP, Lingohr E, Harel J, Gyles CL: **Complete genomic sequence of bacteriophage phiEcoM-GJ1, a novel phage that has myovirus morphology and a podovirus-like RNA polymerase.** *Appl Environ Microbiol* 2008, **74(2)**:516-25.
41. Hong J, Kim KP, Heu S, Lee SJ, Adhya S, Ryu S: **Identification of host receptor and receptor-binding module of a newly sequenced T5-like phage EPS7.** *FEMS Microbiol Lett* 2008, **289(2)**:202-9.

42. Wang J, Jiang Y, Vincent M, Sun Y, Yu H, Bao Q, Kong H, Hu S: **Complete genome sequence of bacteriophage T5.** *Virology* 2005, **332(1)**:45-65.
43. Savalia D, Westblade LF, Goel M, Florens L, Kemp P, Akulenko N, Pavlova O, Padovan JC, Chait BT, Washburn MP, Ackermann HW, Mushegian A, Gabisonia T, Molineux I, Severinov K: **Genomic and proteomic analysis of phiEco32, a novel Escherichia coli bacteriophage.** *J Mol Biol* 2008, **377(3)**:774-89.
44. Morgan GJ, Hatfull GF, Casjens S, Hendrix RW: **Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus.** *J Mol Biol* 2002, **317(3)**:337-59.
45. Lawrence J, Ochman H: **Amelioration of bacterial genomes: rates of change and exchange.** *Journal of Molecular Evolution* 1997, **44**:383-397.
46. Marri PR, Golding GB: **Gene amelioration demonstrated: the journey of nascent genes in bacteria.** *Genome* 2008, **51(2)**:164-8.
47. Labrie S, Moineau S: **Complete genomic sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages.** *Virology* 2002, **296(2)**:308-20.
48. Ford ME, Sarkis GJ, Belanger AE, Hendrix RW, Hatfull GF: **Genome Structure of Mycobacteriophage D29: Implications for Phage Evolution.** *J Mol Biol* 1998, **279**:143-164.
49. Roberts MD, Martin NL, Kropinski AM: **The genome and proteome of coliphage T1.** *Virology* 2004, **318(1)**:245-66.
50. Ma XX, Ito T, Chongtrakool P, Hiramatsu K: **Predominance of clones carrying Panton-Valentine leukocidin genes among methicillin-resistant Staphylococcus aureus strains isolated in Japanese hospitals from 1979 to 1985.** *J Clin Microbiol* 2006, **44(12)**:4515-27.
51. Ma XX, Ito T, Kondo Y, Cho M, Yoshizawa Y, Kaneko J, Katai A, Higashiide M, Li S, Hiramatsu K: **Two different Panton-Valentine leukocidin phage lineages predominate in Japan.** *J Clin Microbiol* 2008, **46(10)**:3246-58.
52. Bae T, Baba T, Hiramatsu K, Schneewind O: **Prophages of Staphylococcus aureus Newman and their contribution to virulence.** *Mol Microbiol* 2006, **62(4)**:1035-47.
53. Narita S, Kaneko J, Chiba J, Piemont Y, Jarraud S, Etienne J, Kamio Y: **Phage conversion of Panton-Valentine leukocidin in Staphylococcus aureus: molecular analysis of a PVL-converting phage, phiSLT.** *Gene* 2001, **268(1-2)**:195-206.
54. Vybiral D, Takac M, Loessner M, Witte A, von Ahsen U, Blasi U: **Complete nucleotide sequence and molecular characterization of two lytic Staphylococcus aureus phages: 44AHJD and P68.** *FEMS Microbiol Lett* 2003, **219(2)**:275-83.
55. Zou D, Kaneko J, Narita S, Kamio Y: **Prophage, phiPV83-pro, carrying panton-valentine leukocidin genes, on the Staphylococcus aureus P83 chromosome: comparative analysis of the genome structures of phiPV83-pro, phiPVL, phi11, and other phages.** *Biosci Biotechnol Biochem* 2000, **64(12)**:2631-43.
56. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW: **Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size.** *J Mol Biol* 2010.
57. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR, Hendrix RW, Hatfull GF: **Origins of Highly Mosaic Mycobacteriophage Genomes.** *Cell* 2003, **113**:171-182.
58. Mediavilla J, Jain S, Kriakov J, Ford ME, Duda RL, Jacobs WR Jr, Hendrix RW, Hatfull GF: **Genome organization and characterization of mycobacteriophage Bxb1.** *Mol Microbiol* 2000, **38(5)**:955-70.
