A meta-analysis of host - phage interaction matrices in *Streptococcus thermophilus*.

Eugenio PARENTE[1,2]*, Teresa ZOTTA[2], Annamaria RICCIARDI [1,2].
[1] Department of Agricultural, Forestry, Food and Environmental Sciences, University of Basilicata, Potenza, Italy, [2] Institute of Food Science, National Research Council, Avellino, Italy

Streptococcus thermophilus is one of the most important dairy starters. Its interaction with bacteriophages in dairy environments (both cheese and yoghurt production plants) has been reviewed recently (Quiberoni et al. 2010), and the high specificity of bacteriophage-host relationships in this species has been attributed to the CRISPR-CAS defence mechanism. Because of the scarcity of plasmid encoded phage resistance mechanisms compared to other dairy species, like *Lactococcus lactis*, bacteriophage control usually relies on traditional approaches, like rotation, use of phage unrelated multiple strain starters, use of direct to vat starters, etc. Phage-host interactions are usually presented as rectangular matrices (Phage Host Interaction Matrix PBIM) with bacteriophages on columns and hosts on lines. These matrices, if not reordered, are of little assistance in individuating groups of bacteriophages with similar host ranges or in analysing the structure of the matrix.
Recently, Flores et al. (2011) described a number of approaches for analysing the statistical structure of Phage bacteria Interaction Networks (PBINs) and pointed out that the significant modularity of PBINs for *S. thermophilus* (with blocks of hosts attacked by limited number of bacteriophages) contrast with the general trend for nestedness. This was postulated to be dependent on the high specificity of the CRISPR-CAS defence mechanism (Weitz et al., 2013). Flores et al. (2011) have developed BiMat (http://arxiv.org/pdf/1406.6732v2.pdf), a MATLAB package for the analysis of bipartite networks (i.e. networks in which two types of nodes exist, like bacteriophages and hosts and in which directed connections are possible only between one type of node and the other). BiMat provides extensive tools for the analysis of PBINs but requires some familiarity with MATLAB. The objective of this study was to examine 7 published studies for which PBIMs are available (Suarez et al., 2002; Quiberoni et al., 2003; Binetti et al., 2005; Quiberoni et al., 2006; Guglielmotti et al., 2009; Zinno et al., 2010; Ma et al., 2014) and an unpublished study (Parente et al., this meeting) in order to find simple and effective alternatives for the analysis of *S. thermophilus* PBINs.

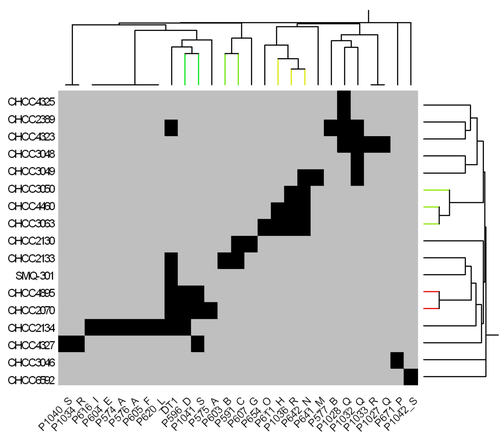
STEP 1. Unsorted Phage Bacteria Interaction Matrices

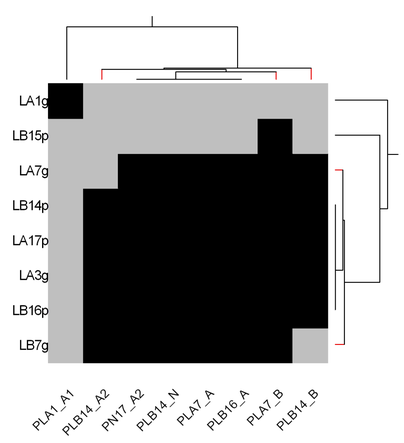
Host	p574_A	p575_A	p576_A	p577_B	p591_C	p596_D	p603_B	p604_E	p605_F	p607_G	p611_H	p616_I	p620_L	p641_M	p642_N	p654_O	p671_P	p1027_Q	p1028_Q	p1032_Q	p1033_R	p1034_R	p1036_R	p1040_S	p1041_S	p1042_S	DT1
CHCC3063	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
CHCC2070	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
CHCC2130	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHCC2133	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CHCC2134	1	0	1	0	0	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CHCC2389	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1
CHCC3048	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
CHCC3049	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0
CHCC3050	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
CHCC3046	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
CHCC4323	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
CHCC4325	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
CHCC4327	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0
CHCC4460	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
CHCC4895	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
CHCC6592	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
SMQ-301	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Host	pLA1_A1	pLA7_A	pLA7_B	pLB14_A2	pLB14_B	pLB14_N	pLB16_A	pN17_A2
LA17p	0	1	1	1	1	1	1	1
LA1g	1	0	0	0	0	0	0	0
LA3g	0	1	1	1	1	1	1	1
LA7g	0	1	1	0	1	1	1	1
LB14p	0	1	1	1	1	1	1	1
LB15p	0	0	1	0	0	0	0	0
LB16p	0	1	1	1	1	1	1	1
LB7g	0	1	1	1	0	1	1	1

Two PBIMs with high (Zinno et al., 2010; left) and low (Parente et al., 2014; right) modularity are shown as an example. Black cells indicate an interaction/infection (1), white cells (0) indicate lack of infection. Unsorted PBIMs are of little help in the analysis but can be used to calculate important features of the network: H number of hosts; P number of phages; S number of species (H+P); M size (HxP); I number of interactions; C connectance (I/M); LH mean number of interactions, host (I/H); LP mean number of interactions, phage (I/P).

