UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE AGRARIE E ALIMENTARI
Department of Food, Environmental and Nutritional Sciences (DeFENS)

Graduate School in Molecular Sciences and Plant, Food and Environmental Biotechnology

PhD programme in Food Science, Technology and Biotechnology

XXVIII cycle

POLYPHASIC APPROACH TO STUDY THE EMERGING PATHOGEN *LACTOCOCCUS GARVIEAE*

Scientific field AGR/16

GIOVANNI ERACLIO

Tutor: Prof. Maria Grazia Fortina

PhD Coordinator: Prof. Maria Grazia Fortina

2014/2015

LIST OF CONTENTS

ABSTRACT 1

RIASSUNTO 3

PREFACE 5

STATE OF THE ART 8

Lactococcus garvieae 8

Insertion sequences and temperate bacteriophages as promoter of genome plasticity 9

Biocontrol of *Lactococcus garvieae* 10

Phage therapy 11

AIM OF THE STUDY 13

REFERENCES 14

1. PHYLOGENETIC RELATIONSHIPS BETWEEN *LACTOCOCCUS GARVIEAE* AND *LACTOCOCCUS LACTIS* 19

MATERIALS AND METHODS 19

Lactococcus garvieae strains 19

DNA extraction 19

Multi Locus Sequence Typing (MLST) 19

Genome analysis and comparison 20

RESULTS AND DISCUSSION 21

Multilocus Sequence Typing (MLST) 21

Genome comparison 27

CONCLUSIONS 30

REFERENCES 31

2. INSERTION SEQUENCE ELEMENTS IN *LACTOCOCCUS GARVIEAE* 33

MATERIALS AND METHODS 34

RESULTS AND DISCUSSIONS 35

CONCLUSIONS 42

REFERENCES 43

3. PHENOTYPIC AND GENOTYPIC DIFFERENCES BETWEEN EUROPEAN AND AMERICAN STRAINS OF *LACTOCOCCUS GARVIEAE* 45

MATERIALS AND METHODS 46

Sampling, selection, culturing and identification of *L. garvieae* isolates............ 46
Carbon source utilization.. 46
Antibiotic resistance... 46
Plasmid and total DNA extraction and quantification.. 47
Genomic fingerprinting of *L. garvieae*.. 47
PCR amplification... 47
Southern blotting analysis and probes preparation 48
Genome analysis... 48
RNA isolation.. 48
cDNA synthesis and RT-PCR experiments... 48
RESULTS AND DISCUSSION 52
Carbon source utilization.. 52
Antibiotic resistance... 54
Genomic fingerprinting 55
Plasmids dissemination and Southern analysis 58
Resistance to sulfonamide compounds.. 62
CONCLUSIONS 67
REFERENCES.. 68
4. CHARACTERIZATION OF TEMPERATE BACTERIOPHAGES INFECTING *LACTOCOCCUS GARVIEAE* REVEALS NEW INSIGHT INTO THE EVOLUTION OF THE SPECIES 71
MATERIAL AND METHODS 72
Induction assay and transmission electron microscope...................................... 72
DNA extraction and sequencing 72
Phage genome analysis and comparison 73
RESULTS AND DISCUSSION 75
Phage induction 75
Ex novo genome sequencing of temperate bacteriophage TB25 75
Genome comparison 80
CONCLUSIONS... 91
REFERENCES.. 92

5. ISOLATION AND CHARACTERIZATION OF NEW VIRULENT PHAGES INFECTING *LACTOCOCCUS GARVIEAE* 95

MATERIALS AND METHODS 96

Bacterial host and phage isolation 96

Microbiological assays 96

Electron microscopy 97

DNA extraction and restriction profile 97

Phage DNA extraction, sequencing, and genome analysis 97

Structural protein profile 98

Proteomic tree design 98

RESULTS AND DISCUSSION 98

Phage isolation and lytic spectrum assessment 98

Morphological, genomic and structural characterization 98

Characterization of phage GE1 101

Genome of phage GE1 and its *cos* site 102

Analysis of phage GE1 genes 104

Proteome of phage GE1 104

Comparison of prolate-headed phages 108

Phylogenetic studies of *L. garvieae* phages 110

CONCLUSIONS 111

REFERENCES 112

ACKNOWLEDGEMENTS 116

LIST OF PAPER, ORAL COMMUNICATIONS AND POSTERS 117

Peer-reviewed publications 117

Posters and oral communications in scientific congress 118

ABSTRACT

This Ph.D. thesis research project aims to increase knowledge on emerging and opportunistic pathogens in food, to improve the current systems of management and control of food safety. During the project, the attention was focused on the study of *Lactococcus garvieae*, the aetiological agent of a hemorrhagic septicemia in aquaculture, but recently recognized as an opportunistic human pathogen.

In the first part of the research, strains of *L. garvieae* coming from different niches were studied through a Multi Locus Sequence Typing approach and genome analysis and comparison. Data highlighted a high level of genetic heterogeneity within the species and the separation in two main genomic lineages, with the presence of a third lineage that might be considered the ancestor branch, and the evolutionary intermediate between *Lactococcus lactis* and *L. garvieae* main clusters. Moreover, the analysis of 11 sequenced genomes of *L. garvieae* and their comparison with *L. lactis* genomic sequences provided a first insight in the core and pan-genome of *L. garvieae.* The core genome, consisting of 1341 genes, contains genes related to virulence factors; about 70% of the total core genes were also conserved in the analysed *Lactococcus* genomes, suggesting a common ancestor. A dispensable genome constitute by many genes, could explain the cosmopolitan lifestyle of *L. garvieae* species.

Subsequently to better clarify the polymorphic character of these strains, insertion sequences distribution and temperate bacteriophages were studied as mechanisms involved in genome plasticity.

Insertion sequences (ISs) were found in many analysed genomes and a substantial homology to the *Lactococcus lactis* elements was found, suggesting the movement of ISs between these two phylogenetically closely related species. Moreover, five new elements occurring in *Lactococcus garvieae* where for the first time described. All the ISs are inserted in non-coding regions and their possible involvement in chromosomal rearrangement was described.

The 50% of the 45 strains of *L. garvieae,* coming from different ecological niches, showed the presence of inducible temperate phages, belonging to *Siphoviridae* family; their differences in tail length and capside width group them in two morphotypes.

Sequencing of a prophage from a dairy *L. garvieae* strain and *in silico* analysis of fourteen sequenced *L. garvieae* strains, revealed 7 complete phages disseminated along six genomes. The phage genomes have a length on average with other lactococcal phages, ranging from 30 to 40 kb; genome comparison revealed in some case homology with other *L. lactis* phages. Interestingly only one phage genome has a %GC similar to *L. garvieae* species: the other genomes showed a GC value similar to that of *L. lactis*. These findings led us to hypothesize that the phages infecting *L. garvieae* originated from *L. lactis*, highlighting that the latter species might have appeared long time ago in the environment.

Being *Lactococcus garvieae* one of the most important pathogen in aquaculture sector, in the last part of this research, we isolated virulent phages to use in future strategies of phage therapy. Eleven lytic phages were isolated, and they showed a narrow lytic spectrum, in some case specific to few hosts. These results highlighted the importance of design a specific phage cocktail to improve the efficiency of phage therapy.

Moreover, the genome sequencing of one of these phages showed homology with other *L. lactis* phages (c2 and Q54), suggesting that phages infecting different lactococcal species may have a common ancestor.

RIASSUNTO

L'impiego di un approccio polifasico per lo studio di *Lactococcus garvieae*, un patogeno emergente.

Lo scopo di questo progetto di tesi di dottorato è quello di incrementare le attuali conoscenze sui patogeni emergenti e opportunistici isolati negli alimenti, al fine di migliorare gli attuali sistemi di gestione e controllo della salubrità dei prodotti alimentari. Durante il progetto, l'attenzione è stata posta sullo studio di *Lactococcus garvieae*, l'agente eziologico di setticemia emorragica nel settore dell'acquaculture, ma recentemente riconosciuto come patogeno emergente nell'uomo.

Nella prima parte della ricerca, differenti ceppi di *L. garvieae*, provenienti da diverse nicchie, sono stati studiati tramite esperimenti di Multi Locus Sequence Typing e mediante analisi e comparazione dei genomi.

I risultati hanno evidenziato un alto livello di eterogeneità genetica all'interno della specie e la separazione dei ceppi in due principali linee evolutive, con la presenza di una terza la quale potrebbe essere considerata la linea evolutiva ancestrale e anche evolutivamente intermedia tra la specie *L. lactis* e *L. garvieae.*

L'analisi di 11 genomi sequenziati e la loro comparazione con i genomi di *L. lactis*, ha fornito una prima indicazione sulla composizione del *core* and *pan-genome* di *L. garvieae*. Il *core genome*, costituito da 1341 geni, contiene proteine con funzioni riconducibili a fattori di virulenza; circa il 70% del totale del *core genome* è anche conservato in altri genomi del genere *Lactococcus*, suggerendo un comune ancestor. Un *dispensable genome* costituito da molti geni, potrebbe spiegare il carattere cosmopolita della specie *L. garvieae*.

In seguito, al fine di chiarire il carattere polimorfico dei ceppi, sono stati studiati la presenza e distribuzione delle sequenze di inserzione e di fagi temperati al fine di valutare il loro impatto sulla plasticità del genoma.

Le sequenze di inserzione (SI), sono state ritrovate in molti genomi analizzati e una sostanziale omologia con elementi trovati in *Lactococcus lactis* suggerisce un loro movimento tra le due specie filogeneticamente vicine. In più, 5 nuovi elementi sono stati per la prima volta trovati e descritti in *Lactococcus garvieae.* Tutte le sequenze di inserzione sono inserite in regioni non codificanti e una loro possibile implicazione in fenomeni di arrangiamento genomico è stata descritta.

Il 50% di 45 ceppi di *L. garvieae*, provenienti da diverse nicchie ecologiche hanno mostrato la presenza di fagi temperati inducibili, appartenenti alla famiglia delle *Siphoviridae*; le loro differenze in lunghezza della coda e larghezza del capside ha permesso di raggrupparli in due morfotipi.

Il sequenziamento di un fago temperato indotto da un ceppo di *L. garvieae* proveniente dal settore caseario e, l' analisi *in silico* di quattordici genomi sequenziati di ceppi *L. garvieae,* hanno evidenziato la presenza di 7 completi fagi disseminati nei genomi di sei ceppi. I genomi fagici hanno una lunghezza in media con quella di altri fagi di lattococchi, compresa tra 30 a 40 kb; la comparazione dei genomi con altri presenti in banca dati, ha rilevato in qualche caso omologia con fagi infettanti *L. lactis.* Interessante notare come un solo fago temperato ha un contenuto GC simile a quello

della specie *L. garvieae*: gli altri genomi hanno mostrato un valore di GC simile a quello di *L. lactis*. Queste scoperte ci hanno portato a ipotizzare che i fagi infettanti *L. garvieae* si siano originati da *L. lactis*, sottolineando che quest'ultima specie potrebbe essere presente nell'ambiente da più tempo.
Essendo *Lactococcus garvieae* uno dei più importanti patogeni nel settore dell'acquacultura, nell'ultima parte della ricerca si è intesi isolare fagi litici da usare in future strategie di biocontrollo, mediante l'impiego della *phage-therapy* (terapia fagica). Undici fagi litici sono stati isolati, tutti mostranti un ristretto spettro d'azione, in molti casi specifico solo verso pochi ceppi. Questi risultati hanno evidenziato l'importanza di mettere a punto specifici cocktail fagici, al fine di incrementare l'efficienza della terapia fagica.
Il sequenziamento del genoma di uno di questi fagi, ha inoltre mostrato omologie con altri fagi litici (c2 e Q54) infettanti la specie *L. lactis*, suggerendo che fagi che infettano differenti specie di lattococchi potrebbero essersi originate da un ancestor comune.

PREFACE

Food is essential for human life and its quality must be guaranteed. Even if the most has been done to ensure the safety of food, e.g. by reducing microbiological and chemical contamination, the situation is far from provide foods without risks.
The World Health Organization (WHO) has estimated that in 2005 1.8 million people died from diarrheal diseases, largely attributable to contaminated food and drinking water. This is an underestimated worldwide problem: about 76 million cases of food-borne diseases, resulting in 325.000 hospitalizations and 500 deaths, are estimated to occur each year in the United States of America (USA) alone (Mead et al., 1999).
The globalization and further liberalization of world trade, while offers many benefits and opportunities, such as a greater variety of foods, presents new safety and quality challenges. Many exporting countries, particularly developing ones, lack of an effective food control system for domestically produced food or for imported and exported foods. Producers or processors are mostly small-scale enterprises that developed informally. Distribution and consumption are largely localized, though large volumes of fresh food are traded in open markets.
In order to harmonize the situation and reduce risks, in the last years the European Union addressed several legislative measures for the implementation of food safety policies. The Regulation (EC) No. 178/2002 of the European Parliament and the Council of 28 January 2002, that laid down the general principles and requirements of food law in matters of food safety, is one of those measures.

The quality and safety of a food is strongly influenced by the presence of a specific microbial population, in particular food-associated pathogenic and spoilage bacterial species. Recent advances in the study of pathogens have contributed to assess the risk posed by these organisms in the food production chain and in many cases have led to the identification of effective control measures. On the other hand, some important mechanisms related to the pathogenic potential of these microorganisms are not known, as the real habitat adaptation capacity, the cellular answer at different cultural conditions, or the influence of the inter and intra cellular interactions on the ability to develop virulence.
The mechanisms of adaptation and virulence in bacteria are related to environmental changes of the habitat in which the microorganisms live: to remain viable, cells must adapt themselves to new conditions. Most human pathogens have two distinct phases of their life cycle, one when they are within their host and the other when they are in their natural environmental niche. For example *Vibrio cholerae*, has developed a mechanism to switch gene expression in order to overcome various stressful barriers. When cells are in the aquatic environment, they have to face fluctuations in various physicochemical parameters. When they are ingested by humans, they have to tolerate a low gastric pH, increased temperature, activities of various intestinal proteases, and other unknown intestinal factors, which profoundly influence the gene expression of the pathogen for combating such stressful stimuli (Datta and Bhadra, 2003). Changes in environmental parameters lead to changes in gene expression, with the activation of specific shock-proteins, necessary to survive in different environmental conditions.

Another problem related to safety concept is the presence and diffusion in foods of emerging or opportunistic pathogen, for which the studies related to their dissemination, impact and role are insufficient. Moreover, the increase in the incidence of horizontal gene transfer events and the detection of microbial strains with multiple resistance to antibiotics and virulence factors not typical of the species, demands an additional effort to ensure quality and healthiness of food.
Emerging pathogens are for example many *Enterococcus* species (Arias et al., 2012) and some species belonging to *Lactococcus* genus, such as *L. garvieae*, known to be a fish pathogen, but isolated in many and different foods (flour, cheese, meat and vegetables) (Ferrario et al., 2012). In recent years these species have been associated with several cases of human infections and several researchers have hypothesized the possibility of transmission of these emerging pathogens from food to humans (Aguado-Urda et a., 2013; Chan et al., 2011; Fihman et al., 2006).

Currently, new areas of growing significance have shown their potential on the development of new technologies to combat food pathogenic bacteria. The application of these new approaches to new emerging species or biovars, should enhance food quality and safety in the whole production chain.
An innovative methodology is based on use of high-throughput, non-targeted and broad scales approaches to analyse gene expression (transcriptomics), protein expression (proteomics) and metabolite content (metabolomics) to provide new insights into mechanisms, causes and effects, of microorganisms; these technics are generally called with abbreviation “omics”. (Davies H., 2010).
One of the most important implication is the ability to acquire a comprehensive understanding of biological processes to identify the various players (e.g. genes, RNA, proteins and metabolites), rather than each of those individually. Among the “omics” techniques, transcriptomics is also correlated with the study of whole genome. This approach, quickly allow to obtain more information to study biodiversity in microorganisms.
An alternative strategy to combat infectious diseases is represented by the use of phages as biocontrol agents of food pathogens (Mahony J. et al., 2011). Phages are omnipresent and are accidentally, yet regularly, consumed through ingestion of water and food. For this reason they are presumed to be safe. This, together with their specificity, makes them good tools for food safety purposes. For these reason in the last years, several researches have been carried out on phages targeting strains of pathogenic species, such as *Salmonella* spp., *Campylobacter* spp. and *Listeria* spp. This strategy, known as “phage-therapy”, could really represent an interesting area of growing significance, in the food safety and health-care sectors.

This PhD thesis reports the study of *Lactococcus garvieae*, an opportunistic human pathogen. The effective role of *L. garvieae* in food is still unclear as well as its potential pathogenicity. Applying “omics” techniques, in particular through a genomic approach, we investigated the phylogenetic history of this species and the mechanisms that promote its genome plasticity; moreover the availability of strains coming from two distinct geographical areas (Europe and America) led us to understand possible

mechanisms involved in environmental adaptation. Finally, studies on virulent phages able to infect *L. garvieae* have been carried out, in order to create alternative strategies of bio-control to replace antibiotics in treatment of *L. garvieae* infections.

STATE OF THE ART

Lactococcus garvieae

Lactococcus garvieae is the aetiological agent of lactococcosis, an acute and haemorrhagic septicaemia that affects a variety of fishes overall the world. The microorganism is a facultative anaerobic, non-motile, non-spore forming, α-hemolytic Gram positive bacterium, occurring in pairs and short chains. It is able to grow in a wide range of temperature (4-45°C. optimum 37°C), in presence of NaCl (6.5%), bile salts, at pH 9.6 and in methylene blue-milk (Vendrell et al., 2006).
Isolated for the first time in United Kingdom from mastitic cows, it has been described as *Streptococcus garvieae* and deposited as reference strain (ATCC43932) of the species; its taxonomic position changed in 1991 in *Enterococcus seriolicida* (Kusuda K et al., 1991) and in 1993 in *L. garvieae* (Prieta et al., 1993).
Generally it can be isolated from septicemic processes in fish, but the host range of *L. garvieae* is not limited to aquatic species. This emerging pathogen has been found in domestic animals such as cows, from subclinical intramammary infections (Devriese et al., 1999), cat and dog tonsils (Pot et al., 1996), bullfrog (Mendoza et al., 2012). In addition, the agent has been isolated from human infections (Fihman et al., 2006; Russo et al., 2012) as well as a component of the intestinal microbiota (Kubota et al., 2010).
Its presence has also been detected in many food matrices: one of the most important sources are artisanal cheeses made with goat and cow raw milk, where in some case *L. garvieae* represents the predominant population (Fortina et al., 2003; Foschino et al., 2006). The ubiquity lifestyle of this microorganism has also allowed its isolation from vegetables (Kawanishi et al., 2007), meat (Santos et al., 2005), cereal and cow litters (our collection).
Pathogenicity of *L. garvieae* in fish is well studied. Different authors characterized this bacterium serologically, identifying an antigen denominated KG in the cellular wall. Two antigenic types of this pathogen, KG^+ and KG^- were found, with the latter possessing a specific envelope-like substance that inhibits agglutination with specific antibodies (Kitao et al., 1983).
Several experiments have been performed in order to determine the possible correlation between pathogenicity of *L. garvieae* isolated from diseased fish and the two antigenic types (KG^- and KG^+). The presence of an envelope-like substance (linked to KG^-) is typical of more virulent strains than non-capsulated one (Barnes et al., 2002). Controversial results were obtained for strains isolated from human with endocarditis: in fact they seemed able to show pathogenicity even though the analysis of their genomes did not show the presence of genes coding for extracellular envelope. Up to now, a capsule gene cluster composed of 15 genes has been only found in one fish-isolated virulent strain (Morita et al., 2011). Moreover, factors and environmental stresses that can influence the synthesis of this capsule and, more in general, the expression of its pathogenicity is still unknown.
A first overview was given by Aguado-Urda et al., 2013: the authors analysed the global transcriptome of two *L. garvieae* strains coming from fish and humans, in

response to different temperatures (18°C and 37°C, corresponding to temperatures at which fish lactococcosis occurs and to physiological temperature in humans, respectively). The data obtained revealed that, in general, at 18°C the response was more evident and the fish isolate up-regulates several genes encoding different cold-shock and cold-induced proteins involved in an efficient adaptive response to low-temperature conditions. These findings show the adaptability of these microorganisms in facing stress, but only few genes, potentially involved in Gram-positive pathogenicity, such as autolytic enzymes and manganese transporters, are described to be up-regulated.

Insertion sequences and temperate bacteriophages as promoter of genome plasticity

Ubiquitous microorganisms are able to survive in different niches principally due to their genomic plasticity. The adaptation to new environments is often accompanied by an acquisition of additional genes and/or genome rearrangements (Caufield et al., 2015; Makarova et al., 2006). Two of the most important mechanisms involved in genome plasticity are related to the presence of insertion sequences and temperate bacteriophages.

Insertion sequences (ISs) are the shortest autonomously Mobile Genetic Elements (MGE). They consist of a unique functional module which catalyses transposition, and its cognate sites of action, the element ends, known as Inverted Repeat sequences (IRs). IRs serve as the sites for recognition and cleavage by transposase in transposition reaction by a cut-and-paste mechanism, usually generating short Direct Repeat sequences (DRs) of the target DNA at the site of insertion (Mahillon and Chandler, 1998; Siguier et al., 2014). ISs participate in chromosome rearrangements, plasmid integration and activation of gene expression (Siguier et al., 2014).

The distribution and abundance of insertion sequences among bacterial genomes are highly variable: many strains of the same species can either contain multiple IS copies or be completely devoid of ISs. Even if many studies have been carried out, the mechanisms related to the enormous variation in genomic abundance and distribution of IS elements still remain poorly understood. IS abundance can be related with the frequency of horizontal gene transfer (HGT), genome size, pathogenicity and human association (Wagner, 2006; Touchon and Rocha, 2007). It is also known that IS can undergo massive expansion and loss; IS expansion has been commonly observed in bacteria with host-restricted lifestyles and has been considered an early step in the genome reduction (Siguier et al., 2014).

Bacteriophages are viruses that infect bacteria and they are classified in two categories based on their lifecycle: the virulent phages, only able to replicate through a lytic cycle, killing the cell by lysis, and the temperate (or lysogenic) phages which are able to integrate in the genome and to be replicated themselves together with the genome of the cells.

Once the genome of a temperate phages is injected into the cytoplasm of the host, depending on the metabolic state of the cell, the phage can switch its genes into lysogenic state instead of follow the lytic lifestyle. The lysogenic lifestyle will be

maintained until stress conditions will cause the activation of the transcription of lytic genes.
Recently, the number of studies regarding the mechanisms of interactions between phages (both lytic and temperate ones) and bacteria is growing. Lytic phages can be used in order to substitute the antibiotics to prevent spreading of pathogen microorganisms (Bakhshinejad 2014; Sulakvelidze et al., 2001), but also a deeper understanding of resistance mechanisms of bacteria to bacteriophages is important to prevent fermentation failure in food industries (Brüssow, 2001; Deveau et al., 2006).
In contrast, studies on temperate phages are more focused on the ability of phages to increase genome plasticity of their hosts, contributing to evolution and adaptation of cells to different environments. Generally, the integration of a temperate phage is responsible of several changes in the host behaviour, among which autoimmunity against infections by the same or closely related phages, disruption of bacterial genes during the integration or over-expression of bacterial genes through the promoter region of the phage.
In many cases temperate phages carry autoreplicative elements-called “morons”- codifying for toxins, regulatory and effector proteins, adhesins, exonucleases and superantigens. These sequences are flanked by a specific transcriptional promoter and a terminator, allowing the expression of gene(s) also during lysogenic cycle (Fortier and Sekulovic, 2013).
To understand the strong impact of morons on genome plasticity, it is worth noting the so called “lysogenic conversion”, where a non-pathogenic strain is converted to pathogen one by the integration of temperate phages carrying toxin or virulent genes. It happened for *E. coli* O157:H7 where new clones having two Shiga toxin-encoding prophages (Sp5 and Sp15) emerged (Hayashi et al., 2001) and for *Vibrio cholera* that acquired the cholera toxin by a filamentous phage (Waldor and Mekalanos, 1996).
The lack of information in literature about these evolutionary mechanisms in *L. garvieae* led us to better investigate their role on *L. garvieae* evolution and adaptability.

Biocontrol of *Lactococcus garvieae*

Antibiotics have been largely used to prevent lactococcosis both in animal and human; in general in fish outbreaks, the most used antibiotics are erythromycin, oxytetracycline, amoxicillin and low-level doxycycline (Vendrell et al., 2006), but different studies revealed that strains coming from different geographical areas have a specific resistance/sensibility pattern. In fact Ravelo et al., (2001) studied strains susceptible to enrofloxacyn and nitrofurantoin and resistant to oxolinic acid and sulphametoxazol-trimethoprim, while Diler et al. (2002), analysing Turkey outbreaks, demonstrated that the isolated were sensitive to erythromycin, ofloxacin, ampicillin and chloramphenicol, but resistant to penicillin and clindamycin.
Human infections, such as endocarditis and spondylodiscitis are cured differently. In Li et al., (2008) penicillin and gentamicin were given for 30 days to a patient affected by endocarditis; for the same pathology Fihman et al. (2006) described a treatment with

amoxicillin (150 mg/kg/day) and gentamicin (2 mg/kg/day), while for spondylodiscitis Chan et al., (2011) described a treatment with ampicillin for 6 weeks.
All together these results highlight, once again, the variability characterizing the species.
Nowadays, the increase of antibiotic resistant strains as a consequence of an indiscriminate use of antibiotics, has led to focus attention in the search for new alternative strategies to prevent pathogen growth.
Regarding *L. garvieae* infections, sanitary measures are the first barrier to prevent the introduction of pathogens in the fish farm: periodic cleaning of the tanks, disinfection of all utensils, reduction of fish manipulation, elimination of dead or diseased fish and maintenance of low densities of the culture, constitute the most important measures.
Several vaccines against *L. garvieae* have been developed, autovaccines (constituted by inactivated microorganisms with formalin) and adjuvanted vaccines. Vaccination protocol consists of an intraperitoneal vaccination one month before water temperature increase over 15°C (Aizpurúa et al., 1999), or an oral vaccination with capsulate and non-capsulated antigens. The use of alginate microparticles as protector was also studied by Romalde et al., (2004) and the authors highlighted a high level of protection even though the primary immunization did not guarantee complete protection.
Bacteriocins are peptides with antimicrobial properties which play an important role in bacteria competition. In *L. garvieae* two different bacteriocin-producer strains were described: *GarML* from a Mallard duck (Gabrielsen et al., 2014) and *GarA* from human clinical isolates (Maldonado-Barragán et al., 2013). These peptides act, against other *L. garvieae* strains, inhibiting the septum formation, and for these reasons they could be used in the prevention or treatment of infections caused by this pathogen.
Moreover, also the bacteriocin nisin Z from *L. lactis* TW34, inhibiting the growth of *L. garvieae*, can be considered as a new alternative strategies to antibiotics (Sequeiros et al., 2015).

Phage therapy

Recently, the problem of antimicrobial resistance strains in the treatment of bacterial infections can be circumvented by the use of phages, in the so called phage therapy. To be used in phage therapy, bacteriophages must be strictly lytic, must have a large spectrum of action and no virulent genes in their genome. Moreover, their viability and the ways of administration should take into consideration when a phage cocktail is designed.
For what concern *Lactococcus garvieae* only a few bacteriophages have been isolated and described by Park et al., respectively in 1997 and 1998; all of them were isolated from diseased fish and showed a broad range of activity against *L. garvieae* strains.
Moreover, Nakai et al., (1999) compared the action of different bacteriophages administrated differently: by an intraperitoneal injection or added in the feed. The results obtained, although preliminary, showed that both methods could prevent *L. garvieae* infections, suggesting a potential use of the phage therapy for controlling the disease.
Up to now, only one *L. garvieae* virulent phage has been investigated at the genomic level. Ghasemi at al., (2014) described in detail a new phage infecting *L. garvieae*: it

was named WP-2, belongs to the family *Podoviridae*, one of the three families of the Caudovirales order, of which lactococcal phages are members, which comprises of: *Siphoviridae* (long, non-contractile tails), *Myoviridae* (long, contractile tails) and *Podoviridae* (short tails) (Veesler and Cambillau, 2001). To develop efficient phage control tools, new phages and new data are needed.

AIM OF THE STUDY

This Ph.D. thesis research project aims to improve knowledge on emerging pathogens. *Lactococcus garvieae*, known to be a fish pathogen and recently recognized as an opportunistic human pathogen, represents a valid example of emerging pathogens in food.

A polyphasic approach was carried out with the aim of studying:

- the phylogenetic relationship existing between *L. lactis* and *L. garvieae,* through genomes analysis and comparison;
- the presence and distribution along the genome of *L. garvieae* of insertion sequences and temperate bacteriophages, two of the most important mechanisms involved in genome plasticity;
- a phenotypic and genotypic comparison between two geographically (European and American) distinct groups of *L. garvieae* strains, focusing the attention on antibiotic-resistance profile;
- isolation and characterization of lytic phages, for a their future use in phage therapy strategies.

REFERENCES

1. Aguado-Urda M., Gibello A., Blanco Mdel M., Fernández-Garayzábal J.F., López-Alonso V., López-Campos G.H. (2013). *Global transcriptome analysis of* Lactococcus garvieae *strains in response to temperature*. PLoS One. 8(11):e79692.

2. Aizpurúa J., Esnault F. Revista de Acuicultura Trouvit Informa. Invierno 1999/2000. 21-3.

3. Arias C.A., Murray B.E. (2012). *The rise of the* Enterococcus*: beyond vancomycin resistance*. Nat Rev Microbiol. 10:266-278.

4. Bakhshinejad B. (2014). *Bacteriophages and their applications in the diagnosis and treatment of hepatitis B virus infection*. World J Gastroenterol. 20:11671-11683.

5. Barnes A.C., Guyot C., Hanse B.G., Mackenzie K., Horn M.T., Ellis A.E. (2002). *Resistance to serum killing may contribute to differences in the abilities of capsulate and non-capsulated isolates of* Lactococcus garvieae *to cause disease in rainbow trout (*Oncorhynchus mykiss L.*)*. Fish Shellfish Immunol. 12:155-168.

6. Brüssow H. (2001). *Phages of dairy bacteria*. Ann Rev Microbiol. 55: 283–303.

7. Caufield P.W., Schön C.N., Saraithong P., Li Y., Argimón S. (2015). *Oral lactobacilli and dental caries: a model for niche adaptation in humans.* J Dent Res. pii: 0022034515576052.

8. Chan J.F., Woo P.C., Teng J.L., Lau S.K., Leung S.S., Tam F.C., Yuen K.Y. (2011). *Primary infective spondylodiscitis caused by* Lactococcus garvieae *and a review of human* L. garvieae *infections*. Infection. 39:259-64.

9. Datta P.P., Bhadra R.K. (2003). *Cold shock response and major cold shock proteins of* Vibrio cholerae. Appl Environ Microbiol. 69:6361-6369.

10. Davies H. (2010). *A role for "omics" technologies in food safety assessment*. Food Control. 21:1601-1610.

11. Deveau H., Labrie S.J., Chopin M.C., Moineau S. (2006). *Biodiversity and classification of lactococcal phages*. Appl Environ Microbiol 72:4338–4346.

12. Devriese L.A., Hommez J., Laevens H., Pot B., Vandamme P., Haesebrouck F. (1999) *Identification of aesculin-hydrolyzing streptococci, lactococci, aerococci and enterococci from subclinical intramammary infections in dairy cows.* Vet Microbiol. 70:87-94.

13. Diler O., Altun S., Adiloglu A.K., Kubilay A., Isikli B. (2002). *First occurrence of Streptococcosis affecting farmed rainbow trout (*Oncorhynchus mykiss*) in Turkey.* Bull Eur Ass Fish Pahol. 22:21-25.

14. Ferrario C., Ricci G., Borgo F., Rollando A., Fortina M.G. (2012). *Genetic investigation within* Lactococcus garvieae *revealed two genomic lineages*. FEMS Microbiol Lett. 332:153-161.

15. Fihman V., Raskine L., Barrou Z., Kiffel C., Riahi J., Berçot B., Sanson-Le Pors M.J. (2006). Lactococcus garvieae *endocarditis: identification by 16S rRNA and sodA sequence analysis*. J Infect. 52(1):e3-6.