59. Ford ME, Stenstrom C, Hendrix RW, Hatfull GF: **Mycobacteriophage TM4: genome structure and gene expression.** *Tuber Lung Dis* 1998, **79(2)**:63-73.
60. Morris P, Marinelli LJ, Jacobs-Sera D, Hendrix RW, Hatfull GF: **Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination.** *J Bacteriol* 2008, **190(6)**:2172-82.
61. Pham TT, Jacobs-Sera D, Pedulla ML, Hendrix RW, Hatfull GF: **Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria.** *Microbiology* 2007, **153(Pt 8)**:2711-23.
62. Kwan T, Liu J, Dubow M, Gros P, Pelletier J: **Comparative genomic analysis of 18 Pseudomonas aeruginosa bacteriophages.** *J Bacteriol* 2006, **188(3)**:1184-7.
63. Mulet M, Lalucat J, Garcia-Valdes E: **DNA sequence-based analysis of the Pseudomonas species.** *Environ Microbiol* 2010.
64. Heo YJ, Chung IY, Choi KB, Lau GW, Cho YH: **Genome sequence comparison and superinfection between two related Pseudomonas aeruginosa phages, D3112 and MP22.** *Microbiology* 2007, **153(Pt 9)**:2885-95.
65. Braid MD, Silhavy JL, Kitts CL, Cano RJ, Howe MM: **Complete genomic sequence of bacteriophage B3, a Mu-like phage of Pseudomonas aeruginosa.** *J Bacteriol* 2004, **186(19)**:6560-74.
66. Ceyssens PJ, Lavigne R, Mattheus W, Chibeu A, Hertveldt K, Mast J, Robben J, Volckaert G: **Genomic analysis of Pseudomonas aeruginosa phages LKD16 and LKA1: establishment of the phiKMV subgroup within the T7 supergroup.** *J Bacteriol* 2006, **188(19)**:6924-31.
67. Uchiyama J, Rashel M, Matsumoto T, Sumiyama Y, Wakiguchi H, Matsuzaki S: **Characteristics of a novel Pseudomonas aeruginosa bacteriophage, PAJU2, which is genetically related to bacteriophage D3.** *Virus Res* 2009, **139(1)**:131-4.
68. Ceyssens PJ, Mesyanzhinov V, Sykilinda N, Briers Y, Roucourt B, Lavigne R, Robben J, Domashin A, Miroshnikov K, Volckaert G, Hertveldt K: **The genome and structural proteome of YuA, a new Pseudomonas aeruginosa phage resembling M6.** *J Bacteriol* 2008, **190(4)**:1429-35.
69. Tan Y, Zhang K, Rao X, Jin X, Huang J, Zhu J, Chen Z, Hu X, Shen X, Wang L, Hu F: **Whole genome sequencing of a novel temperate bacteriophage of P. aeruginosa: evidence of tRNA gene mediating integration of the phage genome into the host bacterial chromosome.** *Cell Microbiol* 2007, **9(2)**:479-91.
70. Ceyssens PJ, Hertveldt K, Ackermann HW, Noben JP, Demeke M, Volckaert G, Lavigne R: **The intron-containing genome of the lytic Pseudomonas phage LUZ24 resembles the temperate phage PaP3.** *Virology* 2008, **377(2)**:233-8.
71. Kropinski AM: **Sequence of the genome of the temperate, serotype-converting, Pseudomonas aeruginosa bacteriophage D3.** *J Bacteriol* 2000, **182(21)**:6066-74.
72. Halling C, Calendar R, Christie GE, Dale EC, Deho G, Finkel S, Flensburg J, Ghisotti D, Kahn ML, Lane KB, *et al*: **DNA sequence of satellite bacteriophage P4.** *Nucleic Acids Res* 1990, **18(6)**:1649.
73. Pajunen MI, Elizondo MR, Skurnik M, Kieleczawa J, Molineux IJ: **Complete nucleotide sequence and likely recombinatorial origin of bacteriophage T3.** *J Mol Biol* 2002, **319(5)**:1115-32.
74. Clark AJ, Inwood W, Cloutier T, Dhillon TS: **Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages.** *J Mol Biol* 2001, **311(4)**:657-79.
75. Scholl D, Merril C: **The genome of bacteriophage K1F, a T7-like phage that has acquired the ability to replicate on K1 strains of Escherichia coli.** *J Bacteriol* 2005, **187(24)**:8499-503.
76. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, Hendrix RW: **Genomic Sequences of Bacteriophages HK97 and HK022: Pervasive Genetic Mosaicism in the Lambdoid Bacteriophages.** *J Mol Biol* 2000, **299**:27-51.
77. Mertens H, Hausmann R: **Coliphage BA14: a new relative of phage T7.** *J Gen Virol* 1982, **62(Pt 2)**:331-41.
78. Dunn JJ, Studier FW: **Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements.** *J Mol Biol* 1983, **166(4)**:477-535.
79. Recktenwald J, Schmidt H: **The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage phiP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved.** *Infect Immun* 2002, **70(4)**:1896-908.
80. Scholl D, Kieleczawa J, Kemp P, Rush J, Richardson CC, Merril C, Adhya S, Molineux IJ: **Genomic analysis of bacteriophages SP6 and K1-5, an estranged subgroup of the T7 supergroup.** *J Mol Biol* 2004, **335(5)**:1151-71.
81. Stummeyer K, Schwarzer D, Claus H, Vogel U, Gerardy-Schahn R, Muhlenhoff M: **Evolution of bacteriophages infecting encapsulated bacteria: lessons from Escherichia coli K1-specific phages.** *Mol Microbiol* 2006, **60(5)**:1123-35.
82. Wietzorrek A, Schwarz H, Herrmann C, Braun V: **The genome of the novel phage Rtp, with a rosette-like tail tip, is homologous to the genome of phage T1.** *J Bacteriol* 2006, **188(4)**:1419-36.

83. Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB: **Nucleotide sequence of bacteriophage lambda DNA.** *J Mol Biol* 1982, **162(4)**:729-73.
84. German GJ, Misra R: **The TolC protein of Escherichia coli serves as a cell-surface receptor for the newly characterized TLS bacteriophage.** *J Mol Biol* 2001, **308(4)**:579-85.
85. Creuzburg K, Recktenwald J, Kuhle V, Herold S, Hensel M, Schmidt H: **The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-like type III effector protein.** *J Bacteriol* 2005, **187(24)**:8494-8.
86. Sato T, Shimizu T, Watarai M, Kobayashi M, Kano S, Hamabata T, Takeda Y, Yamasaki S: **Genome analysis of a novel Shiga toxin 1 (Stx1)-converting phage which is closely related to Stx2-converting phages but not to other Stx1-converting phages.** *J Bacteriol* 2003, **185(13)**:3966-71.
87. Miyamoto H, Nakai W, Yajima N, Fujibayashi A, Higuchi T, Sato K, Matsushiro A: **Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions.** *DNA Res* 1999, **6(4)**:235-40.
88. Plunkett G, Rose DJ, Durfee TJ, Blattner FR: **Sequence of Shiga toxin 2 phage 933W from Escherichia coli O157:H7: Shiga toxin as a phage late-gene product.** *J Bacteriol* 1999, **181(6)**:1767-78.
89. Sato T, Shimizu T, Watarai M, Kobayashi M, Kano S, Hamabata T, Takeda Y, Yamasaki S: **Distinctiveness of the genomic sequence of Shiga toxin 2-converting phage isolated from Escherichia coli O157:H7 Okayama strain as compared to other Shiga toxin 2-converting phages.** *Gene* 2003, **309(1)**:35-48.
90. Cho NY, Choi M, Rothman-Denes LB: **Rothman-Denes, The bacteriophage N4-coded single-stranded DNA-binding protein (N4SSB) is the transcriptional activator of Escherichia coli RNA polymerase at N4 late promoters.** *J Mol Biol* 1995, **246(4)**:461-71.
91. Lobocka MB, Rose DJ, Plunkett G, Rusin M, Samojedny A, Lehnherr H, Yarmolinsky MB, Blattner FR: **Genome of bacteriophage P1.** *J Bacteriol* 2004, **186(21)**:7032-68.
92. Desplats C, Dez C, Tetart F, Eleaume H, Krisch HM: **Snapshot of the genome of the pseudo-T-even bacteriophage RB49.** *J Bacteriol* 2002, **184(10)**:2789-804.
93. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W: **Bacteriophage T4 genome.** *Microbiol Mol Biol Rev* 2003, **67(1)**:86-156, table of contents.
94. Zuber S, Ngom-Bru C, Barretto C, Bruttin A, Brussow H, Denou E: **Genome analysis of phage JS98 defines a fourth major subgroup of T4-like phages in Escherichia coli.** *J Bacteriol* 2007, **189(22)**:8206-14.