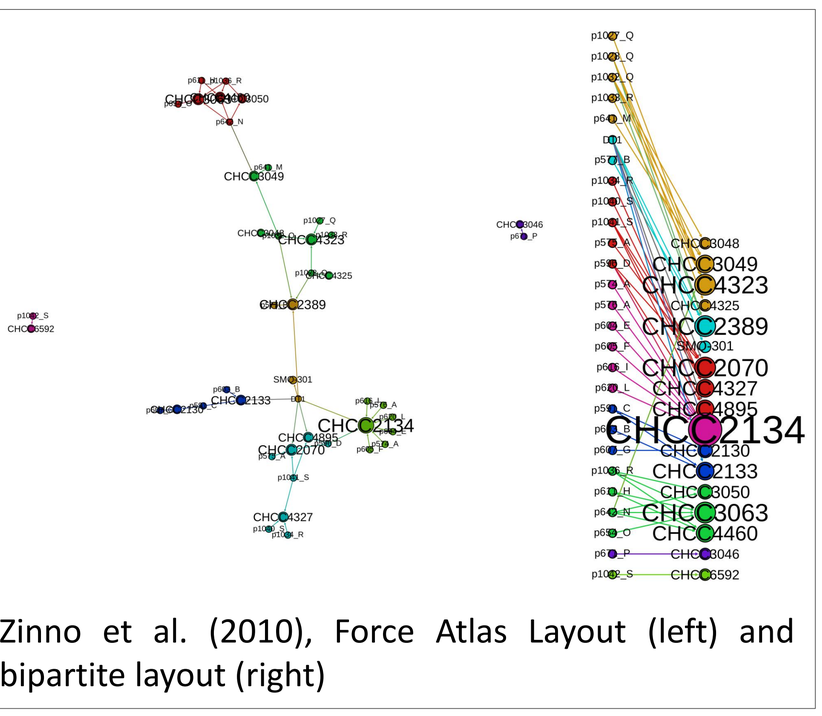
STEP 2. Sorted and clustered Phage Bacteria Interaction Matrices



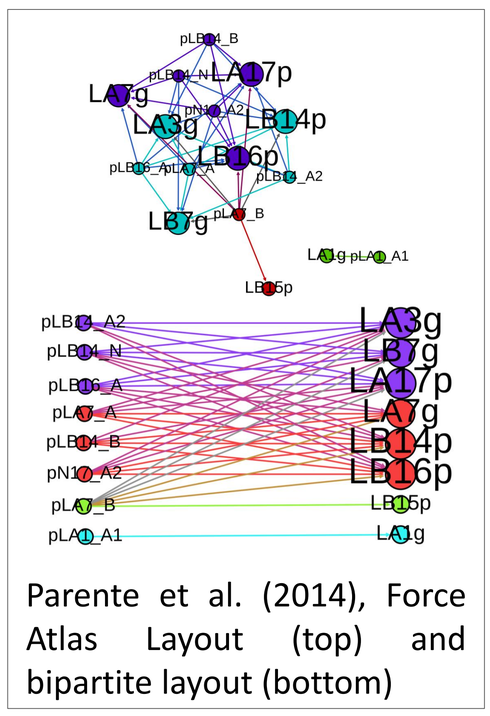


PBIMs can be easily reordered using matrix cluster analysis (here Jaccard's coefficient was used for calculating distances and UPGMA was used for clustering). Matrices for Zinno et al. (2010) and Parente et al. (2014) are shown as an example, but similar results were obtained with all the matrices analysed in this study. Black cells indicate infections and grey cells indicate lack of thereof. Although the ordering is not necessarily optimal, both the structure of the of the network (with small groups or modules of strongly connected phages and hosts) and the grouping of both hosts and phages are more evident and can be used in further applications (i.e. identification of phage unrelated strains).