16. Fortier L.C., Sekulovic O. (2013). *Importance of prophages to evolution and virulence of bacterial pathogens*. Virulence. 4:354–365.

17. Fortina M.G., Ricci G., Acquati A., Zeppa G. Gandini A., Manachini P.L. (2003). *Genetic characterization of some lactic acid bacteria occurring in an artisanal protected denomination origin (PDO) Italian cheese, the Toma piemontese.* Food Microbiol. 20:397-404.

18. Foschino R., Picozzi C., Borghi M., Cerliani M.C., Cresci E. (2006). *Investigation on the microflora of Caprino Lombardo cheese from raw goat milk.* Ital J Food Sci. 18:33-49.

19. Gabrielsen C., Brede D.A., Salehian Z., Nes I.F., Diep D.B. (2014). *Functional genetic analysis of the GarML gene cluster in* Lactococcus garvieae *DCC43 gives new insights into circular bacteriocin biosynthesis.* J Bacteriol. 196:911-919.

20. Ghasemi S.M., Bouzari M., Yoon B.H., Chang H.I. (2014). *Comparative genomic analysis of* Lactococcus garvieae *phage WP-2, a new member of* Picovirinae *subfamily of* Podoviridae. Gene. 551:222-229.

21. Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H. (2001). *Complete genome sequence of enterohemorrhagic* Escherichia Coli *O157:H7 and genomic comparison with a laboratory strain K-12.* DNA Res. 8: 11–22.

22. Kawanishi M., Yoshida T., Kijima M., Yagyu K., Nakai T., Okada S., Endo A., Murakami M., Suzuki S., Morita H. (2007). *Characterization of* Lactococcus garvieae *isolated from radish and broccoli sprouts that exhibited a KG+ phenotype, lack of virulence and absence of a capsule*. Lett Appl Microbiol. 44:481-487.

23. Kitao T. (1983). *Strain variation associated with pathogenesis of* Streptococcus *in cultured yewllowtail (*Seriola quinqueradiata*)*. Proceeding of the second North Pacific Aquaculture Symposium, Tokay University, Tokyo, 231-242.

24. Kubota H., Tsuji H., Matsuda K., Kurakawa T., Asahara T., Nomoto K. (2010). *Detection of human intestinal catalase-negative, Gram-positive cocci by rRNA-targeted reverse transcription-PCR.* Appl Environ Microbiol. 76:5440-5451.

25. Kusuda R., Kawai K., Salati F., Banner C. R., Fryer J. L. (1991). Enterococcus seriolicida *sp. nov., a fish pathogen.* Int J Syst Evol Microbiol. 41:406-409.

21. Li W.K., Chen Y.S., Wann S.R., Liu Y.C., Tsai H.C. (2008). Lactococcus garvieae *endocarditis with initial presentation of acute cerebral infarction in a healthy immunocompetent man.* Intern Med. 47:1143-1146.

26. Mahillon J., Chandler M. (1998). *Insertion sequences*. Microbiol Mol Biol Rev. 62:725-774.

27. Mahony J., McAuliffe O., Ross R.P., van Sinderen D. (2011). *Bacteriophages as biocontrol agents of food pathogens.* Curr Opin Biotechnol. 22:157-163.

28. Makarova K., Slesarev A., Wolf Y., Sorokin A., et al. (2006). *Comparative genomics of the lactic acid bacteria*. Proc Natl Acad Sci USA. 103:15611-15616.

29. Maldonado-Barragán A., Cárdenas N., Martínez B., Ruiz-Barba J.L., Fernández-Garayzábal J.F., Rodríguez J.M., Gibello A. (2013). *Garvicin A, a novel class IId bacteriocin from* Lactococcus garvieae that inhibits septum formation in L. garvieae *strains.* Appl Environ Microbiol. 79:4336-4346.

30. Mead P.S., Slutsker L., Dietz V., McCaig L.F., Bresee J.S., Shapiro C., Griffin P.M., Tauxe R.V. (1999). *Food-related illness and death in the United States*. Emerging Infect Dis. 5:607-625.

31. Mendoza G.M., Pasteris S.E., Ale C.E., Otero M.C., Bühler M.I., Nader-Macías M.E. (2012). *Cultivable microbiota of* Lithobates catesbeianus *and advances in the selection of lactic acid bacteria as biological control agents in raniculture.* Res Vet Sci. 1160-1167.

32. Morita H., Toh H., Oshima K., Yoshizaki M., Kawanishi M., Nakaya K., Suzuki T., Miyauchi E., Ishii Y., Tanabe S., Murakami M., Hattori M. (2011). *Complete genome sequence and comparative analysis of the fish pathogen* Lactococcus garvieae. PLoS One 6:e23184.

33. Nakai T., Sugimoto R., Park K.H., Matsuoka S., Mori K., Nishioka T., Maruyama K. (1999). *Protective effects of bacteriophage on experimental* Lactococcus garvieae *infection in yellowtail*. Dis Aquat Organ. 37:33-41.

34. Park K.H., Matsuoka S., Nakai T., Muroga K. (1997). *A virulent bacteriophage of* L. garvieae *(formerly* Enterococcus seriolicida*) isolated from yellowtail* Seriola quinqueradiata. Dis Aquat Org. 29:145–149.

35. Park K.H., Kato H., Nakai T., Muroga K. *Phage typing of* Lactococcus garvieae *(formerly* Enterococcus seriolicida*) a pathogen of cultured yellowtail*. Fish Sci. 64:62–64.

36. Pot B., Devriese L.A., Ursi D., Vandamme P., Haesebrouck F., Kersters K. (1996). *Phenotypic identification and differentiation of* Lactococcus *strains isolated from animals.* Syst Appl Microbiol. 19:213–222.

37. Prieta J., Doménech A.M., Fernández-Garaizábal J.F., Collins M.D., Rodrígues U.M., Jones D., et al. (1993). *Lactococcosis de la trucha arco iris (*Oncorhynchus mykiss*)*. Med Vet. 10:367-373.

38. Ravelo C., Magariños B., Romalde J.L., Toranzo A.E. (2001). *Conventional versus miniaturizad systems for the phenotypic characterization of* Lactococcus garvieae *strains*. Bull Eur Ass Fish Pathol. 21:136-144.

39. Romalde J.L., Luzardo-Alvarez A., Ravelo C., Toanzo A.E., BLanco-Mendez J. (2004). *Oral immunization using alginate microparticles as a useful strategy for booster vaccination against fish lactococcosis*. Acquaculture 236:119-29.

40. Russo G., Iannetta M., D'Abramo A., Mascellino M.T., Pantosti A. Erario L., Tebano G., Oliva A., D'Agostino C., Trinchieri V., Vullo V. (2012). Lactococcus garvieae *endocarditis in a patient with colonic diverticulosis: first case report in Italy and review of the literature.* New Microbiol. 35:495-501.

41. Santos E.M., Jaime I., Rovira J., Lyhs U., Korkeala H., Björkroth J. (2005). *Characterization and identification of lactic acid bacteria in "morcilla de Burgos"*. Int J Food Microbiol. 97:285-296.

42. Sequeiros C., Garcés M.E., Vallejo M., Marguet E.R., Olivera N.L. (2015). *Potential aquaculture probiont* Lactococcus lactis *TW34 produces nisin Z and inhibits the fish pathogen* Lactococcus garvieae. Arch Microbiol. 197:449-458.

43. Siguier P., Gourbeyre E., Chandler M. (2014). *Bacterial insertion sequences: their genomic impact and diversity*. FEMS Microbiol Rev. 38:865-891.

44. Sulakvelidze A., Alavidze Z., Morris J.G.. (2001). *Bacteriophage therapy*. Antimicrob Agents and Chemother. 45:649–659.

45. Touchon M., Rocha E.P.C. (2007). *Causes of insertion sequences abundance in prokaryotic genomes*. Mol Biol Evol. 24: 969-981.

46. Veesler D., Cambillau C. (2011). *A common evolutionary origin for tailed-*bacteriophage *functional modules and bacterial machineries*. Microbiol Mol Biol Rev. 75:423-433.

47. Vendrell D., Balcázar J.L., Ruiz-Zarzuela I., de Blas I., Gironés O., Múzquiz J.L. (2006). Lactococcus garvieae *in fish: a review*. Comp Immunol Microbiol Infect Dis. 29:177-198.

48. Wagner A. (2006). *Period extinction of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes*. Mol Biol Evol. 23:723-733.

49. Waldor M.K., Mekalanos J.J. (1996). *Lysogenic conversion by a filamentous phage encoding cholera toxin*. Science. 272:1910–1914.

1. PHYLOGENETIC RELATIONSHIPS BETWEEN *LACTOCOCCUS GARVIEAE* AND *LACTOCOCCUS LACTIS*

1. Phylogenetic relationships between *Lactococcus garvieae* and *Lactococcus lactis*

The population structure and diversity of *Lactococcus garvieae* was determined at both gene and genome level. Selected lactococcal isolates of various origins were analyzed by a multi locus sequence typing (MLST) and then the intra species diversity was evaluated through genome analysis and comparison.

MATERIALS AND METHODS

Lactococcus garvieae strains

Lactococcus garvieae strains tested comprise four strains isolated from diseased fish (Lg9; Lg19; V63, V79), four strains isolated from dairy products (G27, G07, TB25, G01), five strains isolated from meat and meat products (Smp3, Po1, Tac2, Bov3, I113), four from vegetables (Ins1, Sed2, Br3, Br4), one from cereals (Far1) and the type strain of the species DSM20684^T. For four of these strains, the whole genome sequence was previously obtained (TB25 - accession number NZ_AGQX00000000; Lg9 - NZ_AGQY00000000; I113 - NZ_AMFD00000000; Tac2 - NZ_AMFE00000000) (Ricci et al., 2012; Ricci et al., 2013). The strains were grown in M17 broth (Difco, Detroit, USA) supplemented with 10 g L^{-1} glucose (M17-G) at 37°C for 24 h. Stock cultures were maintained at -80°C in M17-G with 15% glycerol.

DNA extraction

DNA was extracted as previously described (Ferrario et al., 2012), starting from 100 µL of M17-G broth culture. The concentration and purity of the DNAs were determined with a UV-Vis spectrophotometer (SmartSpecTM Plus, Biorad, Milan, Italy). Nucleo Spin Extract II (Macherey-Nagel GmbH, Düren, Germany) was used to purify PCR products that were sequenced using the dideoxy chain-termination principle (Sanger et al., 1977), employing ABI Prism Big Dye Terminator Kit (Applied Biosystems, Foster City, CA). The reaction products were analysed with the ABI PrismTM310 DNA Sequencer. The database searches were performed by using the basic local alignment tool (BLAST) programs (Altschul et al., 1990) from the National Center for Biotechnology Information website. The phylogenetic tree was constructed using the UPGMA method (Sneath et al., 2011).

Multi Locus Sequence Typing (MLST)

Lactococcus garvieae strains were sequence typed using seven housekeeping genes (*als*, *atpA*, *tuf*, *gapC*, *gyrB*, *rpoC*, and *galP*). The oligonucleotide primers, designed to conserved regions of the selected genes, conditions used and their amplification products are listed in Ferrario et al., 2013. Amplicons were gel purified, sequenced and analysed as reported in the previous section.
Forward and reverse DNA sequences obtained from PCR amplification were trimmed and studied in comparison with sequences from *L. garvieae* genomes deposited in database (strain 8831 - accession number NZ_AFCD00000000; 21881 - NZ_AFCF00000000; ATCC 49156 - NC_015930; LG2 - NC_017490; UNIUD074 -

NZ_AFHF0000000; IPLA 31405 - NZ_AKFO00000000; DCC43 - NZ_AMQS00000000). The most polymorphic regions of 800-850 bp, were selected and then analyzed using MEGA v5 (Takamura et al., 2011). Isolate dataset creation and allele assignation was done using PubMLST.org web tools (http://pubmlst.org/analysis/). Each unique allelic profile, as defined by the allele numbers of the seven loci, was assigned a Sequence Type (ST) number. The same ST number was used for more than one strain if they shared the same allelic profile.
For phylogenetic analysis, concatenated sequences were aligned and analysed with MEGA v5 and the phylogenetic tree was constructed using the neighbor-joining method.
Strains relationships were analyzed using eBURST (Feil et al.,2004) to identify potential Clonal Complexes (CCs), with the default stringent (conservative) definition. To investigate the population structure, the Clonal Frame method was used (Vos and Didelot, 2009). The recombination to mutation ratio (r/m) was calculated as reported by Vos and Didelot. For each dataset, two runs of the Clonal Frame MCMC were performed each consisting of 200,000 iterations. The first half of the chains was discarded, and the second half was sampled every hundred iterations. Split decomposition trees were constructed with 1,000 bootstraps replicates based on parsimony splits as implemented in SplitsTree v4.1 (Huson et al., 2006).

Genome analysis and comparison

Each predicted proteome of the analysed strains (table 1), was searched for orthologues against the total proteome, where orthology between two proteins was defined as the best bidirectional FASTA hits (Pearson, 2000). Identification of orthologues, paralogues, and unique genes was performed following a preliminary step consisting of the comparison of each protein against all other proteins using BLAST analysis (Altschul et al., 1990) (cutoff: E value of 1×10^{-4} and 40% identity over at least 50% of both protein sequences), and then all proteins were clustered into COGs (Clusters of Orthologous Genes) using MCL (graph theory-based Markov clustering algorithm) (van Dongen, 2000).
Following this, the unique COGs where classified by selecting the clusters with members from only one of the *Lactococcus* genomes analysed. COGs shared between all genomes, named core COGs, were defined by selecting the clusters that contained at least one single protein member for each genome. COGs attribution to a specific COG family was made by BLASTp search against the COGs database (http://www.ncbi.nlm.nih.gov/COG/).
In order to provide a highly reliable evolutionary reconstruction, a concatenated protein sequence that includes the product of each core gene from every genome was used to build a *Lactococcus* supertree. Alignment was done using CLUSTAL OMEGA (Sievers et al., 2011), and phylogenetic trees were constructed using the Neighbor joining in PhyML (Guidon et al., 2003). The supertree was visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).
For all genomes used in this study, a pan-genome calculation was performed using the PGAP pipeline (Zhao et al., 2012); the ORF content of each genome was organized in functional gene clusters using the gene family (GF) method. A pan-genome profile and

a core-genome profile were built using all possible BLAST combinations for each genome being sequentially added. Finally, using the pan-genome profile of shared orthologues between the *Lactococcus (garvieae)* genomes, a pan-genome tree was constructed. This tree was visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS AND DISCUSSION

Multilocus Sequence Typing (MLST)

Lactococcus garvieae strains were selected from a greater strain collection previously explored through different genotyping methods (Ferrario et al., 2012) and chosen as representative of the isolation niche and of the different genomic lineages individuated (table 2). They were subjected to a MLST scheme that targeted seven housekeeping genes unlinked, possessing the appropriate levels of sequence diversity, and lacking insertions or deletions that could cause changes in length. The MLST scheme developed in this study was designed to be technically robust, generating high amplicon yields for all genotypes, under the same PCR conditions for all seven loci. MLST analysis of the 26 strains tested has identified 18 different Sequence Types (STs), highlighting a significant heterogeneity in the strain collection. All loci were polymorphic (table 2) and the number of allele assignation varied between eight in *gap*C, the most conserved locus, and 14 in *rpo*C, suggesting a different evolution rate of the different loci, equally distributed along the genome sequences (the minimum distance among the loci was 18 kb). The analysis of allelic profiles highlighted a first relationship among strains. Through the eBURST algorithm that defines Clonal Complexes (CCs) by single-locus variants, three CCs, in which 50% of the strains were distributed, were identified (table 2) indicating a high genotype frequency. CC1 included seven strains grouped in ST3, ST4, and ST13. CC2 grouped ST16 and ST17, corresponding to strains ATCC49156 and LG2. CC3 grouped four strains belonging to ST10 and ST11. As shown, the CCs obtained were not homogeneous with respect to the niche of isolation.

The phylogeny of the 26 *L. garvieae* strains was analysed by constructing a neighbor-joining tree from the 5,713 bp concatenated sequence of the seven loci (figure 1). The tree revealed the presence of two main subgroups, as highlighted in previous work (Ferrario et al., 2012).

Subgroup S_A consisted of strains included in CC1 and CC2 plus three strains with unique STs. Subgroup S_B included strains of CC3 plus eight strains with six different STs. Moreover, in this analysis we found that strain I113 and, particularly, strain DCC43 were the most differentiated of the studied samples, and clustered in independent branches. The Clonal Frame analysis suggests that the two main subgroups appeared at approximately the same time, while ungrouped strains seem to represent the ancestors from which S_A and S_B originate with strain DCC43 that first emerged.

1. Phylogenetic relationships between *Lactococcus garvieae* and *Lactococcus lactis*

The r/m ratio (ratio of probabilities that a given site is altered through recombination and mutation) was calculated for the entire population and for the two main subgroups, to evaluate whether the high genotypic diversity could be due to recombination events. The r/m was 0.978 for the total population, 0.925 for S_A and 1.203 for S_B. These data could indicate distinct inclinations and adaptive abilities to environments of the two subgroups: S_B seems to respond to selective pressure increasing the recombination rate. Recombination events in *L. garvieae* population were also investigated using the SplitsTree program, with the splits decomposition methods on the concatenated sequence of the total population, and for subgroups (figure 2). Interconnected network of phylogenetic relationships, resembling a parallelogram in shape, was observed; also in this case, for members of the subgroup S_B, a major recombinational effect could be highlighted. The tree revealed four major branches, two corresponding to subgroups S_A and S_B and two longer branches, one harboring I113 strain and the other DCC43 strain. The same analysis was also performed using *L. lactis* subsp. *lactis* IL1403 (accession number AE005176), and *L. lactis* subsp. *cremoris* MG1363 (AM406671), phylogenetically related (Morita et al., 2011). The split graph showed the same subdivision of *L. garvieae* population, with the strain DCC43, interconnected with *L. lactis* species by a recombinational event (figure 3).

Table 1**.** General genome features of the *L. garvieae* strains, in comparison with genome features of *L. lactis* and *L. raffinolactis* strains

Lactococcus species	Strain	Genome Size (bp)	GC content (%)	Genome Status	No of protein-coding genes	Average length of protein-coding genes	Protein-coding genes region %	Intergenic region %	Accession Number
L. garvieae	8831	2,085,932	38.0	draft	2016	919	88.9	11.1	NZ_AFCD00000000
	21881	2,164,301	37.9	draft	2216	858	87.9	12.1	NZ_AFCC00000000
	ATCC 49156	1,950,135	38.8	complete	1947	874	87.2	12.8	NC_015930
	DCC43	2,244,387	37.7	draft	2227	882	87.6	12.4	NZ_AMQS00000000
	I113	2,178,733	37.9	draft	2159	893	88.5	11.5	NZ_AMFD00000000
	IPLA 31405	2,052,312	38.5	draft	1778	939	81.4	18.6	NZ_AKFO00000000
	LG2	1,963,964	38.8	complete	1968	871	87.3	12.7	NC_017490
	LG9	2,087,705	38.5	draft	2092	881	88.2	11.8	NZ_AGQY00000000
	Tac2	2,242,863	38.2	draft	2210	896	88.3	11.7	NZ_AMFE00000000
	TB25	2,014,328	38.1	draft	2024	879	88.3	11.7	NZ_AGQX00000000
	UNIUD074	2,171,472	38.7	draft	2146	871	86	14	NZ_AFHF00000000
L. lactis subsp. *cremor*is	A76	2,452,616	35.9	complete	2643	790	85.2	14.8	NC_017492
	MG1363	2,529,478	35.7	complete	2434	856	82.3	17.7	NC_009004
	NZ9000	2,530,294	35.7	complete	2510	837	83	17	NC_017949
L. lactis subsp. *lactis*	Il1403	2,365,589	35.3	complete	2266	883	84.6	15.4	NC_002662
	KF147	2,598,144	34.9	complete	2444	899	84.6	15.4	NC_013656
L. raffinolactis	4877	2,280,761	38.6	draft	2359	803	83.1	16.9	CALL01000000

Table 2. *L. garvieae* strains analyzed and allelic profiles.

Strains	Source	Origin	Year of isolation	Sub-groups	Allele *als*	*atp*A	*tuf*	*gap*C	*gyr*B	*rpo*C	*gal*P	ST	Clonal Complexes (CCs)
DSM20684[T]	Bovine mastitis	UK	1984	B	1	1	1	1	1	1	1	1	Singleton
Smp3	Meat products	Italy	2009	B	2	2	2	2	2	2	2	2	Singleton
Po1	Poultry	Italy	2009	B	3	3	3	2	3	3	3	3	1
Tac2	Turkey	Italy	2009	B	3	3	3	2	3	3	3	3	1
Bov3	Beef	Italy	2009	B	3	3	4	2	3	3	3	4	1
I113	Meat products	Italy	2005	/	4	4	5	3	4	4	4	5	Singleton
Ins1	Salad	Italy	2009	B	5	5	6	2	5	5	5	6	Singleton
Sed2	Celery	Italy	2009	A	6	6	7	2	6	6	6	7	Singleton
Br3	Broccoli	Italy	2009	A	7	6	8	2	7	7	7	8	Singleton
Br4	Broccoli	Italy	2009	A	7	6	8	2	7	7	7	8	Singleton
Far1	Wheat flour	Italy	2006	A	8	7	9	2	8	8	8	9	Singleton
G27	Cow milk	Italy	2003	A	9	7	3	4	7	9	9	10	3
G07	Cow cheese	Italy	2003	A	9	7	3	4	7	9	10	11	3
TB25	Cow cheese	Italy	2001	A	9	7	3	2	7	10	9	12	Singleton
G01	Cow cheese	Italy	2009	A	9	7	3	4	7	9	9	10	3
Lg9	Rainbow trout	Italy	2000	B	10	3	4	2	3	3	3	13	1
Lg19	Rainbow trout	Italy	2000	B	10	3	4	2	3	3	3	13	1
V63	Trout	Italy	2003	A	11	6	10	2	9	11	8	14	Singleton
V79	Trout	Italy	2002	A	11	6	11	5	7	12	11	15	Singleton
8831	Rainbow trout	Spain	/	A	11	6	10	2	9	11	8	14	Singleton
21881	Human blood	Spain	/	A	9	7	3	4	7	9	9	10	3
ATCC49156	Yellowtail	Japan	1974	B	12	8	6	6	10	13	12	16	2
LG2	Yellowtail	Japan	2002	B	12	8	6	7	10	13	12	17	2
UNIUD074	Rainbow trout	Italy	/	B	10	3	4	2	3	3	3	13	1
IPLA31405	Cow milk	Spain	2008	B	3	3	4	2	3	3	3	4	1
DCC43	Mallard duck intestines	/	/	/	13	9	12	8	11	14	13	18	Singleton

ST: Sequence Type
/ : Unknown

1. Phylogenetic relationships between *Lactococcus garvieae* and *Lactococcus lactis*

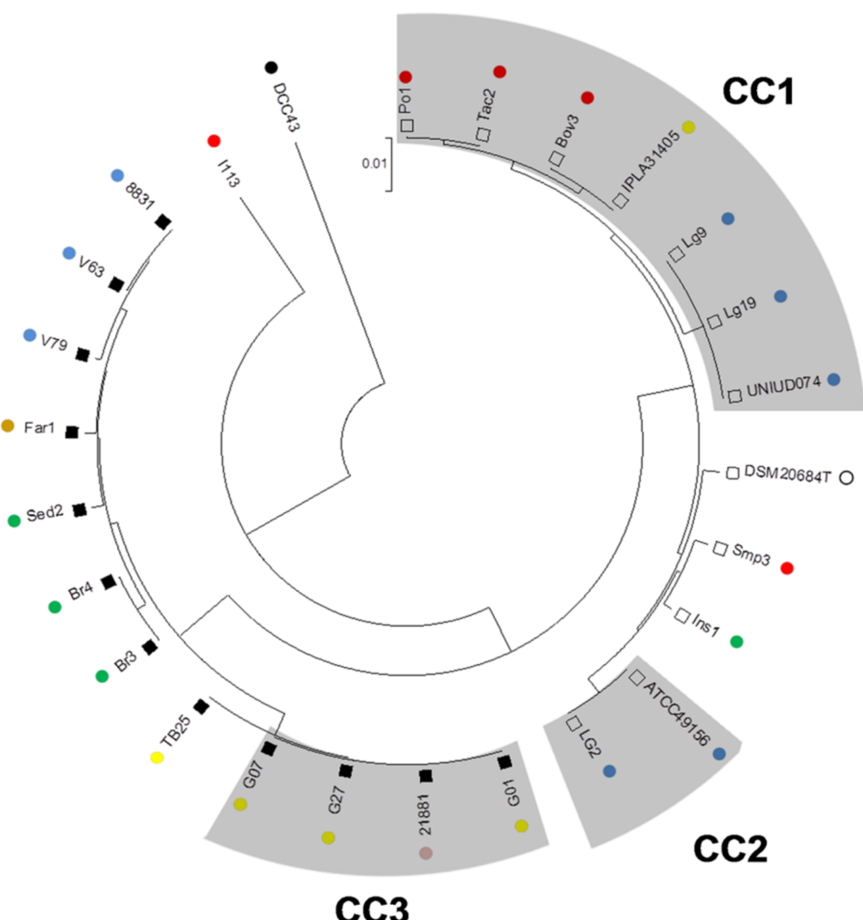


Figure 1. Phylogenetic relationships between *L. garvieae* strains. The unrooted neighbor-joining tree (bootstrap 1,000, Kimura 2-parameter model) was constructed from the 5,713 bp concatenated DNA sequences of the seven loci (*als, atpA, tuf, gapC, gyrB, rpoC and galP*) of *L. garvieae*. Open and closed squares correspond to subgroups SB and SA, respectively. Strain origin is indicated by color code: green = vegetables, brown = cereals, red = meat, yellow = dairy, blue = fish, pink= human, black= animal intestine, white = mastitic cow. Grey shadows represent CCs.

1. Phylogenetic relationships between *Lactococcus garvieae* and *Lactococcus lactis*

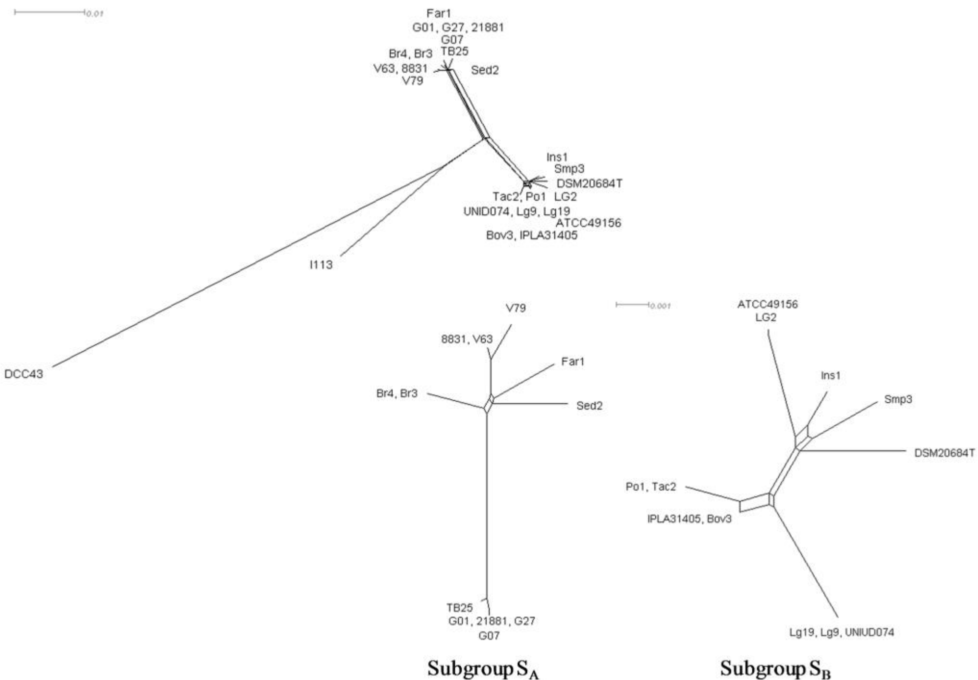


Figure 2. Splits decomposition analysis of *L. garvieae* population and subgroups. Parallelograms identify interconnected network of phylogenetic relationships between strains.

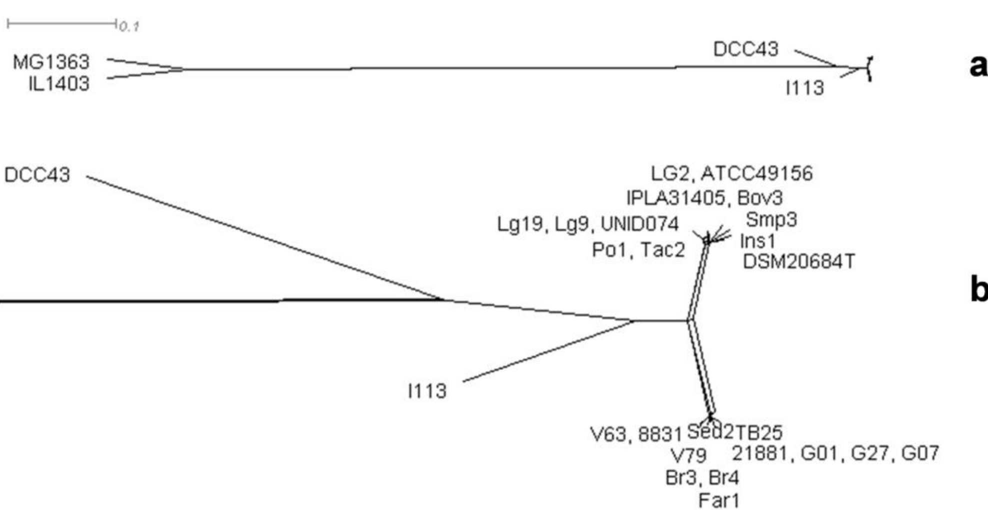


Figure 3. The concatenated sequences of all loci for *L. garvieae* and for the phylogenetically related species L. lactis subsp. lactis and *L. lactis* subsp. cremoris were analyzed using SplitsTree V4.12. a) Overview phylogeny, b) detail of *L. garvieae* population.

1. Phylogenetic relationships between *Lactococcus garvieae* and *Lactococcus lactis*

Genome comparison

General features of the *Lactococcus garvieae* genomes are reported in table 1, in comparison with general genome features of other *Lactococcus* species. For *L. garvieae*, individual genomes varied in size from 1.95 Mb to 2.24 Mb and contained 1778-2227 protein-coding genes. The results include genes that may belong to plasmids or phages (Florez et al., 2012; Aguado-Urda et al., 2012). Overall, the genomic variations in size or the number of the protein-coding genes were < 15 and 21 % respectively between any two strains. A comparison with other *Lactococcus* species, indicated for *L. lactis* a larger genome size than that of *L. garvieae* and a larger number of protein-coding genes. A lower GC content, ranging from 34.90 to 35.90, was also observed in *L. lactis*.

To estimate the number of genes present in each *L. garvieae* strain (core genome), a pan-genome profile and a core-genome profile were built using all possible BLAST combinations for each genome being sequentially added. The pan-genome is defined as a full complement of genes in a species (Mira et al., 2010; Milani et al., 2013). We identified a total of about 4100 orthologous genes (OGs). Figure 4 shows the predicted pan-genome size as a function of the number of genomes sequenced. It appears that the pan-genome size is leveling off (at about 4000-4100 genes), as every extra genome adds less new genes. The core genome is defined as the OGs that are conserved in all strains of the species. Figure 4 displays that the core genome decreases as more genome sequences are added and reaches a minimum of about 1300 genes. In addition, we identified a dispensable genome (present in some but not all 11 strains) of *L. garvieae* consisting of about 2760 genes.

Figure 5 shows the functional classification of the core and pan-genome genes based on COG (Clusters of Orthologous Genes) analyses. The majority of genes of the core genome belonged to the group of housekeeping functions, as well as other intriguing functions, such as metabolism and transport of carbohydrates (G), amino acid metabolism and transport (E), which may suggest that glycans and amino acids have shaped the genome of *L. garvieae* taxon. A gene fraction that appeared enlarged in the dispensable genome corresponds to defense mechanisms and DNA replication and repair. As in common in most bacteria, about 25% of the shared genes fall into the class of hypothetical proteins and protein of unknown function (Mira et al., 2010).