95. Kwan T, Liu J, DuBow M, Gros P, Pelletier J: **The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages.** *Proc Natl Acad Sci USA* 2005, **102(14)**:5174-9.
96. Kaneko J, Kimura T, Narita S, Tomita T, Kamio Y: **Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phiPVL carrying Panton-Valentine leukocidin genes.** *Gene* 1998, **215(1)**:57-67.
97. Iandolo JJ, Worrell V, Groicher KH, Qian Y, Tian R, Kenton S, Dorman A, Ji H, Lin S, Loh P, Qi S, Zhu H, Roe BA: **Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of Staphylococcus aureus 8325.** *Gene* 2002, **289(1-2)**:109-18.
98. Matsuzaki S, Yasuda M, Nishikawa H, Kuroda M, Ujihara T, Shuin T, Shen Y, Jin Z, Fujimoto S, Nasimuzzaman MD, Wakiguchi H, Sugihara S, Sugiura T, Koda S, Muraoka A, Imai S: **Experimental protection of mice against lethal Staphylococcus aureus infection by novel bacteriophage phi MR11.** *J Infect Dis* 2003, **187(4)**:613-24.
99. Yamaguchi T, Hayashi T, Takami H, Nakasone K, Ohnishi M, Nakayama K, Yamada S, Komatsuzawa H, Sugai M: **Phage conversion of exfoliative toxin A production in Staphylococcus aureus.** *Mol Microbiol* 2000, **38(4)**:694-705.
100. Tallent SM, Langston TB, Moran RG, Christie GE: **Transducing particles of Staphylococcus aureus pathogenicity island SaPI1 are comprised of helper phage-encoded proteins.** *J Bacteriol* 2007, **189(20)**:7520-4.
101. Kuroda M, *et al*: **Whole genome sequencing of meticillin-resistant Staphylococcus aureus.** *Lancet* 2001, **357(9264)**:1225-40.
102. O'Flaherty S, Coffey A, Edwards R, Meaney W, Fitzgerald GF, Ross RP: **Genome of staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low G+C content.** *J Bacteriol* 2004, **186(9)**:2862-71.
103. Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME, Gonda RM, Houtz JM, Hryckowian AJ, Kelchner VA, Namburi S, Pajcini KV, Popovich MG, Schleicher DT, Simanek BZ, Smith AL, Zdanowicz GM, Kumar V, Peebles CL, Jacobs WR Jr, Lawrence JG, Hendrix RW: **Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform.** *PLoS Genet* 2006, **2(6)**:e92.
104. Hatfull GF: **Genetic transformation of mycobacteria.** *Trends in Microbiology* 1993, 310-4.
105. Nakayama K, Kanaya S, Ohnishi M, Terawaki Y, Hayashi T: **The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of Pseudomonas aeruginosa: implications for phage evolution and horizontal gene transfer via bacteriophages.** *Mol Microbiol* 1999, **31(2)**:399-419.
106. Zegans ME, Wagner JC, Cady KC, Murphy DM, Hammond JH, O'Toole GA: **Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of Pseudomonas aeruginosa.** *J Bacteriol* 2009, **191(1)**:210-9.
107. Wang PW, Chu L, Guttman DS: **Complete sequence and evolutionary genomic analysis of the Pseudomonas aeruginosa transposable bacteriophage D3112.** *J Bacteriol* 2004, **186(2)**:400-10.
108. Lavigne R, Burkal'tseva MV, Robben J, Sykilinda NN, Kurochkina LP, Grymonprez B, Jonckx B, Krylov VN, Mesyanzhinov VV, Volckaert G: **The genome of bacteriophage phiKMV, a T7-like virus infecting Pseudomonas aeruginosa.** *Virology* 2003, **312(1)**:49-59.
109. Byrne M, Kropinski AM: **The genome of the Pseudomonas aeruginosa generalized transducing bacteriophage F116.** *Gene* 2005, **346**:187-94.
110. Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, Korchevskii R, Robben J, Mesyanzhinov V, Krylov VN, Volckaert G: **Genome comparison of Pseudomonas aeruginosa large phages.** *J Mol Biol* 2005, **354(3)**:536-45.
111. Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, Krylov VN, Volckaert G: **The genome of bacteriophage phiKZ of Pseudomonas aeruginosa.** *J Mol Biol* 2002, **317(1)**:1-19.

doi:10.1186/1743-422X-7-163
Cite this article as: Deschavanne *et al*.: The use of genomic signature distance between bacteriophages and their hosts displays evolutionary relationships and phage growth cycle determination. *Virology Journal* 2010 **7**:163.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central