STEP 3. Analysis of Phage Bacteria Interaction Networks

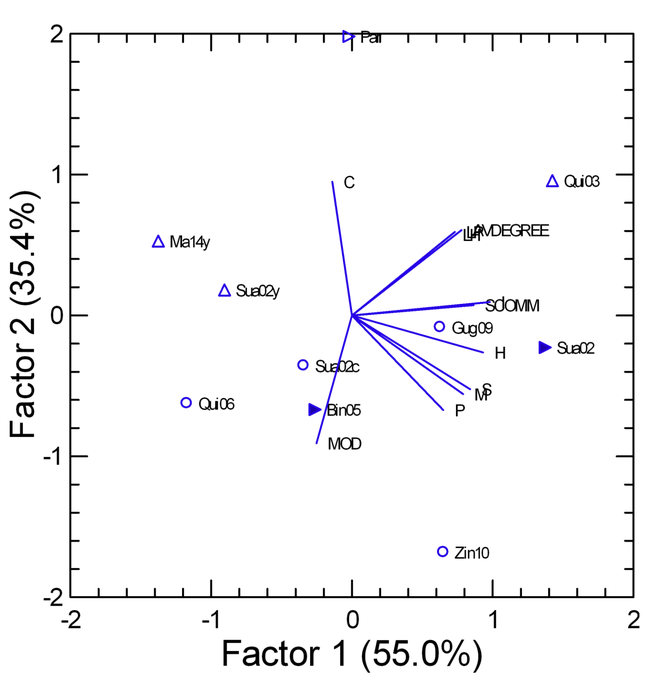


Zinno et al. (2010), Force Atlas Layout (left) and bipartite layout (right)



Parente et al. (2014), Force Atlas Layout (top) and bipartite layout (bottom)

Gephi is a powerful exploratory tool for network analysis. In PBINs two types of networks exist (phages and bacteria) and connections (edges) are possible only between phages and bacteria. Therefore PBINs are directed bipartite networks. Even if Gephi is not specifically designed for the analysis of these networks, it is easy to use to produce meaningful graphs of the PBINs, to calculate a number of important statistics for nodes (**indegree**, number of infecting phages for each host; **outdegree**, number of hosts infected by a phage; **centrality** statistics), network properties (**average degree**, average number of edges connected to a node; **density**, ratio of number of edges to the number of all possible edges; **modularity**, a measure of the strength of division of a network in modules: range -0.5 ÷1, positive values indicate modular structure and to identify communities of nodes). These properties can be used to enhance plots. Two examples for the PBIMs used above are shown on the right. In Zinno et al. (2010) PBIN both Force Atlas and Bipartite layout clearly show the modular structure of the network. Communities detected by network analysis largely overlap those detected by the cluster analysis described in step 2.



STEP 4. How similar are the PBINs for *S. thermophilus*?

An approach similar to that described by Flores et al. (2011) was used to compare several studies describing *S. thermophilus* PBINs. In addition to parameters describing the matrix (see step 1), the average degree, the modularity and the standardized number of communities were used. A Principal Component Analysis was carried out on the correlation matrix and the results are shown in the score and loading plot on the left. Different studies are identified by different symbols (O = cheese; △=yoghurt; ▶ mixed dairy sources; ▷ starters). For one study (Suarez et al., 2002) subsets of data were identified and were used separately. The first factor is mostly related to the size of the matrix and to the number of connections, while the second factor is clearly related to the modularity of the network. Most cheese studies are in the lower part of the graph, while yoghurt studies have an intermediate position. The two PBINS showed in the steps above are clearly at the extremes of the range. Although most studies showed evidence of modularity (Flores et al., 2011) differences were evident but seemed more related to the nature of the study (matrix, time span, isolation strategy). A more comprehensive study is needed to clarify if *S. thermophilus* PBINs are always modular.

References.
Binetti A, Del Rio B, Martin M, & Alvarez M (2005) Appl Environ Microbiol, 71: 6096–6103.
Guglielmotti DM, Binetti A, Reinheimer JA, & Quiberoni A (2009) Int Dairy J, 19: 476–480.
Flores CO, Meyer JR, Valverde S, Farr L, & Weitz JS (2011) PNAS, 108: E288–97.
Ma C, Pan N, Chen Z, Liu Z, Gong G, & Ma A (2014) Int Dairy J 35: 32–37.
Quiberoni A, Auad L, Binetti AG, Suárez VB, Reinheimer JA, & Raya RR (2003) Food Microbiol 20: 461–469.
Quiberoni A, Tremblay D, Ackermann H-W, Moineau S, & Reinheimer JA (2006) J Dairy Sci, 89: 3791–3799.
Quiberoni A, Moineau S, Rousseau GM, Reinheimer JA, & Ackermann H-W (2010) Int Dairy J, 20: 657–664.
Suárez, VB, Quiberoni A, Binetti AG, & Reinheimer JA (2002). J Food Prot, 65:1597-1604
Weitz, J. S., Poisot, T., Meyer, J. R., Flores, C. O., Valverde, S., Sullivan, M. B., & Hochberg, M. E. (2013). Trends Microbiol., 21: 82–91.
Zinno, P., Janzen, T., Bennedsen, M., Ercolini, D., & Mauriello, G. (2010). Int. J. Food Microbiol., 138: 137–144.

LAB11 11th International Symposium on Lactic Acid Bacteria

Scan this to access a copy of this poster