By using the computational procedure described above, we constructed *L.garvieae*- and *Lactococcus*-specific clusters of orthologous genes (LgCOGs and LCOGs respectively) from the proteins encoded in the genome of the 11 sequenced *L. garvieae*, three *L. lactis* subsp. *cremoris*, two *L. lactis* subsp. *lactis* and one *L. raffinolactis* (table 1). A total of 1491 LgCOGs were found to be specific to the 11 *L. garvieae* genomes, with *L. garvieae* DCC43 genome showing the highest presence of unique genes (383), representing 25 % of the total specific LgCOGs found (figure 6 A). About 70% of the total core genes were also conserved in the six sequenced *Lactococcus* genomes, suggesting that these genes may constitute the core genome of lactococci, likely inherited from a common ancestor (figure 6 B).

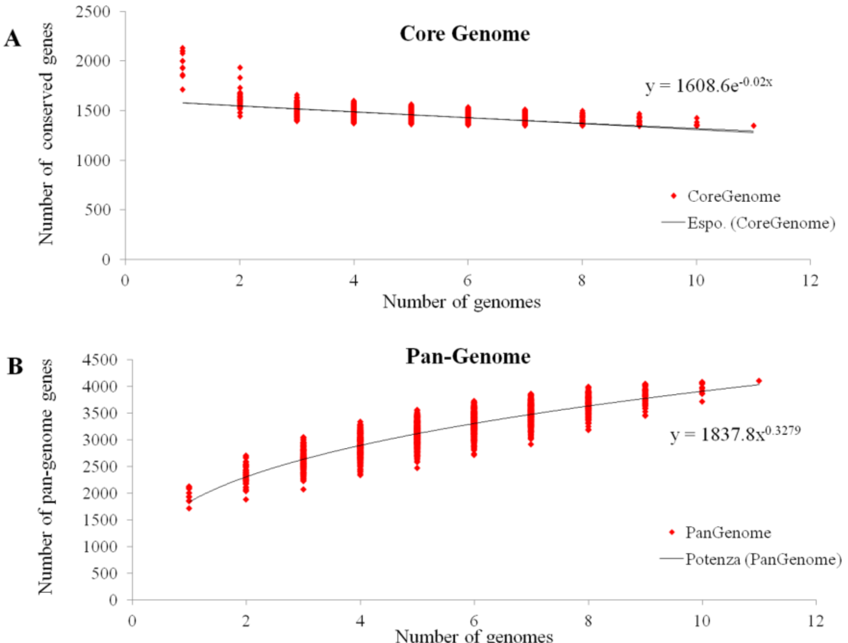


Figure 4. The distribution1 of the number of core COGs (A) and total pangenome COGs (B) found upon sequential addition of *n* genomes. In panel A, an exponential regression to core genome data is shown as a solid curve. In panel B, power law fit to the pan-genome size is shown as solid curve.

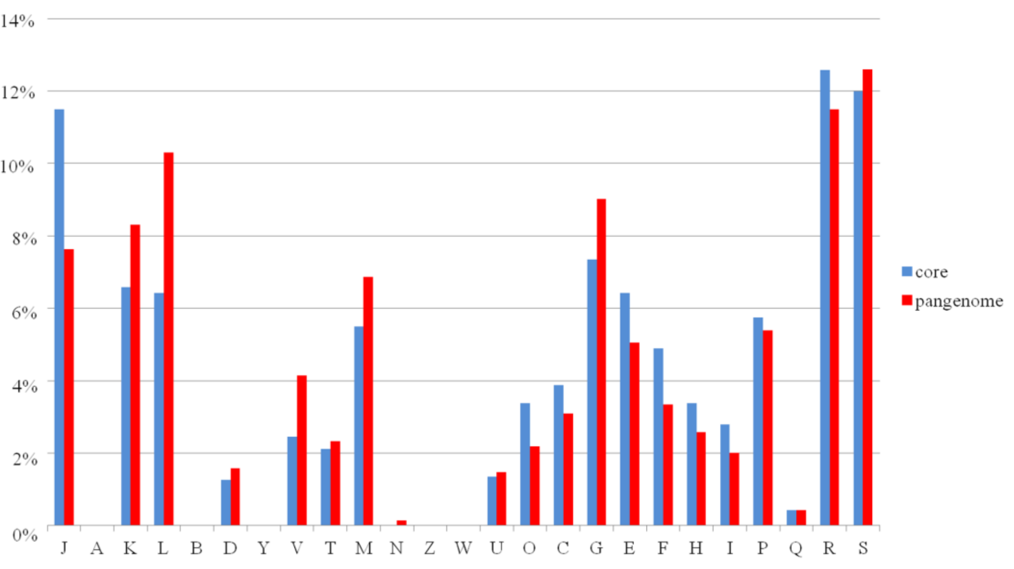
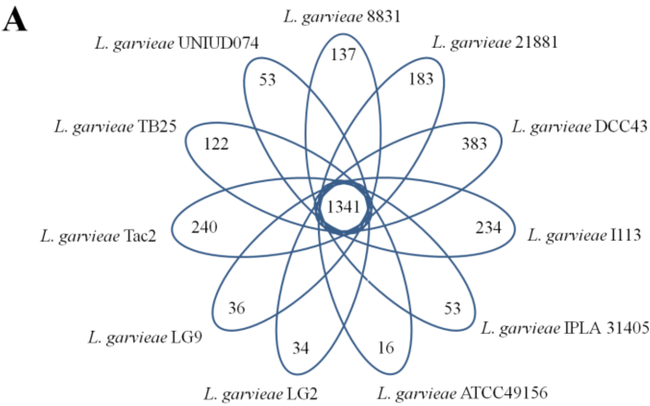


Figure 5. Bar chart showing a representation of COG families annotation of core COGs and whole pangenome COGs. Each COG family is identified by a one-letter abbreviation: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control and mitosis; E, amino acid metabolism and transport; F, nucleotide metabolism and transport; G, carbohydrate metabolism and transport; H, coenzyme metabolism; I, lipid metabolism; J, translation; K, transcription; L, replication and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, and chaperone functions; P, inorganic ion transport and metabolism; Q, secondary structure; T, signal transduction; U, intracellular trafficking and secretion; Y, nuclear structure; V, defense mechanisms; Z, cytoskeleton; W, extracellular structures; R, general functional prediction only; S, function unknown.



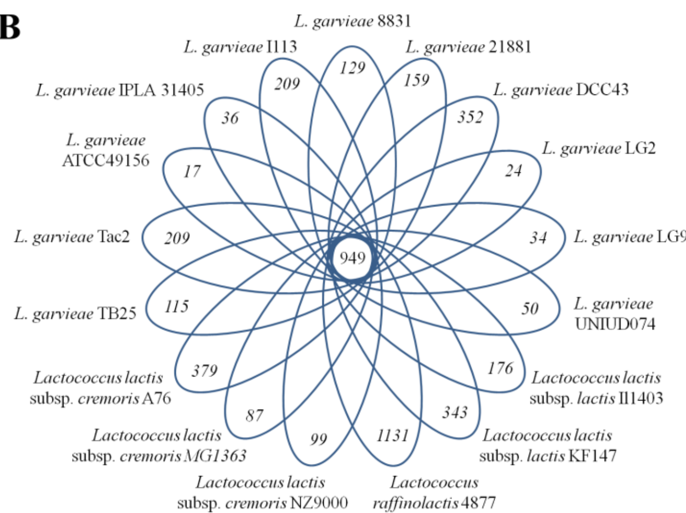


Figure 6. Venn diagram of core COGs (Clusters of Orthologous Genes) shared between all the strain analyzed and COGs unique to each single strain.

CONCLUSIONS

This first chapter draws these following conclusions:

- MLST experiment identified two branches containing the majority of *L. garvieae* strains;
- *L. garvieae* and *L. lactis* are phylogenetically linked;
- two strains, harbour a specific branch and are considered intermediate between the two lactococcal species;
- differences between core genome genes (1341) and dispensable genome pool (2760) could suggest the cosmopolitan lifestyle of *L. garvieae*;
- DCC43 appeared to be the most ancestral lineage of the species.

In general, this first step of the research has highlighted that *L. garvieae* species is characterized by a high degree of genome variability; taking into consideration this statement, the following chapters will be focused on a deeper study of this species analysing the main mechanisms involved in genome plasticity.

Related publication: Ferrario C., Ricci G., Milani C., Lugli G.A., Ventura M., Eraclio G., Borgo F., Fortina M.G. 2013. Lactococcus garvieae*: where is it from? A first approach to explore the evolutionary history of this emerging pathogen*. PLoS One 8, e84796.

REFERENCES

1. Aguado-Urda M., Gibello A., Blanco M.M., López-Campos G.H., Cutuli M.T., Fernández-Garayzábal J.F. (2012). *Characterization of plasmids in a human clinical strain of* Lactococcus garvieae. PLoS One 7:e40119.

2. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). *Basic local alignment search tool*. J Mol Biol. 215:403–410.

3. Feil E.J., Li B.C., Aanensen D.M,. Hanage W.P., Spratt B.G. (2004). *eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data*. J Bacteriol. 186:1518-1530.

4. Ferrario C., Ricci G., Borgo F., Rollando A., Fortina M.G. (2012). *Genetic investigation within* Lactococcus garvieae *revealed two genomic lineages*. FEMS Microbiol Lett. 332:153-161.

5. Flórez A.B., Reimundo P., Delgado S., Fernández E., Alegría A., Guijarro J.A., Mayo B. (2012). *Genome sequence of* Lactococcus garvieae *IPLA 31405, a bacteriocin-producing, tetracycline-resistant strain isolated from a raw-milk cheese*. J Bacteriol. 194:5118-5119.

6. Guindon S., Gascuel O. (2003). *A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood*. Syst Biol. 52: 696–704.

7. Huson D.H., Bryant D. (2006). *Application of phylogenetic networks in evolutionary studies*. Mol Biol Evol. 23:254-267.

8. Milani C., Duranti S., Lugli G.A., Bottacini F., Strati F., Arioli S., Foroni E., Turroni F., van Sinderen D., Ventura M. (2013). *Comparative genomics of Bifidobacterium animalis subsp. lactis reveals a strict monophyletic bifidobacterial taxon*. Appl Environ Microbiol. 79:4304-4315.

9. Mira A., Martín-Cuadrado A.B., D'Auria G., Rodríguez-Valera F. (2010). *The bacterial pan-genome: a new paradigm in microbiology*. Int Microbiol. 13:45-57.

10. Morita H., Toh H., Oshima K., Yoshizaki M., Kawanishi M., Nakaya K., Suzuki T., Miyauchi E., Ishii Y., Tanabe S., Murakami M., Hattori M. (2011). *Complete genome sequence and comparative analysis of the fish pathogen* Lactococcus garvieae. PLoS One 6:e23184.

11. Pearson W.R. (2000). *Flexible sequence similarity searching with the FASTA3 program package*. Methods Mol Biol. 132:185–219.

12. Reimundo P., Pignatelli M., Alcaraz L.D., D'Auria G., Moya A., Guijarro J.A. (2011). *Genome sequence of* Lactococcus garvieae *UNIUD074, isolated in Italy from a lactococcosis outbreak*. J Bacteriol. 193:3684–3685.

13. Ricci G., Ferrario C., Borgo F., Eraclio G., Fortina M.G. (2013). *Genome sequences of two* Lactococcus garvieae *strains Isolated from meat*. Genome Announc 1:e00018-12.

14. Ricci G., Ferrario C., Borgo F., Rollando A., Fortina M.G. (2012). *Genome sequences of* Lactococcus garvieae *TB25, isolated from Italian cheese, and* Lactococcus garvieae *LG9, isolated from Italian rainbow trout*. J Bacteriol. 194:1249-1250.

15. Sanger F., Nicklen S., Coulson A.R. (1977). *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci USA 74:5463–5467.

16. Sievers F., Wilm A., Dineen D.G., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J.D., Higgins D.G. (2011). *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. Mol Syst Biol. 7: 539.

17. Sneath P.H.A., Sokal R.R. (1973). *Numerical Taxonomy*. Freeman, San Francisco.

18. Takamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. (2011). *MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods*. Mol Biol Evol. 28:2731-2739.

19. van Dongen S. (2000). *Graph clustering by flow simulation*. PhD thesis. University of Utrecht, Utrecht, The Netherlands.

20. Vos M., Didelot X. (2009). *A comparison of homologous recombination rates in bacteria and archaea*. ISME. J 3:199-208.

21. Zhao Y., Wu J., Yang J., Sun S., Xiao J., Yu J. (2012). *PGAP: pan-genomes analysis pipeline*. Bioinformatics. 28:416–418.

2. INSERTION SEQUENCE ELEMENTS IN *LACTOCOCCUS GARVIEAE*

2. Insertion sequence elements in *Lactococcus garvieae*

Insertion sequences (ISs) are the simplest intracellular Mobile Genetic Elements which can occur in very high numbers in prokaryotic genomes, where they play an important evolutionary role by promoting genome plasticity.
In this second step of the research, for the first time the presence and dissemination of ISs along the genome of *Lactococcus garvieae* is presented.

MATERIALS AND METHODS

Nucleotide sequences of the twelve *L. garvieae* genomes deposited in the NCBI database (http://www.ncbi.nlm.nih.gov/) were analyzed (table 1). Homology searches were performed using BLASTN and BLASTX (Altschul et al., 1997) at GenBank and ISfinder (http://www-is.biotoul.fr) (Siguier et al., 2006b), to identify the ISs. Positive matches for transposase were confirmed manually to determine which family they belong by the comparison of the elements size, presence of terminal IRs and DRs, number of *orf*s and DDE consensus region with related elements (Mahillon and Chandler, 1998; Siguier et al., 2014). For new elements, IS names were kindly provided by ISfinder. Multiple alignments of DNA and protein sequences were performed using CLUSTAL_W (Thompson et al., 2002). Phylogenetic trees were generated from the matrices using the neighbor-joining method (Saitou and Nei, 1987).

Table 1. *L.garvieae* strains analyzed. nd= not determined.

Lactococcus garvieae Genomic lineage[a]	Strain	Source	Genome status	Scaffolds	Accession Number	Reference
A	21881	Human blood	draft	91	AFCC00000000.1	Aguado-Urda et al., 2011a
	8831	Rainbow trout	draft	87	AFCD00000000.1	Aguado-Urda et al., 2011b
	TB25	Cow cheese	draft	92	AGQX00000000.1	Ricci et al., 2012
B	ATCC 49156	Yellowtail	complete	1	NC_015930	Morita et al., 2011
	LG2	Yellowtail	complete	1	NC_017490	Morita et al., 2011
	UNIUD 074	Rainbow trout	draft	25	AFHF00000000.1	Reimundo et al., 2011
	LG9	Rainbow trout	draft	140	AGQY00000000.1	Ricci et al., 2012
	Tac2	Turkey	draft	97	AMFE00000000.1	Ricci et al., 2013
	IPLA 31405	Cow milk	draft	23	AKFO00000000.1	Flórez et al., 2012
C	I113	Meat products	draft	49	AMFD00000000.1	Ricci et al., 2013
D	DCC43	Mallard duck intestines	draft	68	AMQS00000000.1	Gabrielsen et al., 2012
nd	TRF1	Snake fecal material	draft	112	AVFE00000000.1	McLaughlin et al., 2014

RESULTS AND DISCUSSIONS

Twelve published *L. garvieae* genomes were searched for the presence of IS elements; as reported in table 1, they refer to strains coming from different ecological niches: diseased fish, cheeses, meat products, duck intestine, snake fecal material and human blood. They are representative of the four different genomic lineages characterizing the population structure of the species (Ferrario et al., 2013). Analysis of the genomes revealed the presence of 15 distinct ISs that are members of the IS3, IS982, IS6, IS21 and IS256 families (Mahillon and Chandler, 1998), including five new elements designated ISLgar. The active centre of the transposase, represented by three conserved acidic amino acids, D,D, and E, constituting the DDE motif, was identified in each element. Their frequency and distribution along the 12 genomes of *L. garvieae* is reported in table 2. It is interesting to note a high variability, with strains ATCC 49156 and LG2 carrying multicopies of two distinct ISs, and strains IPLA 31405 and Tac2 carrying the higher number of different IS elements. ISs belonging to IS3 family seemed the most representative in *L. garvieae*, with several copies distributed in 9 of the 12 genomes tested. On the contrary, some ISs seem strain-specific, as ISLgar5 in strain IPLA3 1405 and IS712 in strain DCC43. No evident correlation between genomic lineages/niche of isolation and presence of ISs was found, with the exception of IS6 family elements, not detected in the genomes of strains isolated from diseased fish.

IS*3 family*. Five IS elements were assigned to the IS*3* family, including 39 copies of IS*LL6*, two copies of IS*904*, two copies of IS*981*, one copy of a newly named IS, IS*Lla3,* and two copies of a new element, designed IS*Lgar1*.
The most abundant IS discovered in *L. garvieae* showed the highest nucleotide similarity (95-99%) with an IS element found in *Lactococcus lactis* plasmid pWV04, reported in ISFinder and Genbank (accession U23813.1) as IS*LL6* element, and other unnamed ISs found in *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* (accession AF049873.2; CP000428.1; CP002094.1; AM406671.1; CP002367.1). These IS*LL6*-like elements, including those found in *L. garvieae* strains, show highly conserved and characteristic terminal ends, constituted by two imperfect left inverted repeats (IR-L) (8 and 16 bp respectively, separated by 43 bp) and a 32-bp imperfect right inverted repeat (IR-R). No DR were found. IS*LL6* elements contain two non-overlapping *orf*s in the relative translational reading phases 0 and -1 respectively. *orf* 1*,* highly conserved, encoding a 91 amino acid sequence, was found 11 bp downstream IRL. o*rf* 2*,* showing the typical DD(35)E K motif, ranged in length from 864 to 732 bp, according to the "in frame" start codon found in the intergenic region in the deposited sequences. The GC content of these IS elements is 35.6%, a value coherent with the average GC content of the *L. lactis* genome, but markedly lower than the average content of the *L. garvieae* genome (ranging from 37.7 to 38.8 %). This evidence could indicate a possible transposition from *L. lactis* in *L. garvieae* genome. As previously reported (Ferrario et al., 2013) these two species are phylogenetically linked. The nucleotide sequence of the IS*LL6* element present on *L. garvieae* ATCC 49156 (for which complete genome is available) can be found in figure 1. A potential -10 promoter was located inside the left IR and a -35 hexamer has been found in the right terminal IR; as reported for members

Table 2. IS elements found along the genome of the tested *L. garvieae* strains.

IS family		*Lactococcus garvieae* strains											
		21881	8831	TB25	ATCC 49156	LG2	UNIUD 074	LG9	Tac2	IPLA 31405	I113	DCC43	TRF1
IS*3*													
	IS*LL6*	+ (2)[a]	-	+ (2)	+ 12	+ 12	+ (7)	+ (1)	+ (1)	+ (1)	+ (1)	-	-
	IS*904*	-	-	-	-	-	-	-	+ (1)	+ (1)	-	-	-
	IS*981*	-	-	+ (1)	-	-	-	-	+ (1)	-	-	-	-
	IS*Lla3*	-	-	+ (1)	-	-	-	-	-	-	-	-	-
	IS*Lgar1*	-	-	-	-	-	+ (1)	+ (1)	-	-	-	-	-
IS*982*													
	IS*982*	+ (1)	-	-	-	+ 1	-	-	-	+ (1)	-	-	-
	IS*Lgar2*	-	-	-	+ 7	+10	-	-	-	-	-	-	-
	IS*Lgar3*	-	-	-	-	-	-	-	-	-	-	+ (1)	-
IS*6*													
	IS*S1N*	+ (1)	-	+ (1)	-	-	-	-		+ (1)		-	-
	IS*Teha2*	-	-	-	-	-	-	-	+ (1)	-	+ (1)	-	-
	IS*SS1*	-	-	-	-	-	-	-	-	-	+ (1)	-	-
	IS*SS1S*	+ (1)	-	-	-	-	-	-	-	-	-	-	-
	IS*Lgar4*	-	-	-	-	-	-	-	+ (1)	-	-	-	-
IS*256*													
	IS*Lgar5*	-	-	-	-	-	-	-	-	+ (1)	-	-	-
IS*21*													
	IS*712*	-	-	-	-	-	-	-	-	-	-	+ (1)	-

[a] For draft genomes, the minimum copy number is shown in parentheses.

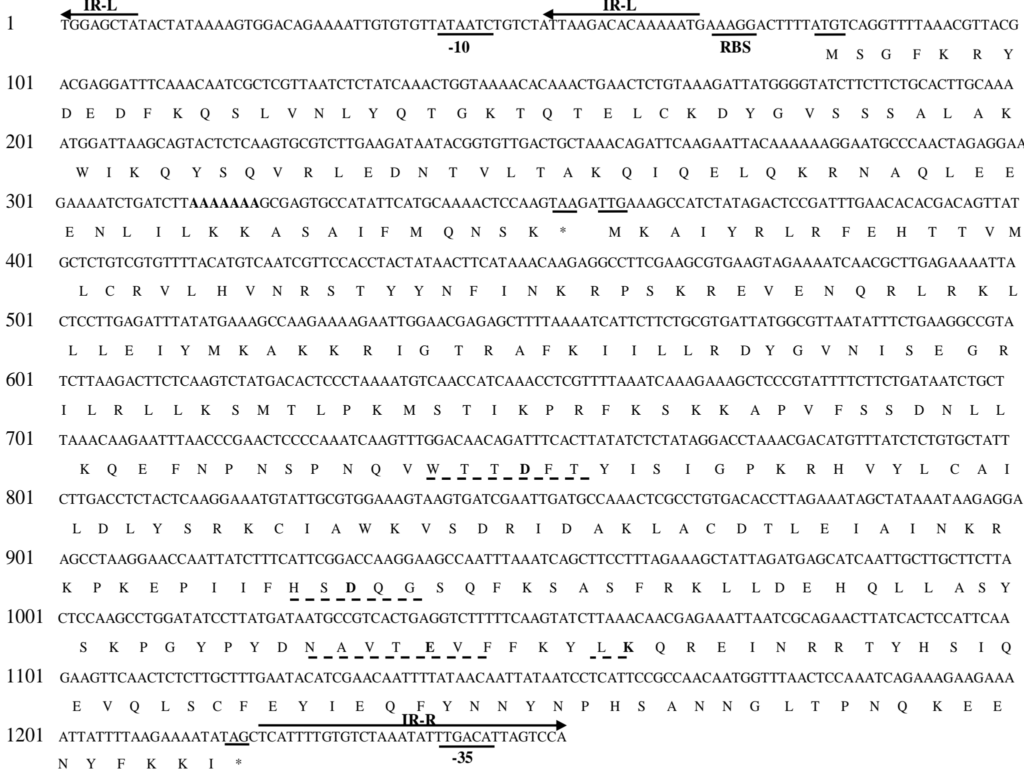


Figure 1. IS*LL6* element presents on *L. garvieae* ATCC 49156.

of this family, the localization of these hexamers could indicate a transposition mechanism “copy-paste”, in which the formation of a circular intermediate results in the assembly of a transient strong promoter composed of a -35 element in the right IS end and a -10 promoter element in the left end (Mahillon & Chandler, 1998). Moreover, a potential frameshift window of the A7 type was present. Twelve identical and intact copies are distributed throughout the ATCC 49156 and LG2 genomes, as previously reported by Morita et al., 2011. To investigate the influence of these ISs on protein-coding regions, the CDSs (Coding DNA Sequences) in the regions adjacent to each IS copy were analyzed. The nucleotide sequence of the 3-kb regions upstream and downstream each of all IS elements were extracted from the entire genome sequence and the 6-kb sequence generated was searched for CDSs by using the BLAST program. For both strains analyzed, IS copies were randomly distributed through the genome and seemed inserted in non-coding regions. An alignment analysis of the ATCC 49156 and LG2 complete genomes revealed a 99% sequence identity and their co-linearity along the whole genome, except for a 113111 bp chromosomal inversion. The chromosomal

fragment is exactly delimited by two full IS*LL6* copies in inverted repeat with respect to each other, that could be involved in this chromosomal rearrangement.
Analysis of the draft genomes of the other *L. garvieae* strains, showed the presence of seven full-length IS*LL6* elements in UNIUD 074 strain, six of which shared a 99% nucleotide similarity each other and 93% nucleotide similarity with the other one. The main differences found (also in respect to previously described *L. garvieae* IS*LL6*) was an intergenic region between the two *orfs* of 135 bp for the identical six copies, and of 45 bp for the other one. At least one complete IS*LL6* copy was found in strains IPLA 31404, LG9, Tac2 and I113; for strains TB25 and 21881 homology search in the DNA database revealed the presence of one IS*LL6* copy with *orf2* interrupted at position 1031 and the existence of two contigs containing portions of another identical copy.

IS*904*, described for the first time in *L. lactis* genome (Dodd et al., 1990), was present in two *L. garvieae* strains, IPLA 31405 and Tac2. A high nucleotide homology (>99%) was found with IS*904* of *L. lactis* subsp. *lactis* (accession X52273.1; M27276.1), *L. lactis* subsp. *cremoris* (accession CP003157.1) and *Streptococcus thermophilus* (accession CP003499.1). In *L. garvieae* strains IS*904* is 1230 bp long and contains two *orfs* of 290 and 761 bp respectively, separated by 95 bp. Within *orf*1, and in proximity of the stop codon, an A AAA AAG frameshift window was present. Within *orf* 2, the typical DD(35)E K motif was detectable. The element was characterized by the presence of 39 bp imperfect IRs, containing 2 mismatches. A comparison with IRs found in the related IS*904* is shown in figure 2. IR-L seemed highly conserved in all sequences examined; the analysis of the IR-R underlines a difference between *L. garvieae* and the other species, the latter showing 8 identical mismatches. In *L. garvieae* strains, examination of the sequences adjacent to the IRs revealed the presence of a TTT sequence that could represent perfect DR sequences; different hypothetical DRs seemed present in the other compared species (figure 2).

DR		
	IS*904* **IR-L**	TGGAAAGGTTATAATAAACTAGACACAAAGTTAAGAGAA
TTT	IS*904* **IR-R** *L.garvieae*	TGGAAAGGTTATAATAAACTAGACACGGAGTTAAGAGAA
TTAT	IS*904* **IR-R** *L. lactis* subsp. *lactis*	TGGAAAGTCAACGAAAAACTAGACACGGAGTTAAGAGAA
TCC	IS*904* **IR-R** *S. thermophilus*	TGGAAAGTCAACGAAAAACTAGACACGGAGTTAAGAGAA
AAT	IS*904* **IR-R** *L. lactis* subsp. *cremoris*	TGGAAAGTCAACGAAAAACTAGACACGGAGTTAAGAGAA

Figure 2. Comparison of left (IR-L) and right (IR-R) inverted repeat sequences at the termini of IS*904* in *L. garvieae* (accession AKFO01000018.1), *L.lactis* subsp. *lactis* (accession X52273.1), *L.lactis* subsp. *cremoris* (accession CP003157.1) and *Streptococcus thermophilus* (accession CP003499.1); mismatch positions are indicated by shaded areas. Potential Direct Repeat sequences (DRs) are also indicated.

IS*981* was originally identified in several lactococci (Polzin and McKay, 1991); this element, present in two *L. garvieae* genomes, showed the highest nucleotide homology with IS*981* of *L. lactis* subsp. *lactis* (accession M33933.1), *L. lactis* subsp *cremoris* (accession CP000425.1) and *Streptococcus iniae* (accession CP005941.1). IS*981* is highly conserved in these species, suggesting possible recent interaction and horizontal transfer events. The element is 1224 bp long and contains two non-overlapping *orfs* of 261 and 838 bp respectively. A potential frameshift window of the A7 type was present in proximity of the stop codon of the *orf*1. The element is flanked by 40 bp imperfect IR (13/14 mismatches). In the majority of the cases, no DRs were detectable.
In *L. garvieae* strain TB25 an IS element showed homology with some nucleotide sequences reported in GenBank, found in *L. lactis* subsp. *cremoris* (accession CP003132.1; CP003133.1), *L. lactis* subsp. *lactis* (accession X92946.1; AE005176.1) and *Enterococcus faecalis* (accession CP002491.1). One of these sequences (AE005176.1) was named IS*1077*. This element has been previously cited in Bolotin et al., 1999 as an IS occurring in *L. lactis* together with IS*904*, but not characterized. This element has no homology with IS elements present in ISfinder database. For this reason, we carried out a more deepened study. Firstly, we submitted the *L. lactis* subsp. *cremoris* IS sequence (accession CP003132.1) to ISfinder for the name attribution. According to IS nomenclature, the element has been named IS*Lla*3;. it is 1484 bp in size and contains 30 bp imperfect IRs at its end. IR-L was conserved in all sequences examined; the analysis of IR-R showed for *L. garvieae, L. lactis* subsp. *cremoris* and *E. faecalis* identical sequences characterized by 7 mismatches. In the related elements present in *L. lactis* subsp. *lactis*, IR-R showed substantial differences. Examination for *orf*s revealed two potential *orf*s, one extending from 57 to 614 bp and the other from 611 to 1429 bp, containing the typical DDE motif. In the published sequence of *L. garvieae* strain TB25, the second *orf* is interrupted at position 1148. It is possible to hypothesize a mutation that render inactive the corresponding transposase, or an error occurring in sequence determination. A search for promoter sequences yielded only a consensus -10 sequence preceding *orf*1. Potential -35 and -10 sequences that may function as weak promoters preceded *orf*2. To verify the relationships of IS*Lla3* and related elements of the IS3 family, we constructed a phylogenetic tree using the best-studied IS3 family members, via multiple alignments of *orf*2 transposase sequences. As shown in figure 3, IS*Lla3* fits within IS3 family, particularly within the subgroup IS150, according to Mahillon and Chandler classification (1998).
In the genome of *L. garvieae* strains UNIUD 074 and LG9 we identified a new element showing 69% nucleotide homology with the *Streptococcus agalactiae* IS*861*, registered in ISfinder, and 83% homology with published sequences found in other *Streptococcus* species. A 78% similarity and 63% identity between the amino acid sequences were found. The new element, belonging to the subgroup IS150 (figure 3), has been registered in ISfinder as IS*Lgar1***.** It is 1431 bp long, flanked by 28-bp imperfect IRs that shared the signature of the family, including the characteristic ends (5'-TG and CA-3'). IS*Lgar1*contains two non-overlapping *orfs*, one extending from 54 to 569 bp and the other from 593 to 1402 bp, displaying the catalytic DDE motif. The potentially functional transposase most likely is translated as a single amino acid

protein via -1 translational frameshifting. A potential ribosome-binding site was only found upstream *orf*2.

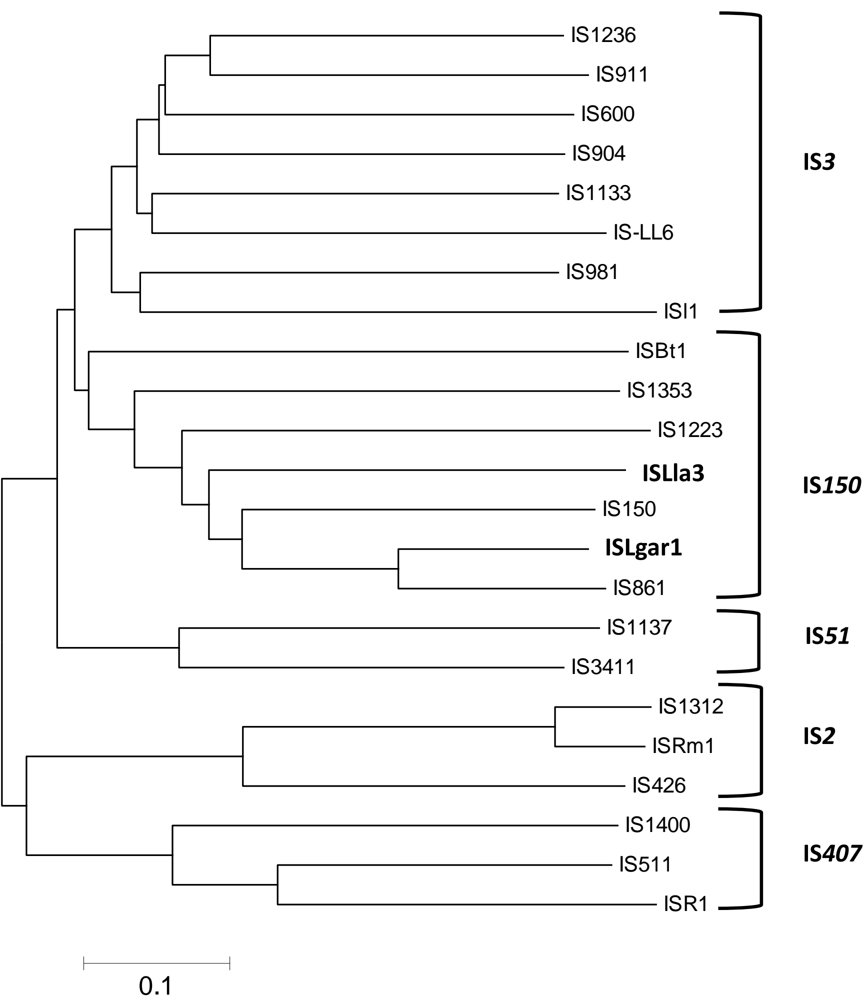


Figure 3. Dendrogram based on the alignment of the amino acid sequences of predicted transposases from several elements of different subgroups of the IS*3* family.

IS*982 family*. IS*982* family, firstly described in lactococci, comprises several IS members spread both in Gram positive and Gram negative bacteria. The family members are between 900 and 1100 bp long and carry similar IRs of between 15 and 35 bp. The elements of this family contain a single *orf* encoding a protein exhibiting low amino acidic identity among the family members.
Three IS elements were assigned to the IS*982* family in *L. garvieae*, including IS*982* and two newly named ISs, IS*Lgar2* and IS*Lgar3* (table 2). Seven and 10 identical copies of IS*Lgar2* were present along the genome of strains ATCC 49156 and LG2 respectively. This IS, previously cited by Morita et al., 2011 as IS*982*, shows 84% identity with IS*982* of *L. lactis* subsp. *cremoris* (accession L34754). IS*Lgar2* is 998 bp long, consists of one unique *orf* encoding a 296 amino acid sequence and contains 20-bp perfect IRs, highly similar to IRs present in *L. lactis* IS*982*. DRs have been found in the majority of the IS copies, ranging from 6 to 9 and characterized by an high A-T presence. IS*Lgar2* copies were randomly distributed through the genome of strains ATCC 49156 and LG2 and inserted in non-coding regions. Interestingly, in LG2 the 16.5 kb capsule gene cluster that characterizes this *L. garvieae* strain (Morita et al., 2011) is flanked on both ends by IS elements, an IS*Lgar2* copy at the left end and an IS*982* copy at the right end.
Strain DCC43 contains a copy of one IS showing 41 and 44 % amino acid identity (61 and 64% positivity) with IS*Lgar2* and IS*982* respectively. For this new element the name IS*Lgar3* has been attributed. IS*Lgar3* contains a single *orf* putatively encoding a 297-amino acid protein and 17-bp imperfect IRs (1 mismatch over 17 nt). In DCC43 strain, IS*Lgar3* is flanked by a 6-bp (ATTAAT) duplication.

IS*6 family*. This family is composed of 45 members, 30 of which are iso-copies (Mahillon and Chandler, 1998). They generally range in length from 789 bp to 880 bp, all carry short related (15 to 20 bp) terminal IRs and generally create 8-bp direct repeats.
As shown in table 2, several *L. garvieae* strains carry single copies of different members of this family, IS*S1N* found in *L. lactis* (accession M37395), IS*teha2* from *Tetragenococcus halophilus* (accession AP012046), IS*S1* found in *E. faecium* (accession U49512) and IS*S1S* found in *L. lactis* (accession M77708).
In Tac2 strain we found another element which represents a new IS within the family. It shows 83% homology with IS*S1S*, with amino acid identity and positivity of 94 and 98% respectively. This new IS, named IS*Lgar4* is 809 bp long and encodes a putative protein of 226 amino acids. The element is characterized by the presence of 18-bp imperfect IRs, containing 1 mismatch; no DRs were detectable.
Two incomplete sequences related to IS elements were also found in IPLA 31405, on two contigs. A preliminary analysis of these sequences indicates a possible belonging to IS*6* family.

IS*21 family*. Only one member of this family is present in the *L. garvieae* genomes analyzed. Specifically, in strain DCC43 an element showing 99% homology with IS*712* from *L. lactis* (accession NC009004) was found. In *L. garvieae,* IS*712* is 2162 bp long with 49 IRs showing 10 mismatches. As reported for the members of the

family, two *orf*s, known as *istA* (1224bp-407aa) and *istB* (759 bp-252aa) were present; the longer putative *istA* carries the DDE motif, lacking the conserved K residue.

IS256 family. Members of the IS*256* family have been identified from a wide range of organisms. In *L. garvieae* one element, showing 81% nucleotide homology with IS*EfmL* from *E. faecium* (AY887085) was found. The new member of the family has been reported in ISfinder as IS*Lgar5*. It is 1336 bp long, with 26 imperfect IRs (4 mismatches). In strain IPLA 31405 IS*Lgar5* generates a 8 bp DRs. The unique *orf* of 1191 bp (396 aa) contains a potential DDE motif with a long spacing between the second D and E residues, typical of many members of the family.

CONCLUSIONS

In this secon chapter, the main results obtained are:

- for the first time, the presence and distribution of IS elements in *L. garvieae* genome have been depicted;
- our data indicate that the distribution and abundance of IS elements among *L. garvieae* genomes is highly variable: 77 IS copies from 15 distinct IS elements were characterized and classified into 5 different families;
- the close similarities of most IS elements in *L. garvieae* and *L. lactis* indicate the possibility of genetic exchange between these species;
- within individual genomes, the analysis of the sequences of the copies of a given IS revealed that the intra-genomic sequence diversity is very low, suggesting that many ISs in *L. garvieae* are evolutionary young and might have been recently acquired;
- this study also revealed 5 new IS elements in *L. garvieae* genomes.

Related publication: Eraclio G., Ricci G., Fortina M.G. (2015). *Insertion sequences elements in* Lactococcus garvieae. Gene 555:291-296.

REFERENCES

1. Aguado-Urda M., López-Campos G.H., Blanco M.M., Fernández-Garayzábal J.F., Cutuli M.T., Aspiroz C., López-Alons V., Gibello A. (2011a). *Genome sequence of* Lactococcus garvieae *21881, isolated from a case of human septicaemia*. J. Bacteriol. 193:4033–4034.

2. Aguado-Urda M., López-Campos G.H., Gibello A., Cutuli M.T., López-Alonso V., Fernández-Garayzábal J.F., Blanco M.M. (2011b). *Genome sequence of* Lactococcus garvieae 8831, *isolated from rainbow trout lactococcosis outbreaks in Spain*. J. Bacteriol. 193:4263–4264.

3. Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997). *Gapped BLAST ans PSI-BLAST: a new generation of protein database search programs*. Nucleic Acid Res. 25:3389-3402.

4. Bolotin A., Mauger S., Malarme K., Ehrlich S.D., Sorokin A. (1999). *Low-redundancy sequencing of the entire* Lactococcus lactis *IL1403 genome*. Ant Van Leeuw. 76:27-76.

5. Dodd H.M., Horn N., Gasson M.J. (1990). *Analysis of the genetic determination for production of the peptide antibiotic nisin*. J. Gen Microbiol. 136:555-566.

6. Ferrario C., Ricci G., Milani C., Lugli G.A., Ventura M., Eraclio G., Borgo F., Fortina M.G. 2013. Lactococcus garvieae*: where is it from? A first approach to explore the evolutionary history of this emerging pathogen*. PLoS One. 8:e84796.

7. Flórez A.B., Reimundo P., Delgado S., Fernández E., Alegría A., Guijarro J.A., Mayo B. (2012). *Genome sequence of* Lactococcus garvieae *IPLA 31405, a bacteriocin-producing, tetracycline-resistant strain isolated from a raw-milk cheese*. J. Bacteriol. 194:5118-5119.

8. Gabrielsen C., Brede D.A., Hernández P.E., Nes I.F., Diep D.B. (2012). *Genome sequence of the bacteriocin-producing strain* Lactococcus garvieae *DCC43*. J. Bacteriol. 194:6976-6977.

9. Mahillon J., Chandler M. (1998). *Insertion sequences*. Microbiol Mol Biol Rev. 62: 725-774.

10. McLaughlin R.W., Cochran P.A., Dowd S.E. (2014). *Isolation of* Lactococcus garvieae strain *TRF1 from the fecal material of a Timber Rattlesnake*. Curr Microbiol. 69:63-68.

11. Morita H., Toh H., Oshima K., Yoshizaki M., Kawanishi M., Nakaya K., Suzuki T., Miyauchi E., Ishii Y., Tanabe S., Murakami M., Hattori M. (2011). *Complete*

genome sequence and comparative analysis of the fish pathogen Lactococcus garvieae. PLoS One. 6:e23184.

12. Polzin K.M., McKay L.L. (1991). *Identification, DNA sequence, and distribution of IS981, a new high-copy-number insertion sequence in lactococci*. Appl Environ Microbiol. 57:734-743.

13. Reimundo P., Pignatelli M., Alcaraz L.D., D'Auria G., Moya A., Guijarro J.A. (2011). *Genome sequence of* Lactococcus garvieae *UNIUD074, isolated in Italy from a lactococcosis outbreak.* J. Bacteriol. 193:3684–3685.

14. Ricci G., Ferrario C., Borgo F., Eraclio G., Fortina M.G. (2013). *Genome sequences of two* Lactococcus garvieae *strains isolated from meat*. Genome Announc. 1:e00018-12.

15. Ricci G., Ferrario C., Borgo F., Rollando A., Fortina M.G. (2012). *Genome sequences of* Lactococcus garvieae *TB25, isolated from Italian cheese, and* Lactococcus garvieae *LG9, isolated from Italian rainbow trout*. J. Bacteriol. 194:1249-1250.

16. Saitou N., Nei M. (1987). *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol. 4:406-425.

17. Siguier P., Gourbeyre E., Chandler M. (2014). *Bacterial insertion sequences: their genomic impact and diversity.* FEMS Microbiol Rev. 38:865-891.

18. Siguier P., Perochom J., Lestrade L., Mahillon J., Chandler M. (2006b). *ISfinder: the reference centre for bacterial insertion sequences*. Nucleic Acid Res. 34:D32-D36.

19. Thompson J.D., Gibson T.J., Higgins D.G. (2002). *Multiple sequence alignement using ClustalW and ClustalX.* Curr Protoc Bioinformatics. 2:Unit2.3. doi: 10.1002/0471250953.bi0203s00.

3. PHENOTYPIC AND GENOTYPIC DIFFERENCES BETWEEN EUROPEAN AND AMERICAN STRAINS OF *LACTOCOCCUS GARVIEAE*

The recent isolation of new strains of *L. garvieae* from two American dairy farms has led us to investigate the potential variability of the species in function of the different geographical area of origin.
The analysis was carried out by comparing European and American isolates both at phenotypic (carbon source utilization and sensitivity to different antibiotics) and genotypic (molecular typing, plasmid isolation, genome analysis) level.

MATERIALS AND METHODS

Sampling, selection, culturing and identification of *L. garvieae* isolates

A total of 59 laboratory strains of *Lactococcus garvieae* were analysed in this part of the research. The isolates come from dairy farms from United States of America, Italian cheeses, meat products, diseased fish, vegetables and human faeces (table 1).
Samples coming from human faeces and dairy farm from USA were searched from the presence of *L. garvieae* as follow: all samples were collected and transported in isothermal boxes to the laboratory; samples (2 g) were enriched in 1:9 (w/w) M17 both (Difco, Detroit, Michigan, USA) supplemented with 1g/L glucose (M17-G) at 30°C for 24h. After enrichment, DNA was extracted as reported elsewhere (Ferrario, 2012) and species-specific PCR assay was carried out (Zlotkin et al.,1998). Positive samples were then plated and incubated at 30°C for 24 h and after incubation, randomly selected colonies were purified and submitted to taxonomic identification, as reported above.
Stock cultures were maintained at -80°C in M17-G with 20% glycerol.

Carbon source utilization

Ability of strains of *L. garvieae* to metabolize different carbon sources were assessed using glucose (as a positive control), lactose and sucrose. All the experiments were carried out using API 50 CHL medium, which composition is reported in table 2; carbon source were prepared separately as stock solutions (10%) and sterilized by filtration. To eliminate traces of growing media, cells to be used as inoculum were washed twice with distilled water, then inoculated in 200 µL of API 50 CHL in a 96 well plate. Carbon source was added at final concentration of 0.2% and cells were inoculated at 1% (10^7 UFC/mL). Two different negative controls, one specific for the strain (API 50 CHL media inoculated with cells, but without any carbon source), and one specific for the media (only 200 µL of API 50 CHL), were used. Experiments were carried out in triplicate. Incubation was carried out at 30°C for 18h.

Antibiotic resistance

Strains of *L. garvieae* were tested for their susceptibility to different antibiotics using disk diffusion methods (Neo-Sensitabs, Rosco-Diagnostica, Denmark). Antibiotics were chosen in order to cover antibiotics normally used to treat *L. garvieae* infections and to test new ones as well. List of antibiotic assessed in this study together with resistance/sensibility threshold is reported in table 3.
Experiments were carried out in Muller Hilton plate agar (Fluka, Germany). An overnight culture (10^9 UFC/mL) was diluted in sterile physiological solution (0.9%

NaCl) until 10^7 UFC/mL, and by using a sterile swab, cells were deposited in the plate. Experiments were carried out in triplicate and incubated at 30°C for 18h.
Minimal inhibitory concentrations (MICs) were determined by broth dilution method in M17-G, with a cell concentration of 10^5 UFC/mL; tubes were incubated at 30°C for 18 h. Experiments were repeated three times.
Growth curves in presence and absence of sublethal concentrations of antibiotics were carried out for selected *L. garvieae* strains. Overnight culture were diluted to 10^7 UFC/mL in physiological solution and then inoculated in fresh media (M17-G with and without antibiotic) at 10^5 UFC/mL; negative controls without antibiotics were predisposed. Cells were harvested after 4, 6, 8, 16 and 24 h of incubation; aliquots were properly diluted in physiological solution, disseminated on M17-g agar plates and incubated for 24 h at 30°C. For each dilution, three plates were predisposed and then the results were averaged.

Plasmid and total DNA extraction and quantification

For strains grown in pure culture, total DNA was extracted according to Borgo et al., 2012. Plasmidic DNA was obtained following Fortina et al., 2009 with some modifications: to increase extraction yield, 6 mL of pure culture were used and the volume of each reagent was duplicate. To allow elimination of cellular debris, one chloroform-isoamylic alcool (24:1-v/v) extraction was followed by three extractions with phenol and one with chloroform-isoamylic alcool.
Concentration and purity of DNA were determinated with a UV-Vis spectrophotometer (Jenway, Genova, Italy).

Genomic fingerprinting of *L. garvieae*

L. garvieae strains were typed by combined analysis of Repetitive Element PCR (REP) using primers BOXA1R (Versalovic et al., 1994; De Urraza et al., 2000) and Random Amplification of Polymorphic DNA-PCR (RAPD) with primer M13 (Rossetti and Giraffa, 2005). The sequences of primers used and the annealing temperature are reported in table 4. Amplicons were separated in a 2% agarose gel.

PCR amplification

PCRs were performed using a PCR-Mastercycler 96 (Eppendorf, Hamburg, Germany) in a 25 µL reaction mixture contained 100 ng of bacterial DNA, 2.5 µL of 10× reaction buffer (Fermentas, Vilnius, Lithuania), 200 µM of each dNTP, 2.5 mM MgCl2, 0.5 µM of each primer, and 0.5 U of Taq polymerase (Fermentas). After incubation for 2 min at 94 °C, samples were subjected to 35 cycles of 60 s at the annealing temperature (table 4), followed by 1 min at 72 °C; the reaction was completed by 7 min at 72 °C and kept at 4 °C. Amplification products were separated on a 1.5% agarose gel stained with ethidium bromide in 1× TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.2) buffer and photographed.

Southern blotting analysis and probes preparation

Probes preparation and Southern analysis were carried out as reported by Ricci and Fortina (2006). Probes are constructed on gene *tetM*, *tetS*, *hly* and *phospho-β-galactosidase* (*P-β-gal*).

Genome analysis

Genome analysis carried out on genomes of *L. garvieae* available in databases; folic acid pathway genes were identified by RAST server (Aziz et al., 2008) and confirmed with BLASTn (Altschul et al., 1990). Comparisons against *Lactococcus lactis* subsp. *lactis* KF174 (CP001834) were carried out through BLASTp (Altschul et al., 1997).

RNA isolation

Aliquots (70 µL) of *L. garvieae* cultures (10^7 UFC/mL) were inoculated in 7 mL of M17-G with or without sublethal concentration of antibiotics. Cells were grown until reaching of exponential phase, and then washed twice in sterile physiological solution. Total RNA was isolated using NucleoSpin RNA II (Macherey-Nagel, Switzerland) according to manufacturer's protocol, with some modifications: to facilitate cells lysis, lysozyme was increased to 8 mg/mL and the mixture incubated at 37°C for 40 min. Once RNA concentration and purity were optically determined at 260 and 280 nm with a UV-Vis spectrophotometer, RNA integrity was verified by visualization of the two bands (16S and 23S rRNA) by gel electrophoresis in a RNAsi-blocking buffer (diethyl pyrocarbonate 1ml/L of 1X TAE buffer). 1.5 µg of total RNA was treated with DNase1, Rnase-Free kit (Fermentas) at 37°C for 1h, and after addition of EDTA 50 mM to stop the reaction, was incubated at 65°C for 10 min. RNA was stored at -80°C and used for cDNA synthesis.

cDNA synthesis and RT-PCR experiments

Total RNA was quantified after DNase treatment and 0.1 µg (100 ng) was subjected to reserve transcription in to cDNA with RevertAid First strand cDNA Synthesis Kit (Fermentas) in accordance with manufacturer's instructions.

The level of gene expression was determined fluorometrically using SYBR Green I dye with a LineGene 9620 (Bioer) thermocycler.

A master mix containing, per sample, 7.5 µL of qPCR Master mix (EUR_x, Poland), 1.78 µL water, 0.36 µL of each primer (final concentration 300 nM), was combined with 5 µL of each cDNA sample. This mixture was then centrifuged to deposit samples at the bottom of each lightcycler glass capillary tube before being placed into the chamber of the thermocycler for real time analysis of PCR amplification. The qPCR temperature procedure was performed according to the manufacture's recommendations and corresponds to an initial cycle of denaturation at 94°C for 5 min, followed by 40 cycles of denaturation 95°C for 30 sec, annealing and extension 58°C for 30 sec. Fluorescence acquisition was done at the end of each amplification. After real time PCR a melting curve analysis was performed by measuring fluorescence during heating from 72°C to 95°C at a transition rate of 0.5°C/s. Amplifications were performed in triplicate.

The housekeeping gene *gapC* (codifying for the essential glycolytic gene glyceraldehyde-3-phosphate dehydrogenase) was used as internal control to which all other gene expressions were normalized. The primers used are reported in table 5.
Primers and reaction efficiency were tested before gene expression quantification. Primers were constructed and then tested by OligoAnalyzer 3.1 (https://eu.idtdna.com/calc/analyzer) to evaluate self and hetero dimers formation; best annealing temperature was evaluated testing a range of temperature ranging from 54°C to 64°C; the efficiency of the reaction was calculated, through serial decimal dilutions of known concentrations of DNA, by using LineGene 9620 software (Bioer). Also in this case, all the experiments were carried out in triplicate.
Samples were examined for differences in gene expression using relative quantification in which relevant gene expression is normalized to a housekeeping gene using the (1) $2^{-\Delta\Delta C}$T method as reported by Livak and Schmittgen, 2001.

(1) $2^{-\Delta\Delta CT}$

$\Delta\Delta C_T$: $(C_{T,\ target}-C_{T,\ control})_{Time\ x} - (C_{T,\ target}-C_{T,\ control})_{Time\ 0}$

Where: C_T: threshold cycle
target: gene of interest
control: housekeeping gene
$Time_x$: condition tested (M17-G with antibiotic)
$Time_0$: basal condition (M17-G)

Table 1. *Lactococcus garvieae* strains used in this study, and their origin.

L. garvieae strains	Source of isolation
[a]DSMZ 20684[T]	Cow with mastitis
G18;G27	Cow milk
TB25; G01; G03; G07; G36, G6;	Italian cheese
Far1	Wheat flour
Br1; Br2; Br3; Br4; Sed2; Ins1; Ins2	Vegetables
V32; V61; V63; V79; Lg9; Lg19; Lg23; Lg28	Diseased fish
Po1; Tac1; Tac2; Bov1; Bov3	Raw meat
I113;Smp3; Sml1; Sml3	Meat products
LgUma10	Human faeces
Usa1÷Usa11	Sand from cow litter, dairy farm 1
Usa12÷Usa23	Sand from cow litter, dairy farm 2
Usa24	Straw from cow litter, dairy farm 2

Table 2. Composition of API 50 CHL medium

Ingredients	g/L
Polypeptone	1.0
Yeastextract	0.5
Tween 80	100 µL
K_3PO_4	0.2
Sodium acetate $3H_2O$	0.5
Ammoniumcitrate	0.2
$MgSO_4$ $7H_2O$	0.02
$MnSO_4$ $4H_2O$	5.0 mg
Bromocresolpurple	17.0
Distilled water	1 L

Table 3: List of antibiotics used in this study, grouped into families; zone diameters are reported, S: sensitive, I: intermediate and R: resistant.

Families	Antibiotics	Zone diameter (mm)		
		S	I	R
β-lactam	Ampicillin (10 µg)[a]	≥17	-	≤16
	Cefazolin (30 µg)[b]	≥24	22-23	≤21
	Penicillin (10 U)[a]	≥15	-	≤14
Tetracycline	Tetracycline (30 µg)[a]	≥19	18-15	≤14
Amino acid	Chloramphenicol (30 µg)[a]	≥18	17-13	≤12
Glycopeptide	Vancomycin (30 µg)[a]	>16	-	<16
Macrolides	Erythromycin (15 µg)[a]	≥23	22-14	≤13
Lincosamides	Clindamycin (2 µg)[b]	≥19	16-18	≤15
Aminoglycoside	Gentamycin (250 µg)[a]	-	-	<16
Chinolon	Ciprofloxacin (5 µg)[a]	≥21	20-16	≤15
	Nitrofurantoin (300 µg)[a]	≥17	16-15	≤14
	Norfloxacin (10 µg)[a]	≥17	16-13	≤12
Diaminopyrimidine	Trimethoprim (5 µg)[c]	≥50		≤21
Diaminopyrimidine + sulfonamide	Trimethoprim + sulfamethoxazole (1,25 µg + 23,75 µg)[d]	≥18		≤15
Oxazolidinone	Linezolid (30 µg)[a]	≥23	22-21	≤20
Rifamycin	Rifampicin (5 µg)[a]	≥20	19-17	≤16
Phosphonic	Fosfomycin (200 µg)[a]	≥16	15-13	≤12

[a] CLSI resistance zone diameter for Enterococci spp. (Clinical and Laboratory Standard Institute, 2006).
[b] CLSI resistance zone diameter for antibiotic cefepime for other *Streptococcus* spp. (2006).
[c] EUCAST resistance zone diameter for Enterococci spp. (European Committee on Antimicrobial Susceptibility Testing, 2011).
[d] EUCAST resistance zone diameter for *Streptococcus* spp. (2011).

Table 4. PCR primers and conditions used in this study.

Gene	Primer pair (5’-3’)	Annealing temperature (°C)	Amplicon (bp)
scrA	F:ATGAACCACAAACAAGTTGC R:TAGCCTGCTTTAATGGAGAT	54	1334
P-β-gal	F:GCTACAGCTGCTTATCAAGCA R:AGGATAAATAATCCAATCCCA	58	1014
tetM	F:GTTAAATAGTGTTCTTGGAG R:CTAAGATATGGCTCTAACAA	56	656
tetS	F:CATAGACAAGCCGTTGACC R:ATGTTTTTGGAACGACAGAG	59	667
hly	F:ATGGAAAAGCCAGCTTCTCG R:AACCATAGATGGAGAACCAC	58	640
BOXA1R	CTACGGCAAGGCGACGCTGACG	48	
M13	GAGGGGTGGCGGTTCT	38	

Table 5. RT-PCR primers and conditions used in this study.

Gene	Primer pair (5’-3’)	Annealing temperature (°C)	Amplicon (bp)
folP	F:GTAAGTCTTCTCGCCCAGG R:TGCCCGCTCGACCATTATTC	58	240
folA	F:CGTGTGCTTCCTGGACGTAT R:GGAACATCTCTGATCCGC	58	141
gapC	F:TATCGGTCGTCTTGCTTTCC R:TCAGCCCAGTTGATGTTAGC	58	225

RESULTS AND DISCUSSION

Carbon source utilization

The ability of the strains tested to metabolize lactose and sucrose were evaluated in order to find a possible discrimination between European and American isolates.
In table 6, these physiological properties with the presence of specific genes (*P-β-gal*, phospho-β-galactosidase and *scrA,* phospho-transferase involved, with different mechanisms, in their metabolisms) is reported.

Table 6. Lactose and sucrose utilization and related genes involved for different *L. garvieae* strains. +: utilization, presence; -: not utilization, absence.

***L. garvieae* strains**	**Lactose**		**Sucrose**		**Cluster[a]**
	Utilization	*P-β-gal*	Utilization	*scrA*	
Usa1	+	+	-	-	/
Usa2	+	+	-	-	/
Usa3	+	+	-	-	/
Usa4	+	+	-	-	/
Usa5	+	+	-	-	/
Usa6	+	+	-	-	/
Usa7	+	+	-	-	/
Usa8	+	+	-	-	/
Usa9	+	+	-	-	/
Usa10	+	+	-	-	/
Usa11	+	+	-	-	/
Usa12	+	+	-	-	/
Usa13	+	+	-	-	/
Usa14	+	+	-	-	/
Usa15	-	-	-	-	/
Usa16	+	+	-	-	/
Usa17	+	+	-	-	/
Usa18	+	+	-	-	/
Usa19	+	+	-	-	/
Usa20	+	+	-	-	/
Usa21	+	+	-	-	/
Usa22	+	+	-	-	/
Usa23	+	+	-	-	/
Usa24	+	+	-	-	/
DSMZ 20684^T	-	-	-	-	A

Continued on next page

***L. garvieae* strains**	**Lactose**		**Sucrose**		**Cluster[a]**
	Utilization	*P-β-gal*	Utilization	*scrA*	
Smp3	+	+	-	-	A
Po1	-	-	-	-	A
Tac2	-	-	+	-	A
Bov3	-	-	-	-	A
Sa113	-	-	-	-	A
Ins1	-	-	-	-	A
Lg9	-	-	-	-	A
Lg19	-	-	-	-	A
Sed2	-	-	+	+	B
Br3	-	-	+	+	B
Br4	-	-	+	+	B
Far1	-	-	+	+	B
G27	+	+	+	+	B
G07	+	+	+	+	B
TB25	+	+	+	+	B
G01	+	+	+	+	B
V63	-	-	+	+	B
V79	-	-	+	+	B
LgUma10	+	+	+	+	/

[a]as reported by Ferrario et al., 2013

All strains coming from United States of America (USA) isolated from dairy farm are able to metabolize lactose, with the exception of strain Usa15, instead European strains using lactose were isolated from cheese (G07 and TB25), meat product (Smp3) and human faeces (LgUma10). At genomic level, all strains lactose positive carry one of the genes involved in lactose utilization. *P-β-gal* was already proposed as a genomic marker distinguishing dairy and fish strains (Fortina et al., 2009); the results obtained in this study also show that not only microorganisms specifically isolated from cheese have this marker, but other strains isolated from environments which can be linked to dairy one (cow litter, meat products and, through food, human faeces) can possess this characteristic.
Regarding sucrose, none strains from USA are able to use this carbon source and this result agrees with the absence of one of the genes (*scrA*) related to its utilization. European strains are separated in two groups in function of their ability to use or not this carbohydrate. The presence of gene *scr*A was demonstrated in all sucrose positive strains, with the exception of strain Tac2 for which an alternative pathway could be involved.
Since previous experiments have confirmed that sucrose utilization is a phenotypic characteristic that discriminate strains between the two subgroups in which *L. garvieae* population can be separated (Ferrario et al., 2013), American strains should belong to group A and consequently have genome similarity with other strains from cow milk, cow with mastitis, diseased fish, vegetables, raw meat and meat products.

Antibiotic resistance

Sensitivity of 20 European and 24 American strains of *L. garvieae* toward 17 different antibiotics, belonging to 13 families was carried out. The results, reported in table 7, compare European and American isolates.

Table 7.Sensitivity of strains of *L. garvieae* to different antibiotics. S: sensitive, I: intermediate and R: resistant; for zone diameter breakpoints see table 3.

Antibiotics	**European strains**	**American strains**
Ampicillin (10 µg)	**S** 20/20	**S** 24/24
Cefazolin (30 µg)	**S** 19/20 **R** 1/20	**S** 20/24 **R** 3/24 **I** 1/24
Penicillin (10 U)	**S** 20/20	**S** 24/24
Tetracycline (30 µg)	**S** 16/20 **R** 4/20	**S** 13/24 **I** 8/24 **R** 3/24
Chloramphenicol (30 µg)	**S** 20/20	**S** 24/24
Vancomycin (30 µg)	**S** 20/20	**S** 24/24
Erythromycin (15 µg)	**S** 19/20 **R** 1/20	**S** 24/24
Clindamycin (2 µg)	**S** 1/20 **R** 19/20	**R** 24/24
Gentamycin (250 µg)	**S** 20/20	**S** 24/24
Ciprofloxacin (5 µg)	**S** 15/20 **I** 5/20	**S** 24/24
Nitrofurantoin (300 µg)	**S** 19/20 **I** 1/20	**S** 24/24
Norfloxacin (10 µg)	**S** 8/20 **I** 11/20 **R** 1/20	**S** 14/24 **I** 8/24 **R** 2/11
Trimethoprim (5 µg)	**R** 20/20	**R** 24/24
Trimethoprim + sulfamethoxazole (1,25 µg + 23,75 µg)	**R** 20/20	**R** 24/24
Linezolid (30 µg)	**S** 20/20	**S** 24/24
Rifampicin (5 µg)	**R** 20/20	**R** 24/24
Fosfomycin (200 µg)	**S** 1/20 **I** 1/20 **R** 18/20	**S** 2/24 **I** 4/24 **R** 18/11

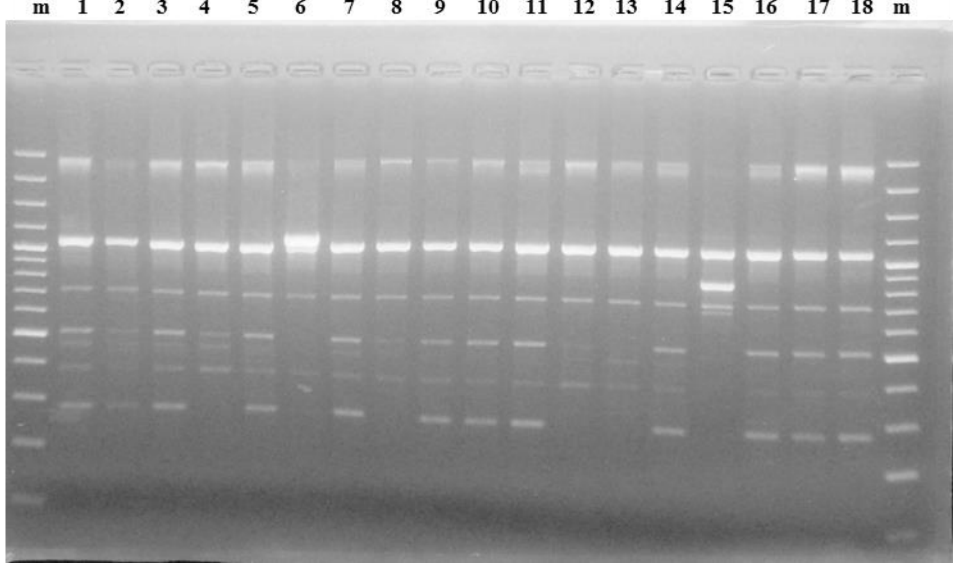
The results reported showed a similar antibiotic sensitivity profile between the strains coming from the two continents. Specifically, all strains are susceptible to many of the tested antibiotics, and resistant to clindamycin, rifampicin, fosfomycin and trimethoprim-sulfamethoxazole. A degree of variability was observed for tetracycline.

Genomic fingerprinting

Genomic fingerprinting is a technique that allows finding differences at the species and strains-level.

24 American strains and some European strains, representative of different ecological niches (dairy, meat, fish and vegetables), were typed by REP and RAPD-PCR and the results reported in figure 1 and 2.

a)



b)

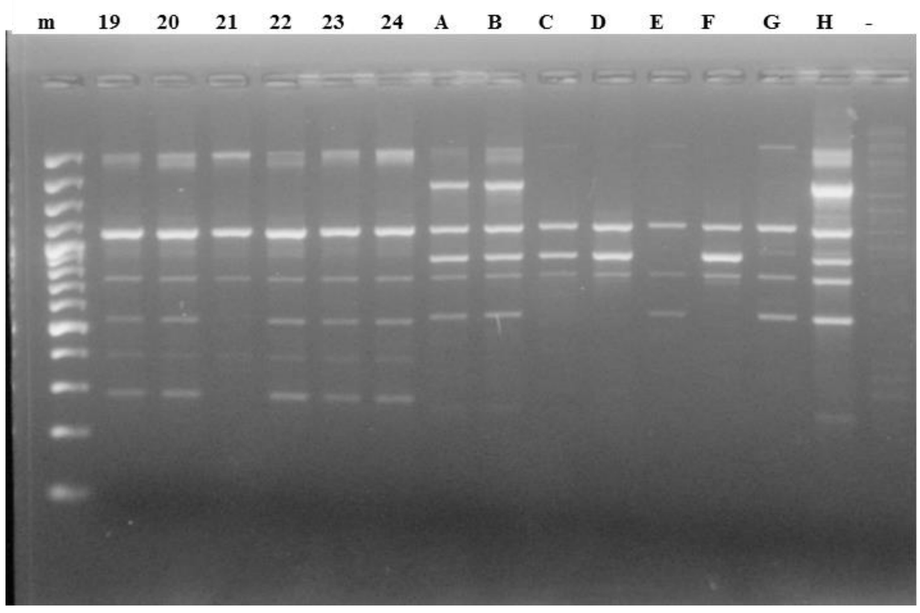
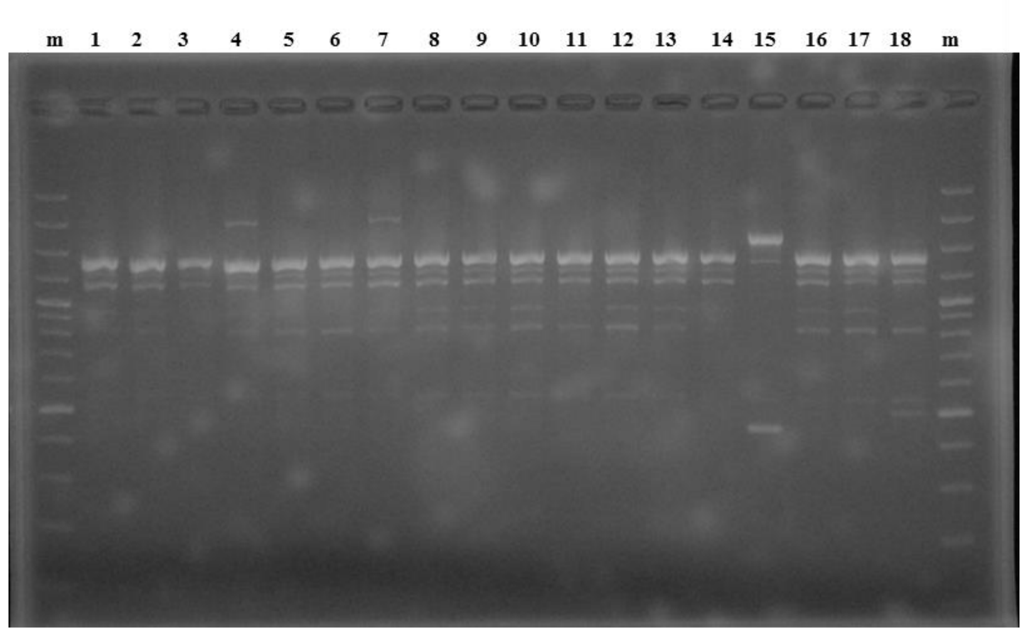


Figure 1 a and b. REP-PRC typing of different *L. garvieae* strains. m: molecular marker (GeneRuler, DNA Ladder Mix, Fermentas), Usa1-Usa24: American strains of *L. garvieae*; A: G07; B: TB25; C: Smp3; D: Lg9; E: Br3; F: Ins1; G: Sed2; H: LgUma10; -: negative control.

a)



b)

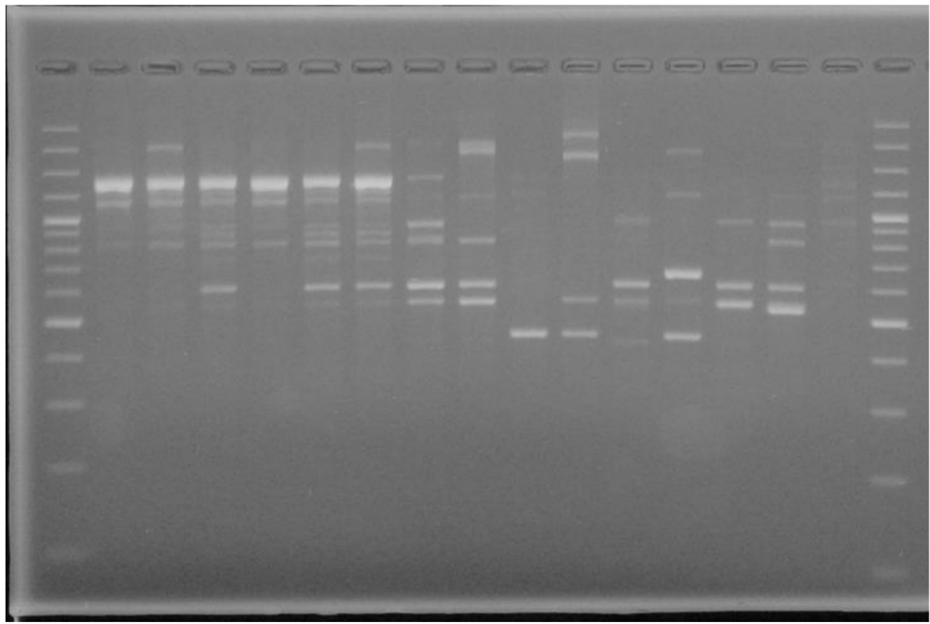


Figure 2 a and b. RAPD-PRC typing of different *L. garvieae* strains. m: molecular marker (GeneRuler, DNA Ladder Mix, Fermentas), Usa1-Usa24: American strains of *L. garvieae*; A: G07; B: TB25; C: Smp3; D: Lg9; E: Br3; F: Ins1; G: Sed2; H: LgUma10; -: negative control.

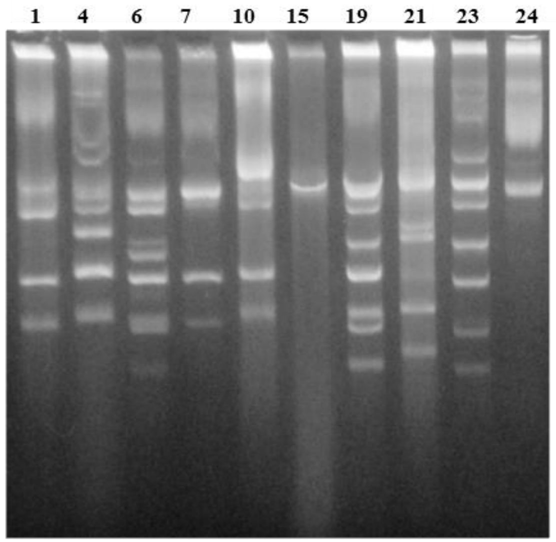
Combined results confirmed the validity of these techniques in order to discriminate the strains within the species. American and European isolates typed differently and this could suggest different evolutionary histories. Analyzing each group, American strains have a more conserved profile, with the exception of strain Usa15, while European isolates show different profiles, in agreement with their different niches of isolation.

The results obtained both at phenotypic (carbon source utilization and antibiotic sensitivity) and at genotypic level, highlight a high variability within the species, not only at geographical level, but also in relation with the ecological niche of isolation.
For this reason, genome plasticity and an overview on the potential pathogenicity of the species were deeper investigated by searching temperate bacteriophages (see chapter 4) and plasmids dissemination (this chapter). These investigations were carried out on selected American (Usa1, 4, 6 7, 10, 15, 19, 21, 23 and 24) and European strains (G03, G6, G18, G36, Br1, Br2, Lg23, Lg28, V32, V61, Ins2, Sml1, Sml3, Tac1, Bov1), chosen for their genomic differences.

Plasmids dissemination and Southern analysis

Plasmids are extra-chromosomal DNA that plays an important role in genome plasticity since they can be exchanged between cells, improving gene dissemination and adaptation. Selected strains of *L. garvieae* were investigated for the presence of plasmids by alkaline DNA extraction.
In figure 3 the results of the experiments are reported.

a)



b)

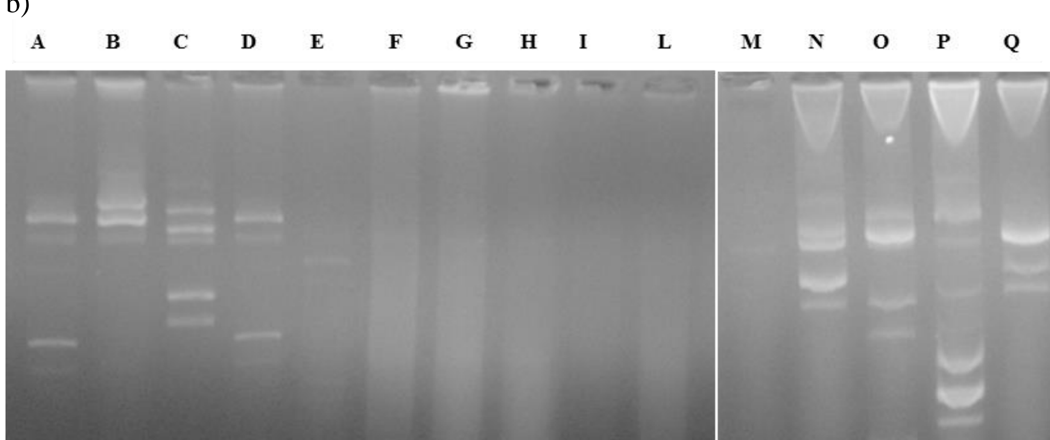


Figure 3 a and b: disseminations of plasmids in *L. garvieae* strains. Ordinal numbers: American strains; A: G03; B: G6; C: G18, D: G36; E: Br1; F: Br2; G: Lg23; H: Lg28; I: V32; L: V61; M: Ins2; N: Sml1; O: Sml3; P: Tac1; Q: Bov1.

Interestingly, all American isolates showed a complex profile, constituted by plasmids of different molecular size, with the exception of strain Usa15. On the contrary, not all European isolates had plasmids: are positive, strains from dairy (raw milk and cheese), and meat niches, while seem plasmid-free strains from vegetables and diseased fishes. Also in this case, the dissemination of plasmids highlights as milk, cow bedding and meat product are linked.

L. garvieae is known to be a pathogen in aquaculture sectors (Vendrell et al., 2006) and recently it has gained attention as opportunistic human pathogen; however, its pathogenicity is not well understood yet.
Since plasmids are able to mobilize genes, we investigated the presences of some genes involved in pathogenicity and their localization at plasmid level.
The study was carried out on *hly* gene codifying a hemolysin (Baida et al., 1996) and two among the genes conferring resistance to antibiotic tetracycline, *tet*M and *tet*S (Teuber at al., 1999 and Huang et al., 2015). Phospho-β-galactosidase gene was used as control.
In table 8 are reported the results of these investigations. Firstly, all strains tested show the presence of *hly* gene. Its localization at chromosomal level (data not shown) indicates that this gene is part of the core genome of the species. Regarding the presence of genes related to tetracycline resistance, it is interesting to note that, according to phenotypic results, the susceptible strains lack both genes, the intermediate strains show the presence of only one of the two genes tested, while the resistant strains harbor both *tetM* and *tetS.*
Southern hybridization experiments (figure 4) indicate that the localization of *tet* genes is at plasmid level. Even if the simultaneous presence also at chromosomal level cannot be excluded, the *tet* genes seems harbored on plasmids of high molecular weight, potentially involved in conjugative transfers.
P-β-gal seems localized both at plasmid al chromosomal level. It is known that together with genes for casein degradation, the genes involved in lactose utilization are essentials for the survival of dairy lactococci in milk and they can also be plasmid-encoded (Wegmann et al., 2012); this probably illustrates the importance of genetic recombination events in lactococci, for their adaptation to different environments, comprising the dairy niche.

Table 8. Antibiotic resistance and *tet*M, *tet*S, P-β galactosidase and haemolysin detection by PCR. +: utilization/presence; -: not utilization/absence. S: sensitive, I: intermediate and R: resistant.

Strains	Tetracycline resistance	*tet*M	*tet*S	*hly*	*P-β gal*
Usa11	S	-	-	+	+
Usa4	I	+	-	+	+
Usa6	I	-	+	+	+
Usa7	I	+	-	+	+
Usa10	R	+	-	+	+
Usa15	S	-	-	+	-
Usa19	R	-	+	+	+
Usa21	S	-	-	+	+
Usa23	I	-	+	+	+
Usa24	S	-	-	+	+
G03	S	-	-	+	+
G6	S	-	-	+	+
G18	S	-	-	+	+
G36	S	-	-	+	+
Sml1	S	-	-	+	-
Sml3	S	-	-	+	-
Tac1	R	+	+	+	-
Bov1	R	+	+	+	-

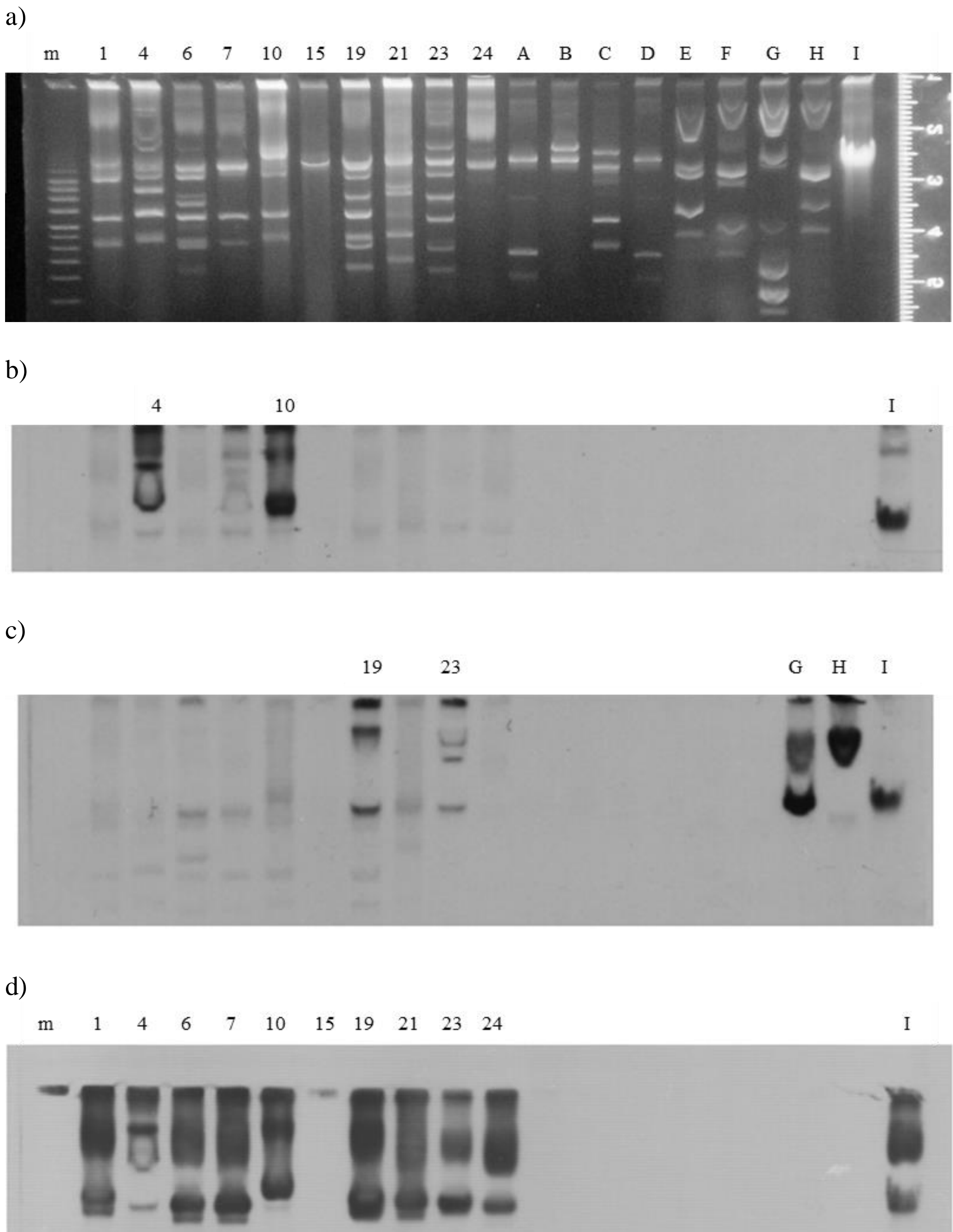


Figure 4 a, b, c and d. Southern analysis on *L. garvieae* strains. Pictures A: electrophoresis run of plasmids from *L. garvieae* strains. B: Southern of *tet*M. gene C: *tet*S gene. D: *phospho-β-gal* gene m: CCC molecular marker (DNA Ladder Supercoiled, Sigma-Aldrich); Ordinal numbers: American strains; A: G03; B: G6; C: G18, D: G36; E: Sml1; F: Sml3; G: Tac1; H: Bov1; I: chromosomal DNA from strain Usa7.

Resistance to sulfonamide compounds

For the analysis of the resistance mechanisms towards trimethoprim, sulfamethoxazole and their mix, two strains (representing the two geographical niches) TB25 (Italian cheese) and USA7 (American dairy farm) were taken into consideration for the determination of minimal inhibitory concentration (MIC), search for genes involved in the resistance and their localization. Moreover, gene expression experiments have been carried out.

<u>Minimal inhibitor concentration (MIC)</u>

Table 9 reports the MIC values determined through broth dilution method in M17-G, starting from a cell concentration of 10^5 UFC/mL.

Results showed a similar MIC values for the two strains in all the condition tested; according to Houvien et al., (1995) strains were resistant to trimethoprim (trmp) (threshold: 0,5 µg/mL), and also to sulfamethoxazole (sulfa) (threshold: 16 µg/mL), according to Enne et al., (2002).

The mix of two antibiotics revealed a MIC value lower than the single antibiotics, highlighting a stronger combined action; this characteristic makes these antibiotics suitable against microbial infections.

Table 9. MIC values for trimethoprim (trmp), sulfamethoxazole (sulfa) and their combination.

Strain	Trmp	Sulfa	Trmp +Sulfa
TB25 (Italian cheese)	12 µg/mL	1300 µg/mL	(3 + 60) µg/mL
USA7 (dairy farm)	12 µg/mL	1300 µg/mL	(3 + 60) µg/mL

A growth curve in presence and absence of sub-lethal concentration of antibiotics were carried out in order to identify the exponential phase of growth, for the subsequent gene expression experiments.

Sub lethal concentrations were chosen to be 10 µg/mL for trimethoprim, 1200 µg/mL for sulfamethoxazole and (2 + 40) µg/mL for trimethoprim plus sulfamethoxazole.

The resulting growth curve allowed to identify the time required to reach the exponential phase, reported in table 10.

Table 10. Incubation times required to reach exponential phase of growth.

Strain	Control without antibiotic	Trpm	Sulfa	Trmp+Sulfa
TB25	8 h	14 h	8 h	8 h
USA7	8 h	14 h	8 h	8 h

<u>Search of genes involved in antibiotics resistance and their expression</u>

In literature different genes involved in trimethoprim and sulfamethoxazole resistance are described (Liu et al., 2015). Regarding trimethoprim, the resistance can be given by *dhfrI* and/or *folA* genes. For sulfamethoxazole resistance, three different genes, named *sulI*, *sulII* and *sulIII* generally located on plasmids or transposons are described for Gram negative bacteria (Enne et al., 2002). Another resistance mechanism involves *folP* gene encoding a DHPS (dihydropteroate synthase).

The analysis of the available genomes of *L. garvieae* allowed verifying the absence of *sul* genes and the presence, in all genomes, of the genes *folA* and *folP* that are part of the complex gene cluster of the folic acid pathway.

RT-qPCR experiments carried out to evaluate the modulation of the expression of *folA* and *folP* did not underline an over-expression in presence of the related antibiotics, as expected (figure 5).

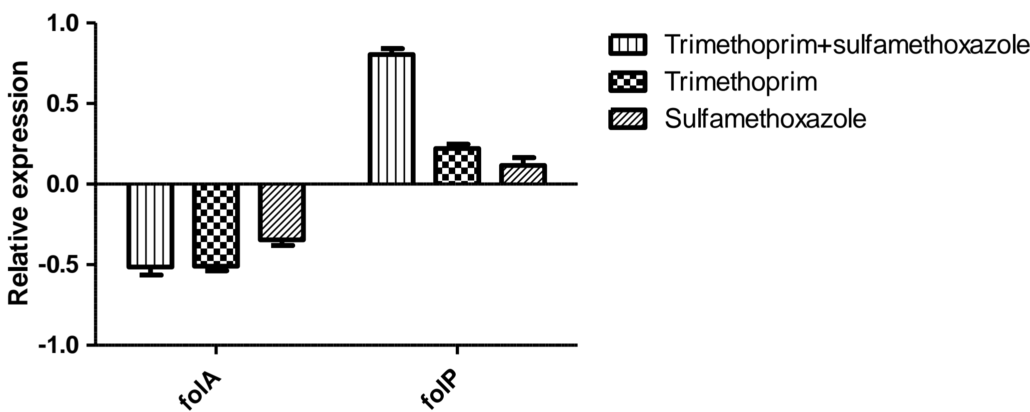


Figure 5. Gene expression of gene *folA* and *folP* in different conditions.

For these reasons, to better understand the mechanisms involved in antibiotic resistance, the metabolic pathway of folic acid, where *folP* and *folA* explicate their function, was deeper studied.

Folic acid pathway

Figure 6 reported the entire pathway of folic acid, adapted from De Crécy-Lagard et al. (2007).

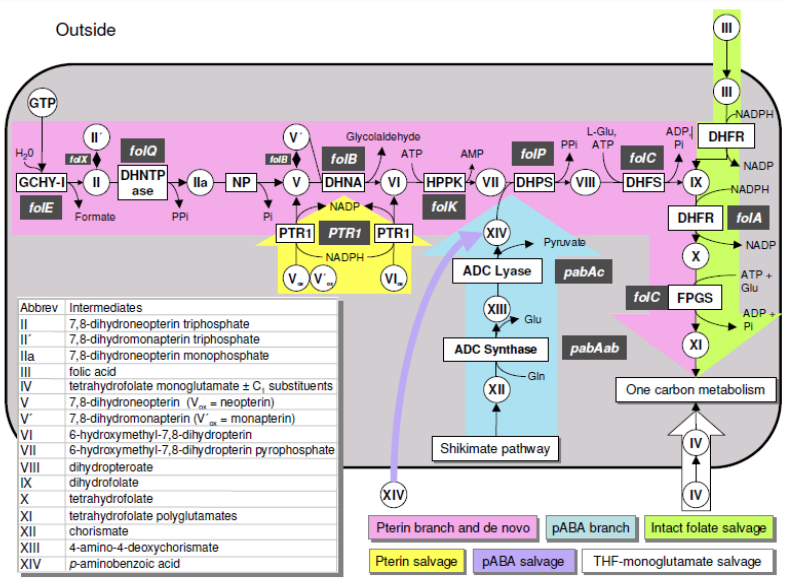
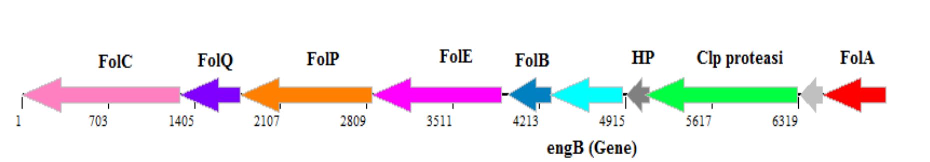


Figure 6. Folic acid pathway.

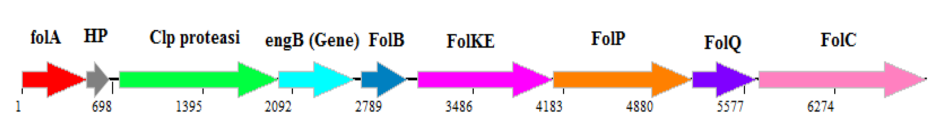
Folic acid synthesis is an essential pathway for most plants, fungi and bacterial cells because tetrahydrofolic acid (THF) is required in different essential activities such as synthesis of purinic bases and N-formylmethionine (de Crécy-Lagard et al., 2007).
Regarding figure 6, different metabolic ways are indicated: *de novo* folate synthesis (purple), folate from the environment (green) and two (yellow and blue) shorter ways whose products are intermediate of the *de novo* synthesis pathway.
Organisms possessing the purple way are able to synthesize acid folic from GTP (guanosine triphosphate), while its absence requires the acquisition of folic acid from the environment (for the cells) or from the diet (animals).
FolP and *folA* genes belong to purple and green way, respectively; inhibition of these two genes blocks the THF synthesis.

Since in literature any information are available about folic acid pathway in *L. garvieae*, the sequenced genomes of strains TB25 and ATCC49156 (complete genome) were searched for the presence of genes involved in folic acid pathway.
Results showed the presence of 8 genes linked to folic acid pathway; this led us to hypothesize firstly, the presence in *L. garvieae* of *de novo* way and secondly that the resistance to trimethoprim and sulfamethoxazole might be explained by mutation in *folP* and *folA* genes.

In figure 7 the genome organization of the folic acid pathway in *L. garvieae* is shown, in comparison with the gene cluster found in the phylogenetically close related *L. lactis*.



Lactococcus garvieae ATCC 49156 DNA, complete genome.



Lactococcus lactis subsp. lactis KF147, complete genome.

Figure 7. Folic acid pathway comparison between *L. garvieae* and *L. lactis*.

Table 11. Amino acid comparison of genes involved in folic acid pathway in *L. garvieae* and *L. lactis*.

Genes	*L. lactis/L. garvieae* ATCC49156	*L. lactis /L. garvieae* TB25
folA	63%	63%
folB	71%	71%
folKE	66%	65%
folP	61%	61%
folQ	65%	65%
folC	66%	66%
Clp protease	85%	84%
engB	82%	82%

The two lactococcal species show a similar gene cluster organization; only in *L. lactis* a gene *folKE* is present which is absent in *L. garvieae*.
Amino acid comparison reveals a high identity for all genes found in *L. lactis* and *L. garvieae,* with an average of 69% (table 11).

The search of this gene cluster in other microorganisms more studied, has led us to find some interesting information. *Bacillus anthracis*,a Gram positive, has a partial grouping of genes involved in folate biosynthesis, while in Gram negative pathogens (such as *Y. pestis* and *H. influenzae*) these genes are dispersed in different regions of the chromosomal DNA.

To complete the analysis of folate biosynthesis in *L. garvieae*, gene *folT* a specific folate transport protein involved in folate acquisition from the environment (green way in figure 6), was searched.

In *L. garvieae* the gene is present showing an aminoacid identity of 32% with the same protein in *Lactobacillus casei*; this species, together with *Lactobacillus rhamnosus* is well known to be auxotrophic for folic acid: the absence of genes involve in *de novo* way makes necessary the acquisition of folic acid from the environment.

<u>Sulfamidic antibiotics</u>

Trimethoprim and sulfamethoxazole belong to sulfamidic family of antibiotics and they have been using since thirty years both for animals and humans.

Both antibiotics interact with two enzymes (DHFR and DHPS, encoded by genes *folA* and *folP*) and replace the physiological substrate, stopping the folic acid pathway.

Since sulfamethoxazole inhibits an enzyme that physiologically occurs before than enzyme DHFR (purple way, figure 6), cell might by-pass this problem using the folic acid from the environment. In *L. garvieae* a high MIC value for sulfamethoxazole (1.3 mg/mL), together with the presence of a specific folate transport protein (*folT*) could confirm this hypothesis.

Different authors state that the resistance to sulfamidic is due to specific modifications in the amino acid of the catalytic site of the enzyme (Liu et al., 2015; Pei et al., 2006).

Enzyme DHPS is composed by different α-helixes and β-sheets, and the amino acid composition of the primary structure shapes tridimensional loops (LOOP1, LOOP2, LOOP5 and αLOOP7), indispensable for substrate recognition (Yun et al., 2012). Moreover in *H. influenzae*, Enne et al., (2002) correlate the resistance of sulfamethoxazole to different point mutations and the presence of extra bases due to a chromosomal insertion.

An alignment of DHPS genes found in *L. garvieae*, *L. lactis, Y. pestis, S. aureus*, *B. anthracis*, in a sensitive strain of *H. influenzae* and one highly resistant for six bases insertion, was carried out in order to deeper study DHPS in *L. garvieae*.

Results show that LOOPs present in DHPS of *L. garvieae* have conserved nucleotides compared to other microorganisms and this confirms the role of this enzyme in folic acid pathway. Moreover, the presence of substituted amino acids (Pro-64 and Ala-243) described to be involved in resistance to antibiotics, together with the results of gene expression, confirms the role of this modified enzyme in conferring antibiotic resistance towards sulfamethoxazole.

CONCLUSIONS

In this chapter, through different approaches:

- *L. garvieae* isolates coming from ecological niches not yet investigated (sand and straw from cow litter) were studied;
- Their phenotypic and genotypic comparison with European strains coming from other niches reveals some differences that could be related to different adaptation mechanisms;
- The plasmid profile of the strains has been investigated. Main results are related to the presence, at plasmid level, of tetracycline resistance genes;
- The folic acid pathway in *L. garvieae* has been, for the first time, investigated and for specific genes of the cluster, a relationship with the sulfamidic resistance has been found.

The availability of new strains, some of which with interesting phenotypic and genotypic features, led us to investigate the impact of temperate bacteriophages both in evolution and spreading of factors (not necessarily virulent) that might have allowed the adaptation of *L. garvieae* to different environments.

Related publication: Eraclio G., Ricci G., Quattrini M., Moroni P., Fortina M.G. (2015). *Virulence-related genes in* Lactococcus garvieae *coming from different ecological niches and geographical areas*. Submitted FEMS Microbiology Letters.

REFERENCES

1. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). *Basic local alignment search tool.* J Mol Biol. 215:403-410.

2. Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997). *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res. 17:3389-3402.

3. Aziz R.K., Bartels D., Best A.A., DeJongh M., Disz T., Edwards R.A., Formsma K.., Gerdes S., Glass E.M., Kubal M., Meyer F., Olsen G.J., Olson R., Osterman A.L., Overbeek R.A, McNeil L.K., Paarmann D., Paczian T., Parrello B., Pusch G.D., Reich C., Stevens R., Vassieva O., Vonstein V., Wilke A., Zagnitko O. (2008). *The RAST Server: rapid annotation using subsystems technology*. BMC genomics. 8:9-75.

4. Baida G.E., Kuzmin N.P. (1996). *Mechanism of action of hemolysin III from* Bacillus cereus. Biochim Biophys Acta. 1284:122-124.

5. Borgo F., Ferrario C., Ricci G., Fortina M.G. (2012). *Genotypic intraspecies heterogeneity of* Enterococcus italicus: *data from dairy environments.* J Basic Microbiol. 52:1-9.

6. Clinical and Laboratory Standards Institute. (2006). Performance standards for antimicrobial susceptibility testing: sixteenth informational supplement M100-S16, vol. 26, no. 3. Clinical and Laboratory Standards Institute, Wayne, Pa.

7. de Crécy-Lagard V., El Yacoubi B., de la Garza R.D., Noiriel A., Hanson A.D. (2007). *Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations*. BMC Genomics. 23;8:245.

8. De Urraza P.J., Gómez-Zavaglia A., Lozano M.E., Romanowski V., De Antoni G.L. (2000). *DNA fingerprinting of thermophilic lactic acid bacteria using repetitive sequence-based polymerase chain reaction.* J Dairy Res. 67:381-392.

9. Enne V.I., King A., Livermore D.M., Hall L.M. (2002). *Sulfonamide resistance in* Haemophilus influenzae *mediated by acquisition of sul2 or a short insertion in chromosomal* folP. Antimicrob Agents Chemother. 46:1934-1939.

10. EUCAST, European committee on antimicrobial susceptibility testing. (2011). European society for of clinical microbiology and infectious diseases.

11. Ferrario C., Ricci G., Milani C., Lugli G.A., Ventura M., Eraclio G., Borgo F., and Fortina M.G. (2013). Lactococcus garvieae*: where is it from? A first approach to explore the evolutionary history of this emerging pathogen*. PLoS One. 8:e84796.

12. Ferrario C. (2012). Lactococcus garvieae *and* Morganella moorganii: *two bacteria models to study quality and safety of fish products.* PhD thesis. Univeristy of Milan, Italy.

13. Fortina M.G., Ricci G. and Borgo F. (2009). *A study of lactose metabolism in* Lactococcus garvieae *reveals a genetic marker for distinguishing between dairy and fish biotypes.* J Food Prot. 72:1248-1254.

14. Huang M., Qi F., Wang J., Xu Q., Lin L. (2015). *Changes of bacterial diversity and tetracycline resistance in sludge from AAO systems upon exposure to tetracycline pressure.* J Hazard Mater. 298:303-309.

15. Huovinen P., Sundström L., Swedberg G., Sköld O. (1995). *Trimethoprim and sulfonamide resistance*. Antimicrob Agents Chemother 39:279-289.

16. Liu Z., Zhang Z., Yan H., Li J., Shi L. (2015). *Isolation and molecular characterization of multidrug-resistant* Enterobacteriaceae *strains from pork and environmental samples in Xiamen, China*. J Food Prot. 78:78-88.

17. Livak K.J., Schmittgen T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408.

18. Pei R., Kim S.C., Carlson K.H., Pruden A. (2006). *Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG).* Water Res. 40:2427-2435.

19. Ricci G., Fortina M.G. (2006). *Characterization of* Lactobacillus helveticus *strains isolated from cheeses by distribution studies of insertion sequences.* Int J Food Microbiol. 112:112-119.

20. Rossetti L., Giraffa G. (2005). *Rapid identification of dairy lactic acid bacteria by M13-generated, RAPD-PCR fingerprint databases*. J Microbiol Methods. 63:135-144.

21. Teuber M., Meile L., Schwarz F. (1999). *Acquired antibiotic resistance in lactic acid bacteria from food.* Ant. Van Leeuw. 76:115-137.

22. Vendrell D., Balcázar J.L., Ruiz-Zarzuela I., de Blas I., Gironés O., Múzquiz J.L. (2006). Lactococcus garvieae *in fish: a review.* Comp Immunol Microbiol Infect Dis. 29:177-198.

23. Versalovic J., Scneid M., de Bruijin F.J., Lupski J.R. (1994). *Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction*. Meth Mol Cell Biol. 5:25-40.

24. Wegmann U., Overweg K., Jeanson S., Gasson M., Shearman C. (2012). *Molecular characterization and structural instability of the industrially important composite metabolic plasmid pLP712.* Microbiol. 158:2936-2945.

25. Yun M.K., Wu Y., Li Z., Zhao Y., Waddell M.B., Ferreira A.M., Lee R.E., Bashford D., White S.W. (2012). *Catalysis and sulfa drug resistance in dihydropteroate synthase*. Science 335:1110-1114.

26. Zlotkin A., Eldar A., GhittinoC. and Bercovier H. (1998). *Identification of* Lactococcus garvieae *by PCR*. J Clin Microbiol. 36: 983-985.

4. CHARACTERIZATION OF TEMPERATE BACTERIOPHAGES INFECTING *LACTOCOCCUS GARVIEAE* REVEALS NEW INSIGHT INTO THE EVOLUTION OF THE SPECIES

The impact of temperate bacteriophages in genome plasticity of *Lactococcus garvieae* is presented in this chapter; the lack of extensive information led us, firstly, to evaluate their dissemination among different strains coming from different niches and then, through genome sequencing and comparison, evaluate their influence on the adaptation of the species.

MATERIAL AND METHODS

Induction assay and transmission electron microscope

Lactococcus garvieae strains (table 1) were grown statically at 30 °C in 5 mL of M 17 broth (Pronadisa, Spain) with 1 % glucose (GM17) until an optical density of 0.3 OD_{600} was reached. Then 5 µg/mL of mitomycin C (Sigma, Italy) was added and changing of optical density was measured (in quadruplicate) every 30 min over 5 hours using a BioTek PowerWave XS2 spectrophotometer (USA). Typical induction curve was characterized by an initial increase followed by a sharp reduction in optical density compared to the control.

The presence of temperate phages was confirmed by transmission electronic microscope (TEM). Lysate was filtered through a 0.45-µm syringe filter and 1 mL was centrifuged for 1 h at 4°C (24,000 × *g*). The supernatant (approximately 800 µL) was gently discarded. The remaining lysate was diluted twice by the addition of 800 µL of ammonium acetate (0.1 M, pH 7.5) and then centrifuged (1 h at 24,000 × *g* and 4°C). The phage solution (10 µL) was mixed with the stain (10 µL of 2% uranyl-acetate) and deposited on a nickel Formvar-carbon-coated grid (Pelco International). The liquid was removed after 1 min by touching the edge of the grid with blotting paper. Phage morphology was observed at 80 kV using a JEOL1230 transmission electron microscope at the Platforme d'Imagerie Moléculaire et Microscopie of the Université Laval. Capside size as well as tail length and width were determined by measuring at least 10 phage specimens.

DNA extraction and sequencing

DNA for restriction profile was isolated by using a previously described protocol (Moineau et al., 1994), with some modifications: DNAse was inactivated at 65°C; to increase DNA release from the capsid 200 µL of SDS was added with 20 µL of proteinase K (stock solution 20 mg/mL) and samples were incubated at 37°C for 15 min and then at 60°C for 30 min. Restriction endonucleases (Roche Diagnostics, Germany) were used as recommended by the manufacturer. The DNA fragments were separated in 0.8% agarose gel in 1X Tris-acetate-EDTA buffer and visualized by UV photography.

DNA for sequencing was extracted from 90 mL of induced lysate adding 10% polyethylene glycol (8000) and 0.6 M sodium chloride, followed by a centrifugation at 18,000 rpm (Beckman) for 1 h at 4°C to increase phage pellets. Pellet was then resuspended in 1 mL of phage buffer (10 mM Tris-HCl pH 7.4, 100 mM NaCl, 10 mM MgSO4) and lysis was completed following the modified protocol above mentioned. DNA was then purified using UltraClean^{TM} Microbial DNA Isolation Kit (MO BIO Laboratories, Inc. USA) to eliminate traces of interfering for the next sequencing step.

To determine the genome of temperate phage TB25, a sequencing library was first prepared with the Nextera XT DNA Sample Prep Kit (Illumina) according to the manufacturer's instructions. The library was sequenced using a MiSeq system (Illumina). De novo assembly was performed with ABySS v1.5.2 assembler (parameters of k:31, aligner: bwamem, b:1000000, p:0.95, s:500 and n:10) and with CLC v7 (with the follow parameters word size:64, minimum contig length: 500", length fraction: 0.9, similarity fraction: 0.95, mapping mode: map reads back to contigs and update contigs: yes).

Phage genome analysis and comparison

Nucleotide sequences of the fourteen *L. garvieae* genomes deposited in the NCBI database (table 2) (http://www.ncbi.nlm.nih.gov/) were searched for the presence of temperate bacteriophages using prophage finder (Bose and Barber, 2006) and the results reinforced with PHAST (PHAge Search Tool) (Zhou et al., 2011). Sequences of at least 30 kb in length together with the presence of genes involved in integration, DNA replication and structural functions were indicative of complete prophages. Homology searches were performed using BLASTn and BLASTp (Altschul et al., 1990, 1997).

Temperate phage genomes which did not show any similarity with other phages in GenBank, were undergone a deeper studies. Open reading frame (ORF) prediction was carried out using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and RAST Server (Aziz et al., 2008). By visual inspection, an ORF was considered valid only if its starting codon was AUG, UUG or GUG and possessed at least 30 amino acids (aa). The research of a ribosomal binding site (RBS) similar to standard Shine-Dalgarno sequence (AGGAGA) was also carried out. Functions and domains were attributed by comparison of the translated products with BLASTp (Altschul et al., 1997); PSI-BLAST and InterProScan at EMBL-EBI (http://www.ebi.ac.uk/) were respectively used to search for more distant homologous proteins and conserved domains when any significant similarity was found by BLAST searches.

Theoretical molecular masses (MM) and isoelectric points (pI) of the phage proteins were obtained using the ProtParam tool (http://web.expasy.org/protparam/). tRNAs were identified using the tRNAscan-SE server (Lowe and Eddy, 1997) and confirmed using ARAGORN program (Laslett and Canback, 2004). Virulence Factor Databases (Chen et al., 2012) together with DBETH (Chakraborty et al., 2012) were used to search virulent factors.

Table 1. *L. garvieae* strains used in this study, and their origin.

***L.* garvieae strains**	**Source of isolation**
[a]NBRC 100934[T]	Cow with mastitis
G18;G27	Cow milk
TB25; G01; G03; G07; G36, G6;	Italian cheese
Far 1	Wheat flour
Br1; Br2; Br 3; Br4; Sed 2; Ins1; Ins2	Vegetables
V32; V61; V63; V79; Lg9; Lg19; Lg23; Lg28	Diseased fish
Po1; Tac1; Tac2; Bov1; Bov3	Raw meat
I113;Smp3; Sml1; Sml3	Meat products
LgUma10	Human feces
Cf1; Cf4; Cf6; Cf7; Cf10; Cf15; Cf19; Cf21; Cf23; Cf24	Cattle farms

[a]NBRC, NBRC NITE Biological Resource Center, Department of Biotechnology, National Institute of Technology and Evaluation, Kisarazu, Chiba, Japan.
[T] type strain of the species.

Table 2. *L. garvieae* genomes analysed in this study.

L. garvieae strain	Source	Genome status (scaffolds)	Accession Number
21881	Human blood	Draft (91)	AFCC00000000.1
8831	Diseased rainbow trout	Draft (87)	AFCD00000000.1
ATCC 49156	Diseased yellowtail	Complete	NC_015930
LG2	Diseased yellowtail	Complete	NC_017490
UNIUD 074	Diseased rainbow trout	Draft (25)	AFHF00000000.1
Lg9	Diseased rainbow trout	Draft (140)	AGQY00000000.1
Tac2	Turkey	Draft (97)	AMFE00000000.1
IPLA 31405	Cow milk	Draft (23)	AKFO00000000.1
I113	Meat products	Draft (49)	AMFD00000000.1
DCC43	Mallard duck intestines	Draft (68)	AMQS00000000.1
TRF1	Snake fecal material	Draft (112)	AVFE00000000.1
NBRC 100934/DSMZ 20684	Cow mastitis	Draft (56)	NZ_BBJW00000000.1
Lg-ilsanpaik-gs201105	Human cholecystitis	Draft (53)	NZ_JPUJ01000000
M14	Unknown	Draft (13)	NZ_CCXC01000000

RESULTS AND DISCUSSION

Phage induction

The presence of temperate bacteriophages in 45 strains of *L. garvieae* was assessed using mitomycin-C. Typical growth curves indicative of lysogeny were exhibited by 22 strains out of the 45 ecotypes (approximately with 50% of incidence): 4 coming from vegetables, 3 from Italian cheeses, 6 from cattle farms, 2 meat isolated, 1 from cow milk, 5 from diseased fishes and 1 isolated from raw meat.

Transmission electronic microscope confirmed the presence of temperate bacteriophages in 20 samples out 22 tested, since two strains (both isolated from diseased fish) did not show any phage particles.

Since morphological analysis revealed that all phages had a non-contractile tails, they were attributed to *Siphoviridae* family (table 3). Capside diameter, tail width and length measures were taken and allowed the separation of phages into two main morphotypes (MT) (I and II), the second with three subgroups (sb) (A, B and C).

The three phages belonging to MT I, all isolated from vegetables, show dimensions (capside diameter and tail length) higher that those characterizing the 8 species in which *Siphoviridae* family can be separated (Deveau et al., 2006).

MT II grouped phages with morphological features on average with other lactococcal phages. Subgroups A, B and C discriminated phages on the base of differences of tail length of 202, 119 and 222 nm, respectively. Each subgroup is characterized by a different number of host strains, (ranging from 3 to 8), also environmentally different (diseased fish, cheese, raw meat and vegetables) as for sb C. To confirm differences among temperate phages belonging to MT II subgroups, a restriction analysis profile was carried out using *HindIII* (figure 1). Each phage showed a specific restriction profile and differences in genome length were detected: 40, 32 and 39 kb for sb A, B and C respectively; also in this case, genome length is on average with other lactococcal temperate phages already reported (Labrie and Moineau, 2002).

Ex novo genome sequencing of temperate bacteriophage TB25

One inducible temperate bacteriophage was found in the cheese isolated *Lactococcus garvieae* strain TB25, and its genome was sequenced ex novo. Heating the digested genome with *HindIII*, at 70°C for 10 min did not change the restriction profile; this observation is typical of DNA packaging that occurs via a headful mechanism (*pac*-site).

As reported in table 4, its genome length is 38.122 bp, with a %GC content of 34.5. 58 ORFs were predicted from the genome and grouped in four distinct modules; function were attributed to 30 predicted proteins, while 50% of the functions remains still unknown; an overall genome comparison also revealed the lacking of any similarity with other lactococcal phages, strongly highlighting the new nature of this phage.

Morphogenesis cluster genes hit with proteins found in other Gram positive bacteria such as *Staphylococcus* spp., *Enterococcus* spp., *Fructobacillus* spp. and *Lactococcus lactis*, even if with a low amino acid identity (31%-54%); these findings could suggest past recombination events.

Table 3. Morphological classification of temperate phages from *L. garvieae* genome. [a] host strain of the temperate phage showed in the picture. [b] host strain of the temperate phage from which DNA was extracted for restriction analysis profile.

Morphotype	Subgroup	Host (source of isolation)	Capside diameter (nm)	Tail width (nm)	Tail length (nm)	Electron micrograph
I	**-**	Br3[a], Br1, Br4 (vegetable)	85±10.2	12.5±1.7	1009±16	Bar, 100 nm
II	**A**	G07[ab], G36 (Italian cheese) Cf4 (cattle farmers)	64.6±3	17.1± 0.86	202.6±4.6	Bar, 50 nm
	B	Smp3[a]-Sml1[b] (meat product), Cf1-Cf6-Cf7-Cf21-Cf24 (cattle farmers), G18 (cow milk)	62±3.3	11.6±1.83	119.6±4.8	Bar, 50 nm
	C	Lg9[a], Lg19, Lg28 (diseased fish) TB25[b] (Italian cheese) Bov3 (raw meat) Sed2 (vegetable)	60.6± 6	13.1± 2.8	222.2± 6	Bar, 50 nm

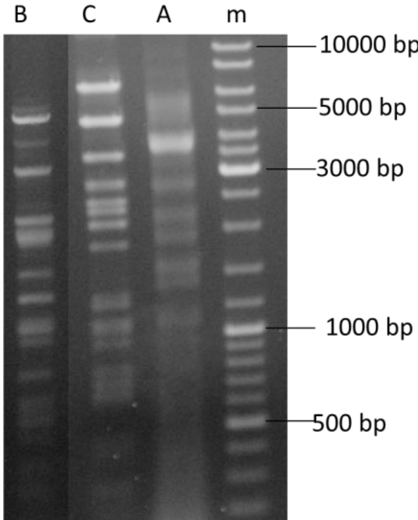


Figure 1. Restriction profile analysis (enzyme: *HindIII*) of the subgroups (A, B and C) belonging to morphotype II. m: marker (GeneRuler, DNA Ladder Mix, Fermentas).

Eight extra genes were located downstream the lysis cluster: among them a gene a codifying cold shock protein was found. Since this region of 6000 bp has a low GC content (31.5), it is attempted to hypothesize the acquisition of these genes through transduction processes.

Table 4. ORFs deduced from the genome of the temperate bacteriophage found in *L. garvieae* strain TB25.

Lysogeny | Replication/Transcription | Morphogenesis | Lysis

TB25

38,122 bp

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58

ORF	Strand[a]	Positions Start	Stop	Size aa	MM[b] kDa	pI	Putative RBS and start codon[c]	Predicted function[d]	Best-match BLASTp result[e]	# aa shared with best match / total # aa in best match (% ID)	E value	Size[f](aa)	Accession numbers
1	-	1140	1	379	43.9	9.45	AGGAGtagaaatcaa**ATG**	Integrase	Integrase, *Lactococcus garvieae*	375/379(99%)	0.0	379	WP_019292401
2	-	1738	1265	157	17.5	9.21	AGGAGtattttt**ATG**	SHOCT domain	H.p., *Lactococcus* phage TPW22	81/204(40%)	4e-33	205	AF066865_5
3	-	2354	1764	196	22.7	4.86	cGGAGgctct**ATG**	-	H.p., *Lactococcus garvieae*	174/196(89%)	4e-124	196	WP_040086243
4	-	2710	2366	114	13	4.92	gaaAGGttgatactc**ATG**	Transcription regulator	XRE trans. regulator, *L. garvieae*	102/114(89%)	2e-67	114	WP_004259391
5	+	3010	3228	72	8.03	9.26	AGGAGAtt**ATG**	-	H.p., *Lactococcus garvieae*	71/72(99%)	3e-42	72	WP_019335571
6	+	3244	3504	86	10.5	9.19	AGGAGtaaaaa**ATG**	-	Excisionase, *Lactococcus garvieae*	85/86(99%)	5e-55	86	WP_019292837
7	+	3514	3732	72	8.73	6.58	AGGAGttaaa**ATG**	-	H.p., *Lactococcus garvieae*	61/72(85%)	3e-37	72	WP_019299714
8	+	3755	3937	60	7.35	8.83	AGGAaAtaaaa**ATG**	Transcription regulator	H.p., *Lactococcus garvieae* DCC43	46/57(81%)	3e-24	58	EKF50671
9	+	4019	4417	132	15.3	7.79	tGGAGAaataaaaa**ATG**	-	H.p., *Lactococcus garvieae*	131/132(99%)	4e-88	132	WP_017370187
10	+	4429	5109	226	26	6.78	AGGAGAataatttt**ATG**	-	Topoisomerase, *L. garvieae*	218/226(96%)	4e-156	226	WP_017370188
11	+	5099	5515	138	15.5	5.25	AGGAGAaagaggaaataa**ATG**	Ss DNA binding p.	Ss DNA-binding p., *L. garvieae*	135/138(98%)	4e-95	138	WP_019299071
12	+	5614	5940	108	12.6	9.69	tGGAGgaatag**ATG**	HNH endonuclease	Endonuclease, *Lactococcus garvieae*	106/108(98%)	6e-72	108	WP_017369953
13	+	5940	6701	253	28.9	7.73	AGGtGgtctaacta**ATG**	Phage protein	DNA replication protein, *L. garvieae*	212/243 (87%)	8e-153	259	WP_017370084
14	+	6710	6874	54	6.34	9.16	AGGtGctt**ATG**	-					
15	+	6876	7787	303	34.4	8.42	AGGttAttgat**ATG**	Primosomal protein	Prepilin peptidase, *L. garvieae*	139/301(46%)	4e-80	297	WP_042217561
16	+	7798	7947	49	5.85	6.57	AGGtGAaaa**ATG**	-	H.p., *Lactococcus garvieae*	45/49(92%)	2e-22	49	CEF50680
17	+	7944	8348	134	15.4	9.63	gcGAGActtggaaa**ATG**	Resolvase	*RusA*, *Lactococcus garvieae*	129/134(96%)	3e-89	134	WP_019293279
18	+	8454	8738	94	10.7	6.40	AGGAaggggaaaa**ATG**	-	H.p., *Lactococcus garvieae*	86/93(92%)	2e-54	93	WP_035002155
19	+	8782	9462	226	25.8	4.67	tGGAGAaacaac**ATG**	5'-deoxyadenosine d.	H.p., *Lactococcus lactis*	213/226(94%)	3e-157	226	WP_003132900
20	-	9994	9605	129	14.7	9.24	AGGtaAat**ATG**	Transmembrane protein	H.p., *Lactococcus garvieae*	121/129(94%)	2e-83	129	WP_019293277
21	+	10187	10393	68	7.34	4.78	AGGAGAataaaac**ATG**	-	H.p., *Lactococcus lactis*	52/68(76%)	5e-25	68	WP_012897654
22	+	10399	10632	77	9.01	10.56	tGGAGAataagtcaa**ATG**	-	H.p., *Lactococcus garvieae*	32/52(62%)	2e-10	73	WP_017370067
23	+	10629	11177	182	20.8	8.91	AGGttAaaca**ATG**	Transmembrane protein	*LemA* family protein, *L. lactis*	136/182(75%)	3e-96	184	WP_046780940
24	+	11186	11995	269	30.8	8.47	AGGtGcaa**ATG**	Transmembrane protein	H.p., *Lactococcus garvieae*	263/269(98%)	0.0	269	WP_017369938
25	+	12075	12497	140	16.4	6.61	AGGgGggaaagtt**TTG**	-	H.p., *Lactococcus garvieae*	136/140(97%)	5e-95	140	WP_017370065
26	+	12670	13500	276	32.3	5.25	AGGAGtgtatt**TTG**	-	ORF27, *Lactococcus* phage Tuc2009	178/276(64%)	5e-132	276	NP_108706
27	+	13587	14039	150	17.2	9.08	AGGtGAgcgattga**GTG**	Terminase	TSs, *Staphylococcus saprophyticus*	81/150(54%)	3e-39	174	WP_041080371
28	+	14036	15283	415	47.4	6.16	tGGAGAaattgaa**ATG**	Terminase	TLs, *Macrococcus caseolyticus*	212/392(54%)	2e-146	416	WP_012656828
29	+	15298	16755	485	55.7	4.97	taGAGAgggtgaggata**TTG**	Portal protein	Portal protein, *E. faecium*	204/482(42%)	8e-114	499	WP_047937716
30	+	16742	17716	324	37.7	9.10	AGGAGAtgtagctc**ATG**	Head morphogenesis	H.p., *Enterococcus dispar*	132/299(44%)	2e-72	296	WP_016173631
31	+	17800	18309	169	18.5	4.39	AGGAGgggcaaat**ATG**	-					
32	+	18312	18704	130	13.9	5.05	AGGAGcataaat**ATG**	-	H.p., *Enterococcus faecalis*	57/106(54%)	6e-31	113	WP_002407384
33	+	18704	19708	334	37	5.17	AGGAacaaaata**ATG**	Major capsid protein	Head protein, *Enterococcus faecium*	132/331(40%)	1e-83	335	WP_002311457

34	+	19729	20043	104	11.8	4.67	AGGAGAggtgcaa**GTG**	Head-tail connector p.	H.p.,*Fructobacillus ficulneus*	43/96(45%)	1e-22	109	GAO99837
35	+	20044	20349	101	11.4	8.80	tgggggtattag**ATG**	-	H.p., *Staphylococcus pasteuri*	30/96(31%)	2e-06	100	WP_023373491
36	+	20342	20683	113	12.7	5.17	AGGtAGtggtc**ATG**	Tail-component	Phage protein, *L. johnsonii*	49/114(43%)	7e-21	116	EEJ59343
37	+	20683	21069	128	14.6	4.54	AGGctttttaaata**ATG**	-	H.p., *Enterococcus faecalis*	35/112(31%)	2e-11	130	WP_016619128
38	+	21081	21659	192	21.1	4.95	AGGAGAaaaaaa**ATG**	Major tail protein	H.p.,*Fructobacillus tropaeoli*	68/184(37%)	2e-30	188	GAP04943
39	+	21677	21913	78	7.9	4.57	AGGtaAcagaaaaa**ATG**	-	H.p., *Lactococcus lactis*	37/70(53%)	3e-11	70	WP_023189578
40	+	21928	22275	115	12.8	5.04	AGGgtAaatc**ATG**	Tail assembly	H.p., *Enterococcus faecalis*	42/116(36%)	2e-07	132	WP_002363376
41	+	22350	22634	94	11.1	4.87	AGaAattgaccgc**ATG**	-					
42	+	22634	26710	1358	143.2	9.00	AGGAGgcata**ATG**	Tape measure	Phage tail, *E. faecalis*	441/1369(32%)	0.0	1348	WP_042888997
43	+	26803	27426	207	24.3	4.89	AGGAaAag**TTG**	-	H.p., *Fructobacillus fructosus*	68/210(32%)	3e-26	239	WP_010691880
44	+	27426	28874	482	55.3	5.54	AGGtaAtgatgta**ATG**	Tail endopeptidase	Partial endolysin, *F. fructosus*	186/439(42%)	1e-113	612	WP_010691878
45	+	28887	29771	294	34	4.47	AGGgagttac**ATG**	-	H.p., *Enterococcus faecalis*	79/188(42%)	3e-38	243	ETU52076
46	+	29773	30024	83	9.44	4.67	AGGgGAattaat**ATG**	-	H.p., *Enterococcus faecium*	30/78(38%)	7e-12	110	WP_005874742
47	+	30027	30143	38	4.47	4.71	AGGgGAaataatc**ATG**	-					
48	+	30143	30565	140	15.2	5.02	AGaAGAagggtggttcaacta**ATG**	-	H.p., *Lactococcus garvieae*	87/140(62%)	2e-49	133	WP_019292915
49	+	30582	30944	120	13.8	5.16	AGGAaAaataaaaa**TTG**	-	Holin, *Lactococcus garvieae*	118/120(98%)	4e-77	120	WP_019293253
50	+	30928	31995	355	38.5	5.39	AGGAGAtgaaa**ATG**	Lysin	1,4-beta-N-acetylmuramidase, *Lactococcus garvieae*	346/355(97%)	0.0	355	WP_019292912
51	+	32066	32728	220	25.7	5.83	tGGAGActaaca**ATG**	Glucose-1-dehydrogenase	H.p., *Lactococcus lactis*	68/216(31%)	2e-31	223	WP_023163727
52	+	32824	33339	171	19.8	4.96	AGGtGcttaga**ATG**	Helix-hairpin domain	H.p., *Lactococcus lactis*	69/180(38%)	4e-21	180	WP_046781535
53	+	33593	34423	276	31	4.80	AGGAGctatt**ATG**	P-loop nucleoside triphosphate hydrolase	H.p., *Lactococcus lactis*	187/274(68%)	4e-128	279	WP_003132960
54	+	34425	34832	135	16.3	8.79	AGGAGgtgtgat**ATG**	-	H.p., *Lactococcus lactis*	74/136(54%)	3e-34	146	WP_003132961
55	+	34835	35590	251	29.5	4.89	AGGAGttaaaa**TTG**	-	H.p., *Lactococcus lactis*	163/248(66%)	2e-111	248	WP_003132962
56	+	36058	36228	66	7.27	5.16	AGGgaAatatatatact**ATG**	Cold-shock protein	Cold-shock protein, *L. garvieae*	65/66(98%)	2e-38	66	WP_017369912
57	+	36845	36967	40	4.51	9.99	AGGAtAtgat**ATG**	-	ORF1091, *L. garvieae* ATCC 49156	39/40(98%)	9e-18	40	BAK58604
58	+	37081	37677	198	22.5	5.49	AGGAGctagtg**ATG**	Histidine phosphatase family	H.p., *Lactococcus garvieae*	196/198(99%)	4e-144	198	WP_019293168

Trans. Regulator, transcriptional regulator; *L. garvieae*, *Lactococcus garvieae*; Ss DNA binding p., single strand DNA binding protein; 5'-deoxyadenosine d., 5'- deoxyadenosine deaminase; TSs, terminase small subunit; TLs, terminase large subunit; *E. faecium*, *Enterococcus faecium*; Head-tail connector p., head-tail connector protein; *F. fructosus*, *Fructobacillus fructosus*.

[a]Orientation of the gene in the genome.

[b]MM, molecular mass.

[c]RBS, ribosomal binding site: uppercase letters represent the hypothetical RBS sequences, bold letters the starting codons.

[d] – indicates no significant matches.

[e]BLASTp result corresponds to second best alignment.

[f]Total size of the aligned proteins

Genome comparison

In order to investigate the presence of other *L. garvieae* temperate phages and their genomic features, fourteen sequenced strains of *Lactococcus garvieae* were investigated for the presence of temperate bacteriophages along their genomes. As reported in table 5, 7 complete temperate bacteriophages were found in the genome of six strains.
Together with phage from strain TB25, the length ranged from 30 to 40 kb and, where host genome status permitted, integration site (*att* core) sequences were also determined. Interestingly for six temperate phages, %GC (34.0-35.9%) is lower compared to %GC of the species (37-38%) and more similar to *L. lactis* species. tRNAs were found in 5 genomes with a frequency on average with that already reported for lactococcal virulent phages (Samson and Moineau, 2010).
A genome comparison of *L. garvieae* phage genomes with other lactococcal phage genomes available in GenBank is reported below. When no homology was found, a deeper analysis of each gene was carried out.

As reported in figure 2 temperate phage genomes found in strain ATCC 49156 and UNIUD074 (both from diseased fish) showed the highest nucleotide and aminoacid homology with temperate phage ϕTP712 from widely used plasmid free *L. lactis* MG 136 (Roces et al., 2013). The three genomes have a similar size and genome organization. Morphogenesis module is the most conservative region between *L. garvieae* and *L. lactis* phages and for this reason we can hypothesize similar morphological features among the three phages. The average aminoacid identity homology of the morphogenesis module is respectively 71% (UNIUD 074) and 86% (ATCC 49156) and these differences led us to speculate that the moment of infection by ϕTP712 and/or the sharing of same ecological niches might have happened differently. Our hypothesis can also be confirmed by the presence in UNIUD 074 phage genome of three interrupted genes (white arrows), where one of this (*cro*-like) conferred a strict lysogenic nature of phage; these interruptions should have happened after infections and carried out by the cell to increase the fitness between phage and host (Desiere et al., 2001).
Only 11 orfs present in ATCC 49156 and UNIUD 074 have an aminoacid homology each others higher than 80%.
One of the temperate bacteriophage found in strain IPLA 31405 isolated from cow milk, showed the highest homology with *Lactococcus lactis* phage ul.36 and its mutant ul36.k1 resistant to abortive infection mechanism AbiK (Labrie and Moineau, 2007); both of them are virulent phages members of the P335 group which contains both temperate and lytic phages (Deveau et al., 2006). Moreover, the dUTPase found in ul36.k1 and proposed to be a target for P335 phages detection, was not found in IPLA genome (Labrie and Moineau, 2002).
As reported in the figure 3, 50% of the predicted orfs, (26/52) from IPLA showed homology with ul36.k1, in some case with an aminoacid identity higher than 90%.
Interestingly, the same DNA segment described in ul36.k1 (Labrie and Moineau, 2007) and ul36.1 (Bouchard and Moineau, 2000) both resistant to AbiK, was found in IPLA genome (between the XRE regulator and Rep-like genes).

Table 5. Position, orientation, length, *att* core sequence, %GC and tRNA of the prophage in different *L. garvieae* genomes.

Host (source of isolation)	Research procedure	Temperate phage(s)	Extremities	Length	*Att* core sequence	%GC	tRNA
ATCC49156 (diseased yellowtail)	*Silico*	1	1,146,793-1,106,521	40,273 bp	AACTCCCCTCGCCTCCATTGTAT[a]	35.4	2 (Lys, Met)
Lg2 (diseased yellowtail)	*Silico*	1	1,160,852-1,120,580	40,273 bp	AACTCCCCTCGCCTCCATTGTAT[a]	35.4	2 (Lys, Met)
UNIUD 074 (diseased rainbow trout)	*Silico*	1	40,669-2,192	38,478 bp	-[b]	35.9	2 (Ser, Met)
8831 (diseased rainbow trout)	*Silico*	1	31,965-251	31,715 bp	-[b]	34.6	1 (Arg)
IPLA 31405 (cow milk)	*Silico*	2 a	204,469-239,454	34,986 bp	AACTCCCCTCGCCTCCATTG[a]	36.4	-[c]
		b	509,217-478,639	30,579 bp	TTGTGCCAAATTTGTGCCAAA[a]	35.1	-[c]
NBRC 100934(cow mastitis)	*Silico*	1	48,776-12,512	36,265 bp	ATGGGTGGCATGATGTA[a]	37.5	1 (Lys)
TB25 (Italian cheese)	*Ex novo* sequencing	1	1-38,122	38,122 bp	-[c]	34.5	-[c]

[a]*Att* core sequences has been determined by searching for perfect direct repeats in the vicinity of the phage genome.
-[b]: none sequences more than 10bp were found.
-[c]: not found.
Lys: lysine, Met: methionine Ser: serine, Arg: arginin

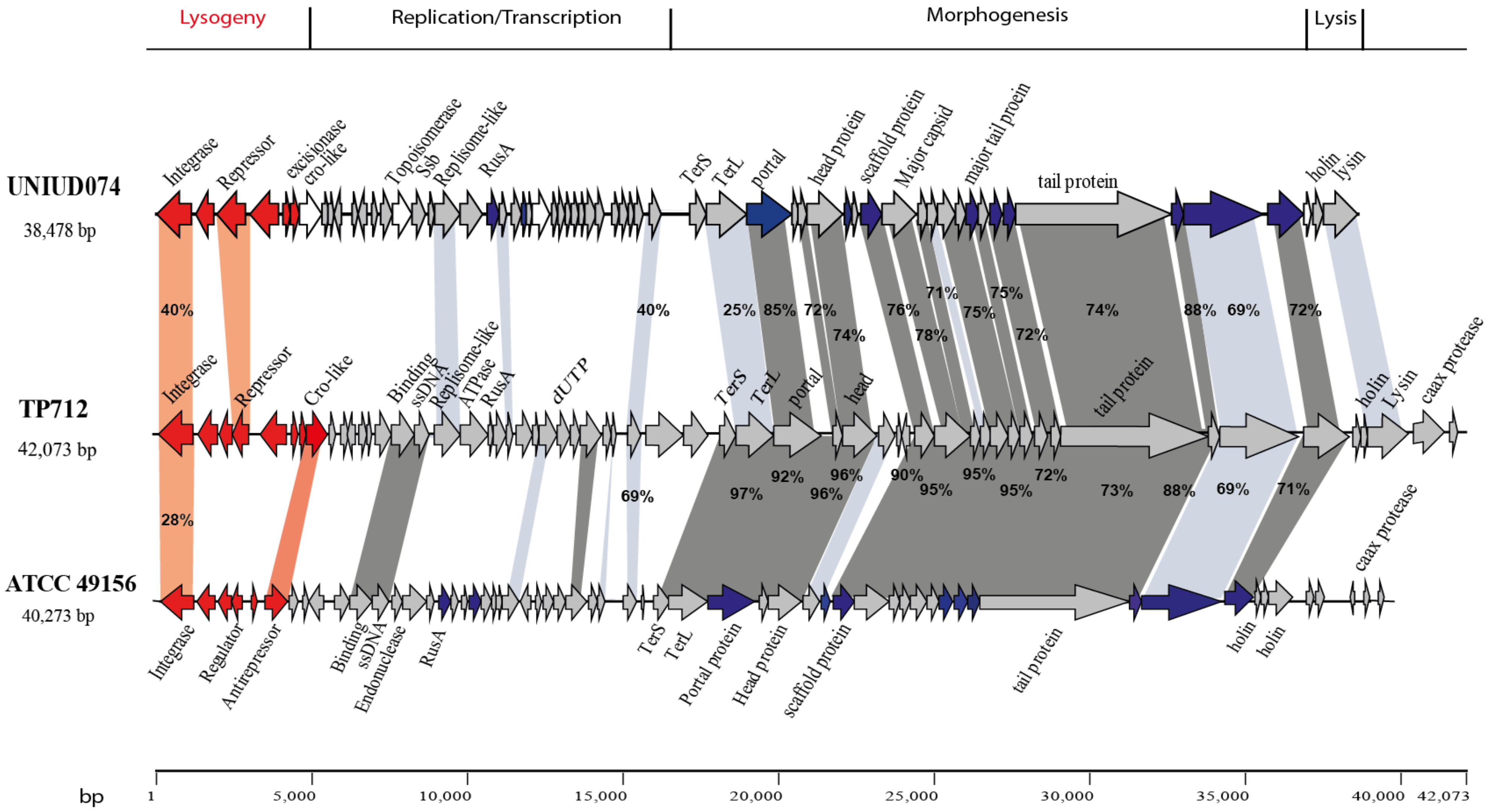


Figure 2. Comparative genomic between temperate phage TP712, ATCC 49156 and UNIUD074.Color shading was used to discriminate between similarity ≥70% amino acid identity (dark color) and similarity of ≤ 69 amino acid identity (light color); the absence of shading means no significant similarity. Blue arrows identify amino acid similarity ≥80% and white arrows identify interrupted orfs. The percent aminoacid acid identity inside the shading is representative of the aligned region only. Abbreviations: ssb: binding protein, *RusA*: Holiday junction resolvase, TerS: terminase small subunit, TerL: terminase large subunit, ssDNA: single strand DNA binding protein.

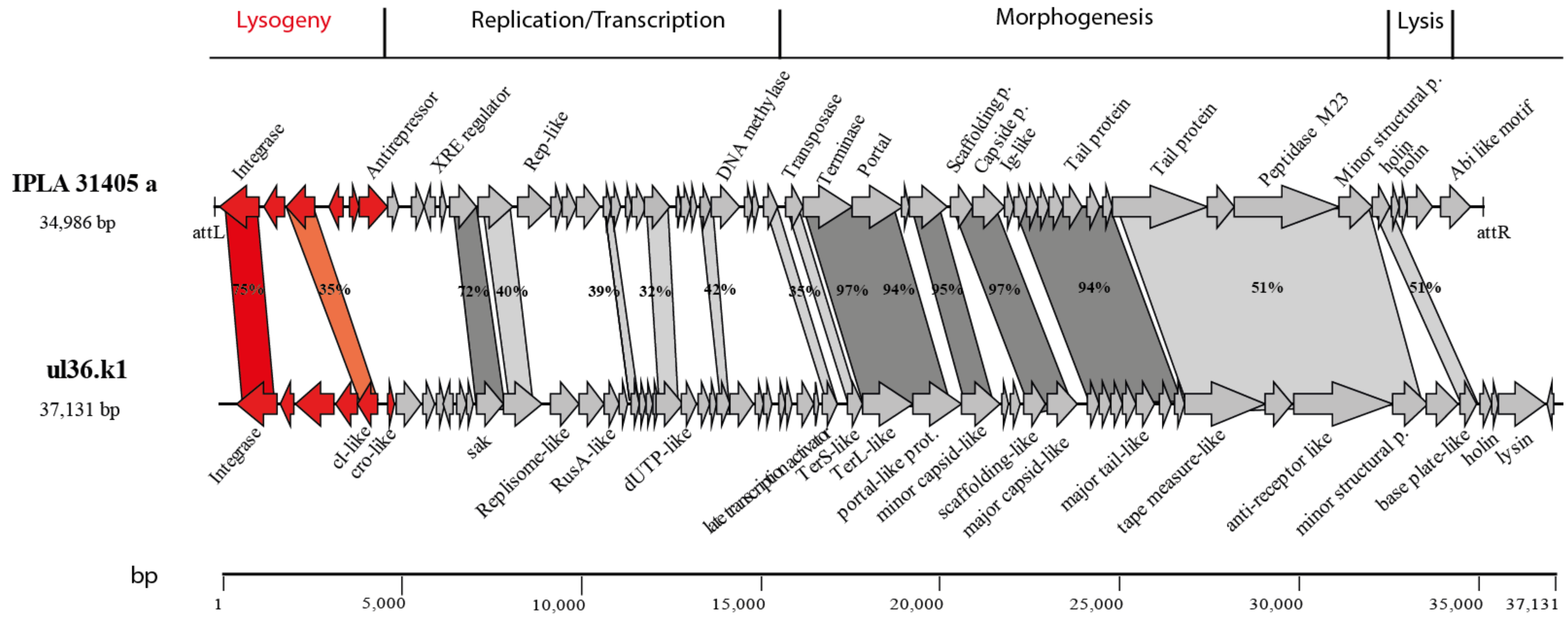


Figure 3. Comparative genomic between temperate phage IPLA31405a and ul36.k1. Color shading was used to discriminate between similarity ≥70% amino acid identity (dark color) and similarity of ≤ 69 amino acid identity (light color); the absence of shading means no significant similarity. The percent aminoacid acid identity inside the shading is representative of the aligned region only.

Abbreviations: p: protein, *RusA*: Holiday junction resolvase, terminase small subunit, TerL: terminase large subunit.

In table 5 are depicted a potential *att* core domain and a Abi-like domain protein was found along the genome, which could be involved in bacteriophage resistance causing cell death upon infection (Chopin et al., 2005).
A phylogenetic tree was constructed using aminoacid sequences of Abi-like protein found in IPLA3405 and 20 lactococcal Abi systems (Chopin et al., 2005; Durmaz and Klaenhammer, 2007; Haaber et al., 2008) using neighbour-joining method, boostraps from 100 replicates with a random seed (figure 4). Results highlighted that Abi-like protein found in *L. garvieae* temperate phage harboured an independent branch; since no similarity to other phage resistance mechanisms was found, as a novel Abi mechanism can be supposed.
The other temperate phage found in *L. garvieae* strain IPLA 31405b shows homology with phage r1t from *L. lactis* subsp. *cremoris* R1 (van Sinderen et al., 1996) (figure 5), a temperate bacteriophage belonging to P335 group, subgroup III (Samson and Moineau, 2010).
As reported for the other phages above described, the highest aminoacid identity was found in morphogenesis clusters (75%). Also in this case a gene codifying a dUTPase was not found in *L. garvieae* genome.
The identification of the hypothetical *att* core domain allowed us to consider belonging to phage genome the gene, located 700 bp downstream the lysis module, encoding a protein showing a conserved cold-shock DNA-binding domain (pfam00313).

Temperate phages found in strain 8831 (diseased fish) and NBRC 100934 (cow with mastitis) showed low nucleotide homology with other genomes in GenBank. The nature of these two new phages was deeper studied through genome analysis, as reported in the tables 6 and 7.
In both phage genomes organization is similar to those reported for other genomes; the major part of the orfs related to replication and transcription module showed hits with proteins from *L. garvieae*, highlighting the adaptation of these phages to this lactococcal species. On the contrary the morphogenesis cluster presents the highest nucleotide variability. In 8831, seven orfs (23, 24, 26, 27, 28 and 38) match (with an aminoacid identity ranging from 53 to 82%) with proteins found in three different species of genus *Weissella* (*hellenica*, *oryzae* and *koreensis*) (Fusco et al., 2015) and species of *Enterococcus* (*gilvus* and *faecalis*) frequently isolate from human-clinical samples (Tyrrell et al., 2002; Zhang et al., 2015). In NBRC 100934, out of 19 orfs found, seven hit with *L. lactis* genomes (aminoacid identity average of 75%) and eight (orf 37, 38, 39, 40, 41, 42, 45 and 46) (aminoacid identity average of 51%) match with different species of *Lactobacillus* (Athalye-Jape et al., 2015; Petrova et al., 2015; Tiequn et al., 2015).

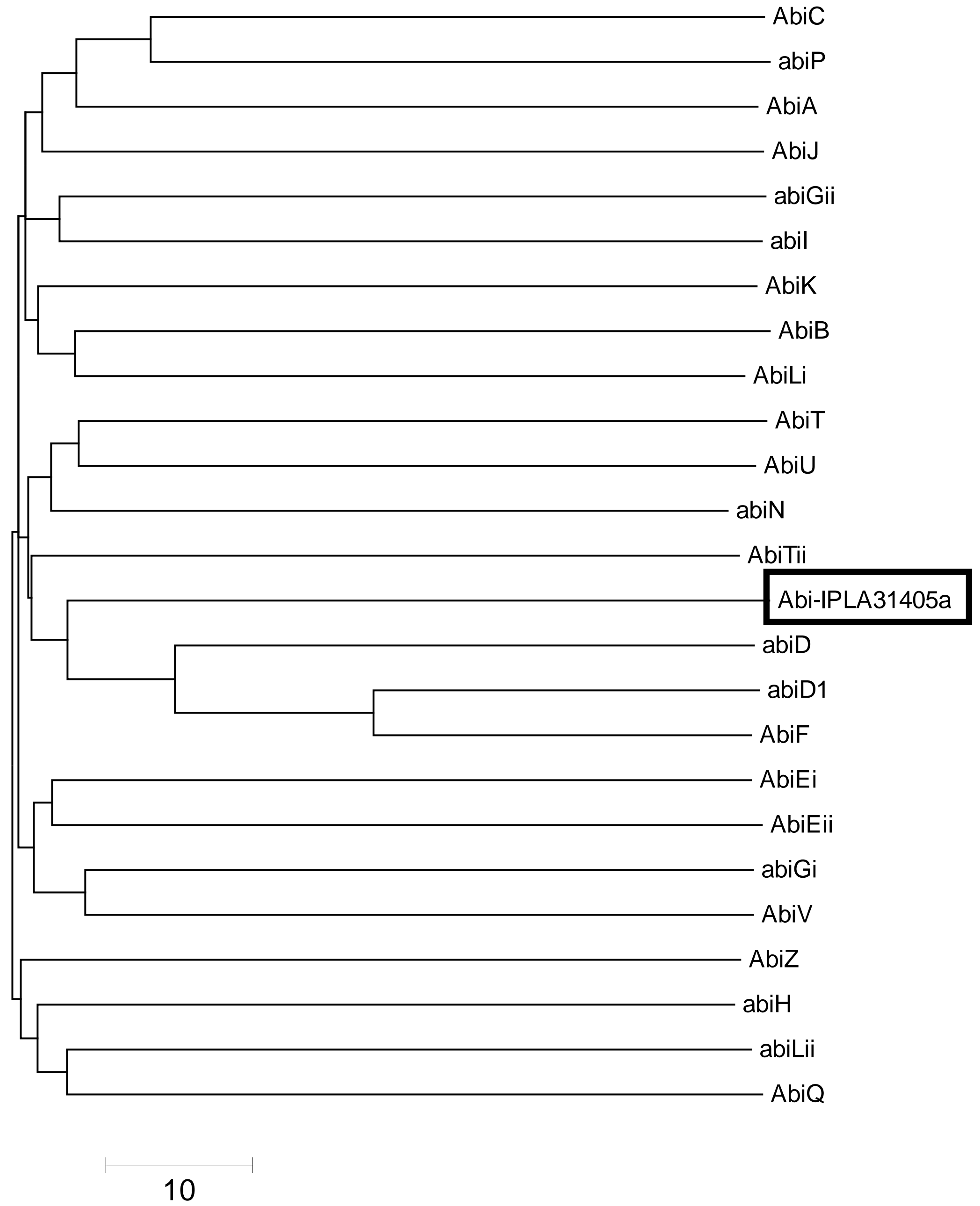


Figure 4. Phylogenetic tree of different Abi mechanisms found in different *Lactococcus* species. Abi found in temperate phage IPLA 31405 is highlighted by the rectangle.

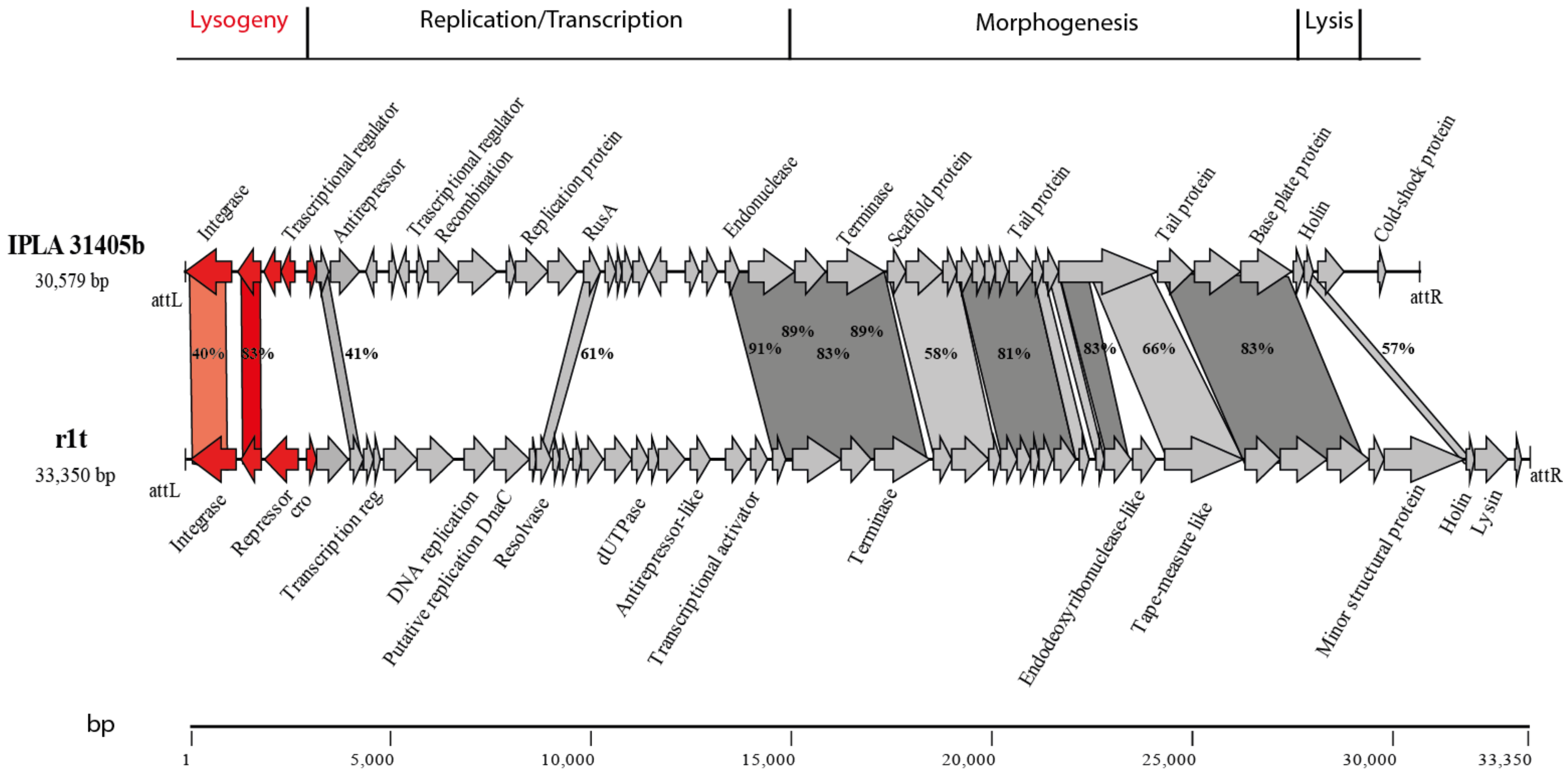


Figure 5.Comparative genomic between temperate phage IPLA31405b and r1t. Color shading was used to discriminate between similarity ≥70% amino acid identity (dark color) and similarity of ≤ 69 amino acid identity (light color); the absence of shading means no significant similarity. The percent aminoacid acid identity inside the shading is representative of the aligned region only.
Abbreviations: *RusA*: Holiday junction resolvase, terminase small subunit, TerL: terminase large subunit.

Table 6. ORFs deduced from the genome of temperate bacteriophage found in *L. garvieae* strain 8831.

Lysogeny | Replication/Trascription | Morphogenesis | Lysis

8831
31,715 bp

ORF	Strand[a]	Position Start	Position Stop	ORF size (aa)	MM[b] (kDa)	pI	Putative RBS and start codon[c]	Predicted function[d]	BLASTp result[e]	# aa shared /total # aa in BLASTp match (% ID)	E value	Size[f] (aa)	Accession numbers
1	-	1125	1	374	43.7	9.53	aAGGAGtagaaatcaa**ATG**	Integrase	Int, *L. lactis*	258/380 (68%)	0.0	380	WP_021037034.1
2	-	1834	1256	192	20.4	7.8	AGGAGAaaaattttta**ATG**	-	H.p., *L. garvieae*	171/192 (89%)	1e-117	192	WP_003134907.1
3	-	2477	1887	196	22.9	5.03	ttGAAGActaatcgggggctct**ATG**	-	H.p., *L. garvieae*	190/196 (97%)	6e-136	196	WP_040086243.1
4	-	2842	2489	117	13.3	4.80	AGGAGAcctatata**ATG**	Transcriptional regulators	XRE transcriptional regulator, *L. garvieae*	115/117 (98%)	2e-75	117	WP_019292841.1
5	-	3930	3136	264	29.6	5.44	AGGAGAtttt**ATG**	-					
6	+	4085	4315	76	9.04	7.97	gagaaAGGAtAaaa**ATG**	-	H.p., *L. garvieae*	68/75 (91%)	1e-42	75	WP_003134901.1
7	+	4328	4579	83	9.62	9.36	AGAAAGGGGtaaac**ATG**	-					
8	+	4596	4745	49	6.08	7.94	tttAAGGAGcaact**ATG**	-	H.p., *L. garvieae*	46/49 (94%)	1e-23	49	CEF51378.1
9	+	4962	5738	258	29.5	4.96	GGAGGAAttaaaa**ATG**	Recombination protein	Recombination protein Bet, *L. garvieae*	250/258 (97%)	0.0	258	WP_004257695.1
10	+	5739	6704	321	37.7	5.11	GGAGGActtttaa**ATG**	-	H.p., *L. garvieae*	293/321 (91%)	0.0	321	WP_004257691.1
11	+	6937	7263	108	12.6	9.52	GGAGGaatat**ATG**	HNH endonuclease	Endonuclease *L. garvieae*	107/108 (99%)	9e-72	108	WP_017369953.1
12	+	7263	8048	261	30	8.75	AGGtGGtccagcta**ATG**	Helix-turn-helix domain	HTH, *L. garvieae*	245/262 (94%)	2e-173	261	WP_026063845.1
13	+	8032	8316	94	11.2	9.39	GAGGAGAtagcag**ATG**	-	H.p., *L. garvieae*	83/93 (89%)	6e-56	93	WP_017369951.1
14	+	8313	8717	134	15.4	9.50	GGGtAAGtaagac**ATG**	Resolvase *RusA*	*RusA*, *L. garvieae*	129/134 (96%)	2e-89	134	WP_019335580.1
15	+	8824	9105	93	10.5	9.30	AAGAAAGgggagaa**ATG**	HD nuclease superfamily	H.p., *L. garvieae*	74/90 (82%)	2e-45	90	WP_017369948.1
16	-	9599	9210	129	14.8	9.03	AGAAAGtaggtaaat**ATG**	-	H.p., *L. garvieae*	121/129 (94%)	2e-83	129	WP_019335578.1
17	-	10115	9717	132	15.6	7.85	aatAGGAGGgccactat**ATG**	Arsenate reductase	*ArsR*, *L. garvieae* TRF1	126/132 (95%)	1e-87	132	ETD04164.1
18	+	10306	10707	133	15.5	9.18	AGGAAGtttattc**ATG**	-	H.p., *L. garvieae*	132/133 (99%)	3e-90	133	WP_017369937.1
19	+	11005	12135	376	41.4	5.92	atAAGAGGtaaaca**ATG**	Lysozyme-like	H.p., *L. garvieae*	159/180 (88%)	3e-103	376	WP_042218703.1
20	+	12355	12987	210	22.6	4.89	taaAGGAGGtaggaa**ATG**	-	H.p., *L. garvieae* TRF1	131/138 (95%)	4e-89	210	ETD04990.1
21	+	13272	13820	182	21	8.91	aattaaAGGGGAAAat**ATG**	-	H.p., *L. garvieae*	77/164 (47%)	6e-44	162	WP_042217496.1
22	+	14032	14523	163	18.2	7.78	atctAAAGGAGggaggt**ATG**	Terminase	Terminase, *L. garvieae*	118/150 (79%)	3e-78	150	WP_014024903.1
23	+	14513	15796	427	49.2	7.66	TATAAAGGGGttcaaag**ATG**	Terminase	Terminase, *Weissella hellenica*	290/426 (68%)	0.0	426	WP_042492710.1
24	+	15798	17393	531	61	4.55	AGGAAAttaat**ATG**	Portal protein	PPP, *Weissella hellenica*	328/497 (66%)	0.0	496	WP_042492712.1
25	+	17377	18267	296	34.1	8.53	AAGAGGAgtcagaag**ATG**	Head morphogenesis	Morphogenesis protein, *E.gilvus*	132/292 (45%)	2e-85	296	WP_010780109.1
26	+	18410	18979	189	21.1	4.75	AGGAGGAAGaatt**ATG**	-	H.p., *Weissella hellenica*	90/144 (63%)	3e-52	168	WP_042492716.1
27	+	18990	19856	288	32.2	4.99	AGGAGAaaaatt**ATG**	-	H.p., *Weissella oryzae*	236/287 (82%)	9e-176	290	WP_027699892.1
28	+	19934	20392	152	15.4	4.80	tgAGGAGGgctt**ATG**	Bacterial Ig-like domain	Head protein, *Weissella koreensis*	81/147 (55%)	5e-37	150	WP_006845687.1
29	+	20405	20749	114	13.1	4.80	atAGGAGGtaatg**ATG**	Phage head-tail	H.p., *Weissella oryzae*	72/110 (65%)	8e-44	111	WP_027699890.1

								connector					
30	+	20746	21021	91	10.7	8.57	aatAGcGAGGtttat**ATG**	-	H.p., *Enterococcus faecalis*	34/92 (37%)	4e-09	92	WP_016624515.1
31	+	21021	21371	116	13.1	9.92	ttGGAGGtgattta**ATG**	Putative tail-component	H.p., *Enterococcus faecalis*	40/108 (37%)	4e-19	112	WP_010708429.1
32	+	21368	21730	114	12.8	5.03	ttAAAAAGGtggttga**ATG**	Zinc peptidases	H.P., *Enterococcus gilvus*	61/113 (54%)	5e-37	119	WP_010780116.1
33	+	21741	22388	215	23.3	4.25	AGGAGAaaat**ATG**	Phage major tail protein	MTP, *Enterococcus faecalis*	129/191 (68%)	3e-87	192	WP_010708431.1
34	+	22458	22901	147	16	5.11	aaAGAGGAaaaatt**ATG**	-	H.p.,*Enterococcus faecalis*	52/137 (38%)	1e-17	135	WP_010826655.1
35	+	22946	23227	93	10.7	9.81	aacAGAtatcaatcttatt**ATG**	-					
36	+	23227	25926	899	95.2	9.61	gaAGGGAGgtaaata**ATG**	Phage protein	Tail protein, *Lactococcus lactis*	361/883 (41%)	0.0	833	WP_011676059.1
37	+	25927	26772	281	32.5	5.15	aagcGGAAAGGaatataa**ATG**	Phage tail protein	H.p., *Enterococcus gilvus*	139/282 (49%)	4e-93	282	WP_010780120.1
38	+	26769	28871	700	77.6	4.72	AAAGtGGAgttttactt**ATG**	Endopeptidase tail	H.p., *Weissella oryzae*	377/717 (53%)	0.0	696	WP_027699884.1
39	+	28861	30204	447	48.6	4.83	AAGAGGaaatatgA**ATG**	-	H.p., *Lactococcus garvieae*	195/406 (48%)	2e-103	401	WP_019299182.1
40	+	30245	30607	120	13.8	5.16	AGAGAGGaaaaataaaaa**TTG**	-	Holin, *Lactococcus garvieae*	118/120 (98%)	4e-77	120	WP_019299702.1
41	+	30591	31715	374	40.5	5.33	AGGAGAtgaaa**ATG**	Lysine motif	Lysin, *Lactococcus garvieae*	342/374 (91%)	0.0	374	WP_019291755.1

Int, Integrase; H.p., hypothetical protein; HTH, helix turn helix; RusA, resolvase; ArsR, arsenate reductase; PPP, phage portal protein, E. gilvus, Enterococcus gilvus; MTP, major tail protein

[a]Orientation of the gene in the genome.

[b]MM, molecular mass.

[c]RBS, ribosomal binding site. The uppercase letter represent the hypothetical RBS sequences and in bold the starting codon, whereas the lowercase letter represent the nucleotides between the RBS sequence and the starting codon and also nucleotide after the RBS.

[d] – indicates no significant matches.

[e]BLATp result corresponds to second best alignment.

[f]Total size of the aligned protein.

Table 7. ORFs deduced from the genome of temperate bacteriophage found in *L. garvieae* strain NBRC 100934.

Lysogeny | Replication/Transcription | Morphogenesis | Lysis

NBRC 100934

36,265 bp

attL ... attR

ORF	Strand[a]	Position Start	Position Stop	ORF size (aa)	MM[b] (kDa)	pI	Putative RBS and start codon[c]	Predicted function[d]	BLASTp result[e]	# aa shared /total # aa in BLASTp match (% ID)	E value	Size[f] (aa)	Accession numbers
1	-	1146	1	381	4.40	9.55	AGGAGtAgaaatcaa**ATG**	Integrase	Int, prophage L54a of *L. piscium*	252/381 (66%)	0.0	381	CEN29374.1
2	-	1455	1279	58	6.9	9.21	AGAtAtGGAgttcgacta**ATG**	-					
3	-	1790	1455	111	13.3	4.59	ttAGGAGGcaat**ATG**	-	H.p., *L. lactis*	56/111 (50%)	3e-34	111	WP_014570766.1
4	-	2240	1806	144	16.4	5.23	AGGAGAtttt**ATG**	-	H.p., *L. lactis*	81/144 (56%)	9e-44	144	WP_021722776.1
5	-	2597	2265	110	12.6	7.86	ttGGAGGtgcaaaa**ATG**	Transcriptional regulator	cro/CI transcriptional regulator *E. faecalis*	48/103 (47%)	2e-26	107	WP_010826721.1
6	+	3015	3254	79	8.94	9	GAAAGGAGcatat**ATG**	-					
7	+	3247	3552	101	12.1	8.92	ttGAAAGAGGtagtcA**ATG**	-					
8	+	3555	4247	230	26.4	7.84	taaaGAAAGGAatttataaga**ATG**	Anti-repressor protein	*AntA*, *L. garvieae*	213/233 (91%)	1e-149	233	WP_040086487.1
9	+	4265	4381	38	4.46	6.04	AAAGGGatattc**ATG**	-	H.p., *L. garvieae* ATCC 49156	32/38 (84%)	3e-12	38	EOT33274.1
10	+	4392	4679	95	11.2	5.88	AGGAGAaaata**ATG**	-	H.p., *L. garvieae*	30/92 (33%)	9e-08	95	WP_042753188.1
11	+	4681	4881	66	7.86	9.42	gtAAAGGtaaataat**ATG**	-	H.p., *L. garvieae*	55/65 (85%)	3e-32	73	WP_003136214.1
12	+	4874	4987	37	4.27	4.60	AAGAGGAGcacaca**GTG**	-	H.p., *L. lactis*	30/37 (81%)	1e-10	37	KGF76748.1
13	-	5675	5004	223	25.3	5.35	atGGAGAaaat**ATG**	-	H.p., *L. lactis*	143/223 (64%)	4e-100	223	WP_033900168.1
14	+	5904	6725	273	31.8	9.39	ttGGAGGtgataga**TTG**	Integrase	Int., *L. garvieae*	198/273 (73%)	4e-138	273	WP_003136217.1
15	+	7125	7304	59	7.03	9.99	caAAAGAAtttttt**GTG**	Ribbon-helix-helix p.	-				
16	+	7429	7809	126	14.8	8.67	tttcAAGGtGttttc**ATG**	-					
17	+	7806	8543	245	27.8	9.34	AGGAGAagcaat**ATG**	Helix-turn-helix domain	H.p., *L. lactis*	167/245 (68%)	7e-87	249	WP_046782081.1
18	+	8553	8717	54	6.25	8.06	ttAGGAGAtttgt**ATG**	-	H.p., *L. lactis*	23/49 (47%)	2e-05	54	WP_010905352.1
19	+	8717	9610	297	34.5	6.93	ttGGAGGAaagta**ATG**	Chromosomal replication initiator protein DnaA	Prepilin peptidase, *L. garvieae*	289/297(97%)	0.0	297	WP_003136220.1
20	+	9622	10335	237	28	4.88	AGGAGAaacaag**ATG**	-					
21	+	10332	10631	99	11.7	4.61	gAGGtGAAaga**ATG**	-	H.p., *L. lactis*	48/106 (45%)	2e-11	95	WP_010905950.1
22	+	10628	10957	109	12.1	4.51	gcctGtGGAGGaaga**ATG**	-					
23	+	10950	11105	51	6.2	5.21	cgGAGGGGaag**ATG**	-					
24	+	11098	11325	75	9	4.75	gaAGGGAGcggcg**ATG**	-	H.p., *L. lactis*	69/73 (95%)	5e-39	145	WP_012897650.1
25	+	11352	11561	69	7.88	5.20	aagcAttGcGGAtgag**TTG**	-	H.p., *L. lactis*	59/68 (87%)	7e-37	80	WP_041168739.1
26	+	11655	11843	62	7.48	9.78	ttAAGGAGtgttggga**ATG**	-					
27	+	11948	12478	176	20.7	5.20	AGGAGAaagaa**ATG**	-	sugar-phospahte nucleotidyltransferase, *C. maltaromaticum*	118/176 (67%)	2e-84	180	WP_035065602.1
28	+	12475	12690	71	8.57	11.51	cgAAGAGGtagtcaa**ATG**	-	H.p.,*L. garvieae*	55/71(77%)	1e-28	71	WP_042753341.1
29	+	12877	13314	145	17.1	8.50	aAGGAGGttaaagtt**TTG**	-	H.p., *L. lactis* subsp. *cremoris*	57/89 (64%)	5e-33	138	KKW72760.1
30	+	13821	14321	166	19.8	9.61	AGGAAGGaatttaaaatt**ATG**	HNH endonuclease	HNH endonuclease *L. lactis*	106/170 (62%)	7e-71	170	WP_047205846.1
31	+	14442	14894	150	17.4	5.42	AGGAGAttattt**TTG**	Terminase	Terminase *L. lactis*	125/147 (85%)	1e-87	150	WP_014573149.1

32	+	14887	16800	637	73.6	5.13	cgGAAtGGGggacttcttcg**ATG**	Terminase	Terminase large subunit, phage P335	525/635 (83%)	0.0	646	NP_839922.1
33	+	16769	16978	69	7.72	8.14	gtAGAAGAttttg**ATG**	-	H.p., *L. lactis*	55/69 (80%)	2e-28	69	WP_015966929.1
34	+	16981	18186	401	44.5	5.91	aaGAAGAGAggtgattaatt**TTG**	Phage portal	Phage portal protein *L. lactis*	282/391(72%)	0.0	390	WP_023189043.1
35	+	18207	18911	234	25.5	4.64	agaaAGGAGGtcaagg**ATG**	Protease	Peptidase *L. lactis*	184/234 (79%)	5e-131	234	WP_032398674.1
36	+	18924	20177	417	46.2	4.94	aaaaatAAAGGAGactcaaa**ATG**	Phage capside	Phage capsid protein *L. lactis*	313/408 (77%)	0.0	409	WP_012898017.1
37	+	20188	20523	111	12.5	4.26	AAAtAAGGggtgacag**ATG**	Phage gp6-like head-tail connector protein	DNA packaging *L. plantarum*	48/106 (45%)	8e-22	112	CDN29021.1
38	+	20516	20863	115	13.2	9.15	aaAtAcGGAGAcggtg**ATG**	Phage head-tail joining protein	Head-tail joining protein *L. sakei*	61/113 (54%)	2e-39	115	WP_035147673.1
39	+	20866	21276	136	15.5	5.96	aaaaaGAAAGggtgact**ATG**	Putative tail-component	H.p.,*Lactobacillus sakei*	68/128 (53%)	9e-40	136	WP_035147674.1
40	+	21276	21659	127	14.6	5.48	aagtGAAAGGtgagaatta**ATG**	-	H.p., *Lactobacillus sakei*	58/126 (46%)	2e-37	128	WP_035147676.1
41	+	21661	22317	218	23	5	AAAGGAGcattaaa**ATG**	Phagetail	Tail protein, *Lactobacillus sakei*	142/215 (66%)	7e-97	222	WP_035147678.1
42	+	22342	22755	137	15.5	4.64	ataGAAGGAActtacgaca**ATG**	-	H.p., *Lactobacillus reuteri*	45/123 (37%)	2e-16	128	WP_011953494.1
43	+	22962	26834	1290	139.4	9.30	gtAGAAAGGGgaaaac**ATG**	Phage-related tail protein	phage tail length tape measure protein *Oenococcus oeni*	531/1275 (42%)	0.0	1218	WP_032816177.1
44	+	26855	27085	76	8.82	4.73	AGGAGAagaga**ATG**	-	H.p., *Lactococcus lactis*	35/69 (51%)	4e-13	68	WP_021214423.1
45	+	27458	28294	278	31.8	5.15	AGAAAGGagtccca**TTG**	Phage tail protein	H.p., *Lactobacillus sakei*	154/278 (55%)	2e-107	278	WP_035147684.1
46	+	28287	29789	500	56.3	4.94	gacctgacctttagtttcccttttctct**ATG**	Prophage endopeptidase tail	H.p., *Lactobacillus sakei*	238/431 (55%)	1e-173	508	WP_035147685.1
47	+	29786	30823	345	38.8	5.11	taAGGAGAaaac**ATG**	Hydrolase	Phage MSP (*E. faecalis* MTUP9)	264/345 (77%)	0.0	345	KAJ80425.1
48	+	30834	31493	219	24.6	4.41	AAGGAAaagaat**ATG**	-	Prophage Lp1 protein 54, *E.faecalis*	170/219 (78%)	5e-122	219	KAJ80426.1
49	+	31507	32046	179	19.5	4.92	GAAAGGaaaattaca**ATG**	-					
50	+	32068	32334	88	10.4	8.01	gaaAGGGGGAgaaat**ATG**	Haemolysin XhlA	H.p., *Lactobacillus farraginis*	30/83 (36%)	5e-06	88	WP_035180475.1
51	+	32354	33409	351	38.1	6.42	AGGAGAattttagaaa**ATG**	Lyzozyme	1,4-beta-N-acetylmuramidase, *L. garvieae*	321/355 (90%)	0.0	355	WP_019299026.1
52	+	34275	35279	334	38.4	5.14	GAGAGGTAGGataaa**ATG**	-	H.p., *Lactococcus garvieae*	229/231 (99%)	1e-161	231	AEK12129.1
53	+	35281	35949	222	25.7	9.49	AAAAAGGtgatggattaaa**ATG**	-	H.p., *Lactobacillus plantarum*	101/228 (44%)	4e-46	220	KGH42505.1

L. piscium, *Lactococcus piscium*; H.p. *L. lactis*, hypothetical protein *Lactococcus lactis*; *L. garvieae*, *Lactococcus garvieae*; *E. faecalis*, *Enterococcus faecalis*; *C. maltaromaticum*, *Carnobacterium maltaromaticum*; *L. plantarum*, *Lactobacillus plantarum*; *L. sakei*, *Lactobacillus sakei*; MSP, minor structural protein.

[a]Orientation of the gene in the genome.

[b]MM, molecular mass.

[c]RBS, ribosomal binding site. The uppercase letter represent the hypothetical RBS sequences and in bold the starting codon, whereas the lowercase letter represent the nucleotides between the RBS sequence and the starting codon and also nucleotide after the RBS.

[d] – indicates no significant matches.

[e]BLATp result corresponds to second best alignment.

[f]Total size of the aligned protein.

CONCLUSIONS

The main conclusions of this chapter are:

- 50 % of tested strains showed the presence of at least one temperate bacteriophages, belonging to *Siphoviridae* family;
- genome sequencing and analysis of an induced temperate from strain TB25 phages highlighted the nature of the phage;
- 7 complete phage genomes were found in 14 sequenced genomes of *L. garvieae*. Their %GC content is similar to that of *L. lactis* and this finding led us to hypothesize that the phages infecting *L. garvieae* originated from *L. lactis*;
- presence of cold shock protein in extra gene modules might highlight the role of temperate phages in the adaptability of *L. garvieae* species to different environments.

Related publication: Eraclio G., Fortina MG, Tremblay DM, D'Incecco P., Pellegrino L., Moineau S. (2015). *Characterization of temperate bacteriophages infecting* Lactococcus garvieae *reveals new insight into the evolution of the species*. Submitted Applied and Environmental Microbiology.

REFERENCES

1. Athalye-Jape G., Rao S., Patole S. (2015). Lactobacillus reuteri *DSM 17938 as a probiotic for preterm neonates: a strain-specific systematic review*. J Parenter Enter Nutr. doi: 0148607115588113.

2. Altschul S.F., Gish W., Miller W. Myers E.W., Lipman D.J. (1990). *Basic local alignment search tool*. J Mol Biol. 215:403–410.

3. Altschul S.F., Madden T.L., Schäffer A.A. Zhang J., Zhang Z., Miller W., Lipman D.J. (1997). *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Res. 25:3389–3402.

4. Aziz R.K., Bartels D., Best A.A., DeJongh M. Disz T., Edwards R.A., Formsma K., Gerdes S., Glass E.M., Kubal M., Meyer F., Olsen G.J., Olson R., Osterman A.L., Overbeek R.A., McNeil L.K., Paarmann D., Paczian T., Parrello B., Pusch G.D., Reich C., Stevens R., Vassieva O., Vonstein V., Wilke A., Zagnitko O. (2008). *The RAST server: rapid annotations using subsystems technology*. BMC Genomics. 9:75.

5. Bose M., Barber R.D. (2006). *Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences*. In Silico Biol. 6:223–227.

6. Bouchard J.D., Moineau S. 2000. Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. Virology. 270:65–75.

7. Chakraborty A., Ghosh S., Chowdhary G., Maulik U, Chakrabarti S. (2012). *DBETH: database of bacterial exotoxins for human.* Nucleic Acids Res. 40:615-620.

8. Chen L., Xiong Z., Sun L., Yang J., Jin Q. (2012). *VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors*. Nucleic Acids Res. 40:D641-645.

9. Chopin M.C., Chopin A., Bidnenko E. (2005). *Phage abortive infection in lactococci: variations on a theme.* Curr Opin Microbiol. 8:473-479.

10. Desiere F., McShan W.M., van Sinderen D., Ferretti J.J., Brüssow H. (2001). *Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic streptococci: evolutionary implications for prophage-host interactions*. Virology. 288:325–341.

11. Deveau H., Labrie S.J., Chopin M.C., Moineau S. (2006). *Biodiversity and classification of lactococcal phages*. Appl Environ Microbiol. 72:4338–4346.

12. Durmaz E., Klaenhammer T.R. (2007). *Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected* Lactococcus lactis. J Bacteriol. 189:1417-1425.

13. Fusco V., Quero G.M., Cho G.S., Kabisch J., Meske D., Neve H., Bockelmann W., Franz C.M. 2015. *The genus* Weissella: *taxonomy, ecology and biotechnological potential*. Front Microbiol. 6:155. doi: 10.3389/fmicb.2015.00155.

14. Haaber J., Moineau S., Fortier L.C., Hammer K. (2008). *AbiV, a novel antiphage abortive infection mechanism on the chromosome of* Lactococcus lactis *subsp.* cremoris *MG1363*. Appl Environ Microbiol. 74:6528–6537.

15. Labrie S., Moineau S. (2002). *Complete genomic sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of Lactococcal phages*. Virology. 296(2):308–320.

16. Labrie S.J., Moineau S. (2007). *Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages*. J Bacteriol. 189:1482-1487.

17. Laslett D., Canback B. (2004). *ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences*. Nucleic Acids Res. 32:11–16.

18. Lowe T.M., Eddy S.R. (1997). *tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence*. Nucleic Acids Res. 25:955–964.

19. Moineau S., Pandian S., Klaenhammer T.R. (1994). *Evolution of a lytic bacteriophage via DNA acquisition from the* Lactococcus lactis *chromosome*. Appl Environ Microbiol. 60:1832–1841.

20. Petrova M.I., Lievens E., Malik S. Imholz N., Lebeer S. (2015). Lactobacillus *species as biomarkers and agents that can promote various aspects of vaginal health*. Front Physiol 6:81. doi: 10.3389/fphys.2015.00081.

21. Roces C., Wegmann U., Campelo A.B., García P., Rodríguez A., Martínez B. (2013). *Lack of the host membrane protease FtsH hinders release of the* Lactococcus lactis *bacteriophage* TP712. J Gen Virol. 94:2814–2818.

22. Samson J.E., Moineau S. (2010). *Characterization of L*actococcus lactis *phage 949 and comparison with other lactococcal phages*. Appl Environ Microbiol. 76:6843–6852.

23. Tiequn B., Guanqun C., Shuo Z. (2015). *Therapeutic effects of* Lactobacillus *in treating irritable bowel syndrome: a meta-analysis*. Intern Med 54:243–249.

24. Tyrrell G.J., Turnbull L., Teixeira L.M., Lefebvre J., Carvalho G.S., Richard R., Lovgren M., Facklam R.R. (2002). Enterococcus gilvus *sp. nov. and* Enterococcus pallens *sp. nov. isolate from human clinical specimens*. J Clin Microbiol. 40:1140–1145.

25. van Sinderen D., Karsens H., Kok J., Terpstra P., Ruiters M.H.J., Venema G., Nauta A. (1996). *Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t*. Mol Microbiol. 19:1343–1355.

26. Zhang C., Du J., Peng Z. (2015). *Correlation between* Enterococcus faecalis *and persistent intraradicular infection compared with primary intraradicular infection: a systematic review*. J Endod 41:1207-1213

27. Zhou Y., Liang Y., Lynch K.H. Dennis J.J., Wishart D.S. (2011*). PHAST: A Fast Phage Search Tool.* Nucleic Acids Res. 39:W347-352.

5. ISOLATION AND CHARACTERIZATION OF NEW VIRULENT PHAGES INFECTING *LACTOCOCCUS GARVIEAE*

Recently, phage therapy has gained attention as a new and alternative strategy to prevent the growth of undesirable microorganism. In this chapter the isolation and characterization of new virulent phages infecting *L. garvieae*, to be used as biocontrol strategies in aquaculture sector, are presented. For one of this a comparative genome analyses was carried out.

MATERIALS AND METHODS

Bacterial host and phage isolation

A total of 73 *L. garvieae* strains previously isolated from different ecological niches (fish and dairy farms, vegetables, meat and cereals) (Ferrario et al., 2012) and 58 industrial *L. lactis* strains were used to determine the host range of the phage. Lactococcal strains were grown statically at 30°C in M17 broth (Oxoid) supplemented with 0.5% glucose (GM17). We tested samples of infected fish, raw and whey milk, wastewater, soil compost, aquaculture water, poultry and beef meat and sand from animal litter to isolate phages infecting *L. garvieae*.

All samples were kept at 4°C until used and the same phage isolation protocol was used (Haddad and Moineau, 2013). Five ml of each environmental sample (for the solid samples 5 g were resuspended in 5 ml of sterile water) were added to 5 ml of double strength GM17 supplemented with 10 mM $CaCl_2$, and inoculated with 100 µL of an overnight culture of an *L. garvieae* strain. Nineteen representative strains of previously studied *L. garvieae* (Ferrario et al. 2013) were used for lytic phage isolation. The mixtures were incubated overnight at 30°C. The cultures were centrifuged at 7,000 rpm for 10 min using an IEC Clinical Centrifuge (USA) and the supernatants were filtered (0.45 µm cellulose acetate filter). Five ml of the filtered supernatant were added to double strength GM17 inoculated as above. This second amplification procedure was repeated three more times. The presence of phages in the supernatants was tested by spot test, depositing 5 µL of the last amplification on a GM17 plate with 10 mM $CaCl_2$ containing a single *L. garvieae* strain. Clear plaques were picked, purified three times, and amplified as described (Emond et al., 1997). To obtain concentrated phage preparations, 1 L of lysate was mixed with polyethylene glycol 8000 (10%) and separated on a discontinuous CsCl gradient followed by a continuous CsCl gradient, as previously reported (Samson and Moineau, 2010).

Microbiological assays

The host ranges of phages isolated were assessed by spotting 5 µL of a 10^{-2} dilution of a phage lysate on top agar containing one lactococcal strain. For phage isolated from compost soil (named GE1), one-step growth assays were performed in triplicate with a multiplicity of infection (MOI) of 0.05 and the host strain *L. garvieae* INS1 (Ferrario et al., 2012). The burst size was calculated by dividing the average phage titer after the exponential phase by average titer before the infected cells began to release virions (Moineau et al., 1993).

Electron microscopy

Ten µl of uranyl-acetate (2%) was deposited on a Formvar carbon-coated grid (200 mesh; Pelco International). Cesium chloride-purified phages (10 µL) were mixed with the stain by pipetting up and down. After 2 min, the grid was deposited on blotting paper and dried for at least 5 minutes. Phages were observed at 80 kV using a JEOL1230 transmission electron microscope at the Plateforme d'Imagerie Moléculaire et Microscopie of the Université Laval. Capsid size and tail length were determined by measuring at least 10 phage specimens.

DNA extraction and restriction profile

DNA for restriction profile was extracted by using a previously described protocol (Moineau et al., 1994); *EcoRI* and *EcoRV* restriction endonucleases (Roche Diagnostics) were used as recommended by the manufacturer. The DNA fragments were separated in 0.8 agarose gel in 1X Tris-acetate-EDTA buffer and visualized by UV photography.

Phage DNA extraction, sequencing, and genome analysis

Phage DNA for sequencing, was extracted from 500 ml of lysate using the Qiagen Lambda Maxi DNA purification kit with modifications as described previously (Deveau et al., 2002). To determine the genome sequence of phage GE1, a sequencing library was first prepared with the Nextera XT DNA Sample Prep Kit (Illumina) according to the manufacturer's instructions. The library was sequenced using a MiSeq Reagent Kit v2 on a MiSeq system (Illumina). De novo assembly was performed with Ray assembler (Boisvert et al., 2010). Genome extremities were amplified using converging primers, and the PCR product was sequenced with an ABI 3730xl at the sequencing platform of the Centre Hospitalier de l'Université Laval. The *cos* site was also confirmed by direct sequencing of the phage DNA using forward (5'-GCAAGGAGGTAATCAGATGCA-3') and reverse (5'-GAACGCATTCTGTGAGCTTG-3') primers.

Open reading frame (ORF) prediction was carried out using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and RAST Server (Aziz et al., 2008). An ORF was considered valid only if its starting codon was AUG, UUG or GUG and it was at least 30 amino acids (aa) in length. Ribosomal binding sites (RBS) were also identified. Functions and domains were attributed by comparison of the translated products using BLASTp (Altschul et al., 1997). PSI-BLAST and InterProScan (http://www.ebi.ac.uk/) were used to search for more distant homologous proteins and conserved domains, respectively, when significant similarity was found by BLAST searches. Theoretical molecular masses (MM) and isoelectric points (pI) of the phage proteins were obtained using ProtParam (http://web.expasy.org/protparam/). tRNAs were identified using the tRNAscan-SE server (Lowe and Eddy, 1997) and confirmed using ARAGORN (Laslett and Canback , 2004). Virulence Factor Databases (Chen et al., 2012) together with DBETH (Chakraborty et al., 2012) were used to identify virulence factors.

Structural protein profile

Purified phages (~10^{11} PFU/mL) were dialyzed (0.02 M Tris-HCl pH 7.4, 0.1 M NaCl, 0.1 M $MgSO_4$), mixed with 4X loading buffer (0.25 M Tris-HCl pH 6.8, 40% [vol/vol] glycerol, 8% [wt/vol] SDS, 20% [vol/vol] β-mercaptoethanol, and 0.1% [wt/vol] bromophenol blue), and boiled for 5 min. Proteins were separated on an AnykD Mini-PROTEAN TGX Precast Protein Gel (Bio-Rad) and stained with Coomassie blue.

Proteomic tree design

The predicted proteome of *L. garvieae* phage GE1 was compared with other lactococcal phage proteomes, including the *L. garvieae* phage WP-2 (Ghasemi et al., 2014) and members of the 10 *L. lactis* phage groups (Deveau et al., 2006). The genomes were downloaded from NCBI and the proteins were extracted using the published annotations. The terminase was used as the starting point to generate a long phage protein concatemer for each phage (35 phages in total). The order of the proteins was identical to the gene order presented in the genome. For the two *L. lactis* phage genomes (KSY1 and P087) in which a terminase was not found, the protein concatemers were constructed following the gene order as deposited in GenBank. MEGA 5 (Tamura et al., 2011) software and a BLOSUM matrix were used for multiple alignments. An unrooted phylogenetic tree was constructed using the neighbor-joining algorithm (Dupuis and Moineau, 2010). The phylogeny was tested with 100 bootstraps replicates.

RESULTS AND DISCUSSION

Phage isolation and lytic spectrum assessment

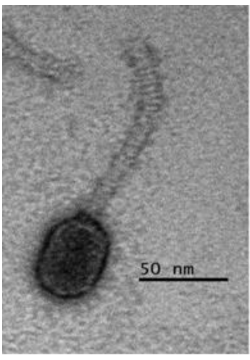
A total of 11 virulent phages were isolated from different environments (1 from compost soil, 5 from waste water and litter bedding, respectively).

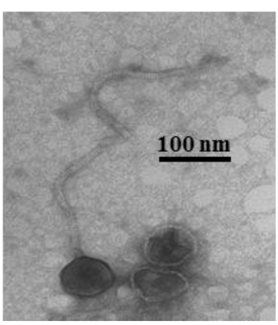
The lytic spectrum was determined by testing phage growth on 73 different *L. garvieae* strains. Each phage showed a different lytic spectrum profile, suggesting differences among the phages.

As reported for other common lactococcal phages, the *L. garvieae* phages are highly specific to a few host strains and this result highlights the design of particular phage cocktail to increase the efficiency of phage therapy.

Morphological, genomic and structural characterization

As reported in figure 1, the phages showed a typical *Siphoviridae* morphology (long noncontractile tail), with differences in capsid shape (11 icosahedral-shape and 1 prolate-head), and tail length.





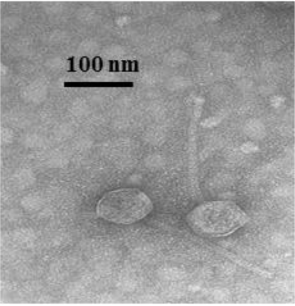
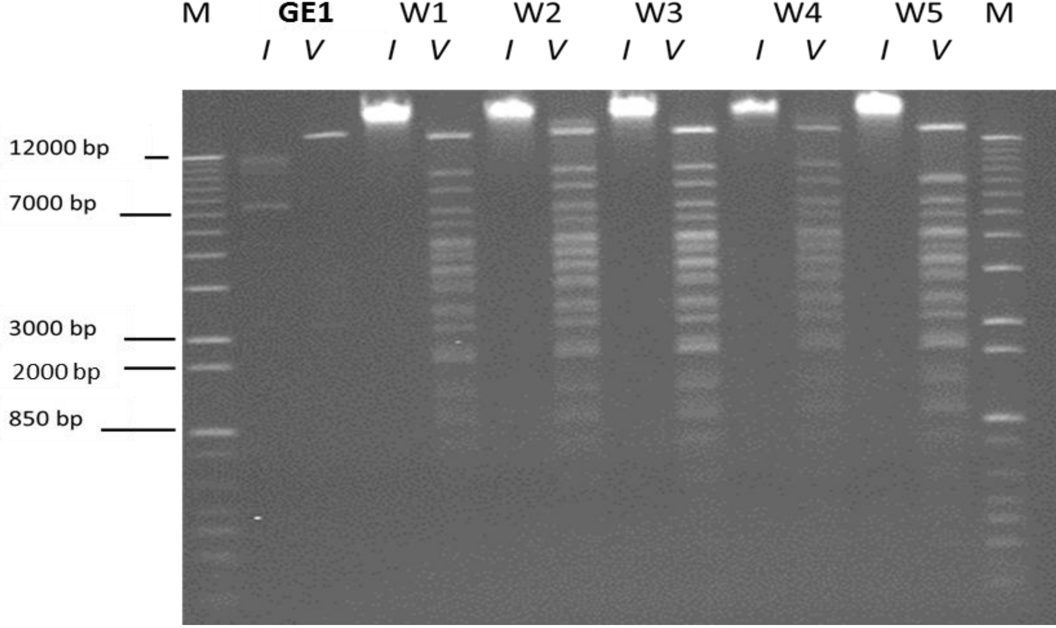


Figure 1. Morphological features of *L. garvieae* lytic phages. Left to right: a prolate head phage, an icosahedral capsid with a long tail (>400 nm) and a short tailed phage (<200 nm).

Restriction profile analysis (figure 2 A and B) revealed genome size ranging from 24 to 60 kb, in common to other lactococcal phages (Brussow, 2001; Labrie and Moineau, 2002; Samson and Moineau, 2010).
Samples from waste water (W) have a similar restriction profile, contrarily to isolated from animal litter (USA), showing for each sample a different restriction profile; since phages evolve together with their host, these findings might highlight the presence in animal litter of strains with a higher genomic variability compared to the other niche.
A)



B)

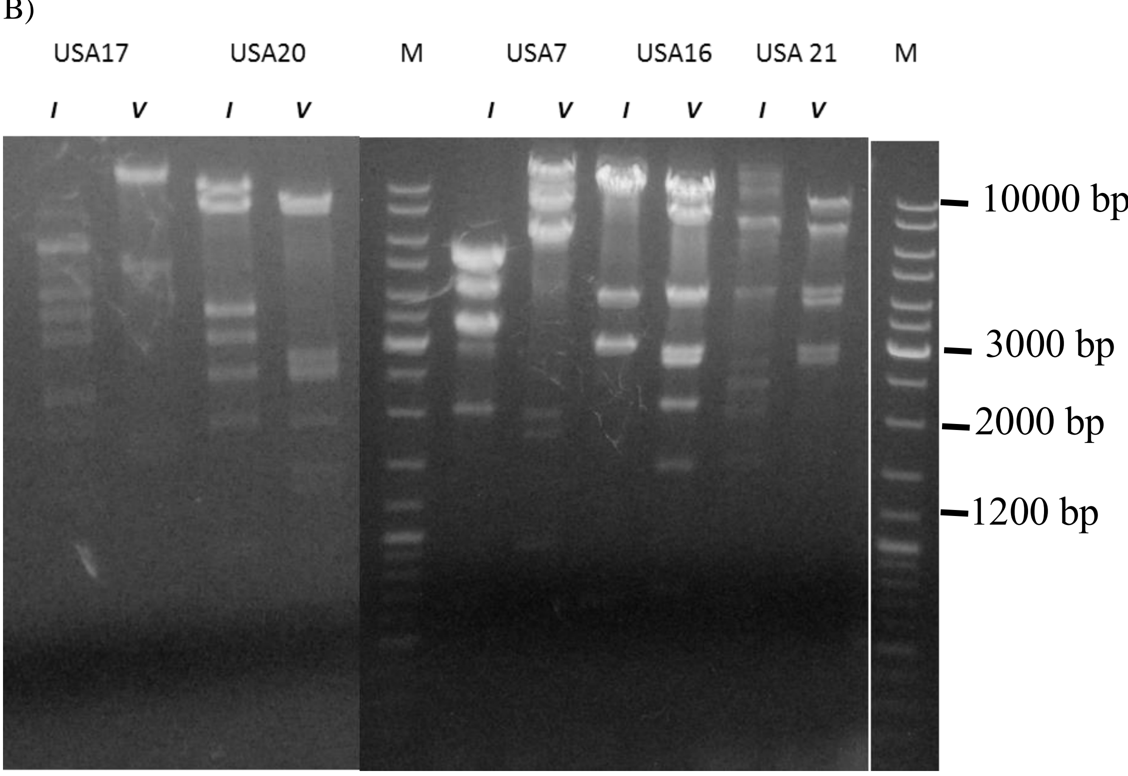


Figure 2 A and B. Restriction profile analysis of lytic phages. GE1, phage from compost soil; W, waste water and USA, from animal litter. *I: EcoRI* and *V:EcoRV*. M, molecular marker (GeneRuler, DNA Ladder Mix, Fermentas).

A structural profile was then carried out by randomly select three phage from waste water and the phage from compost soil. The result is reported in figure 3.
Even if the restriction profile revealed a high similarity among their genomes, phage W1 has a different structural profile and interestingly highly similar to that of phage from compost soil with which shares also the same host. It is possible to hypothesize that, even if the genome composition between W1 and GE1 is different, they might have evolved their morphological structures in order to infect specifically the host.

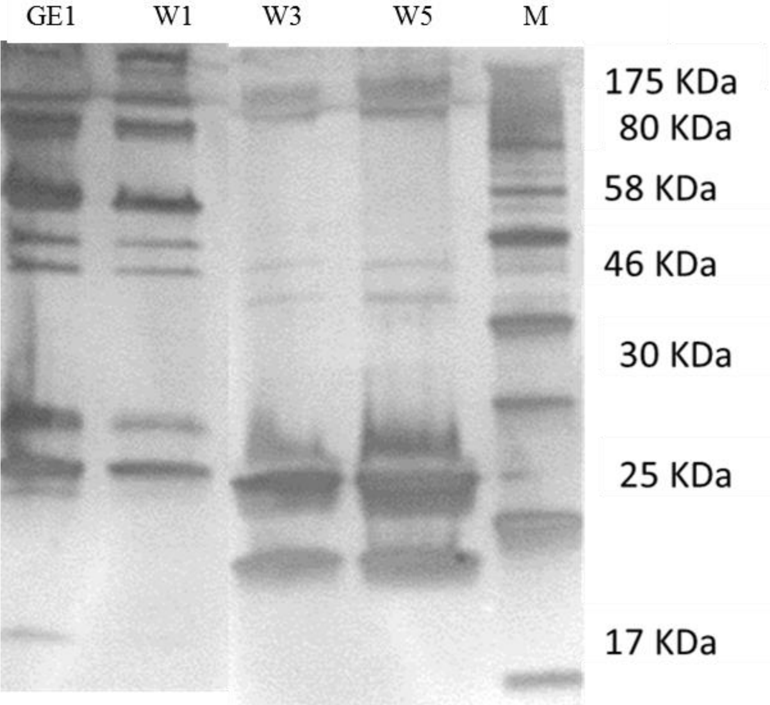


Figure 3. Structural protein profiles of phages. M, Molecular Markers (Bio-Rad). Numbers on the right indicate the molecular mass of the markers (Lane M).

In literature, little information is available on the genomic characterization of lytic phages infecting *L. garvieae*. For this reason, we sequenced the genome of phage GE1, isolated from a sample of compost soil, and the results are reported below.

Characterization of phage GE1

L. garvieae phage GE1, a virulent phage isolated from a compost soil sample, belongs to the Siphoviridae family (figure 4). It has a prolate capsid of 56nm ±3 nm by 38 nm ± 3 nm and a non-contractile tail of 123 nm ± 12 nm in length and 11 nm ± 2 nm in width. The phage morphology is highly reminiscent of phages c2 and Q54, members of two distinct *L. lactis* phage groups that bear their names (Deveau et al., 2006). However, the tail length of phage GE1 is slightly longer (Deveau et al., 2006).
The burst size of phage GE1 was calculated to be 139±13virions per infected cell and its latency period was 31±4 min, similar to other lactococcal phages (Dupuis and Moineau, 2010). The host range of phage GE1 was assessed using a collection of 58 industrially *L. lactis* strains. Phage GE1 was highly specific as it infected only two *L. garvieae* strains, its host INS1, as well as strain INS2, which was isolated from a similar ecological niche (vegetables) (Ferrario et al., 2012). Phage GE1 did not infect any of the 58 *L. lactis* strains tested.

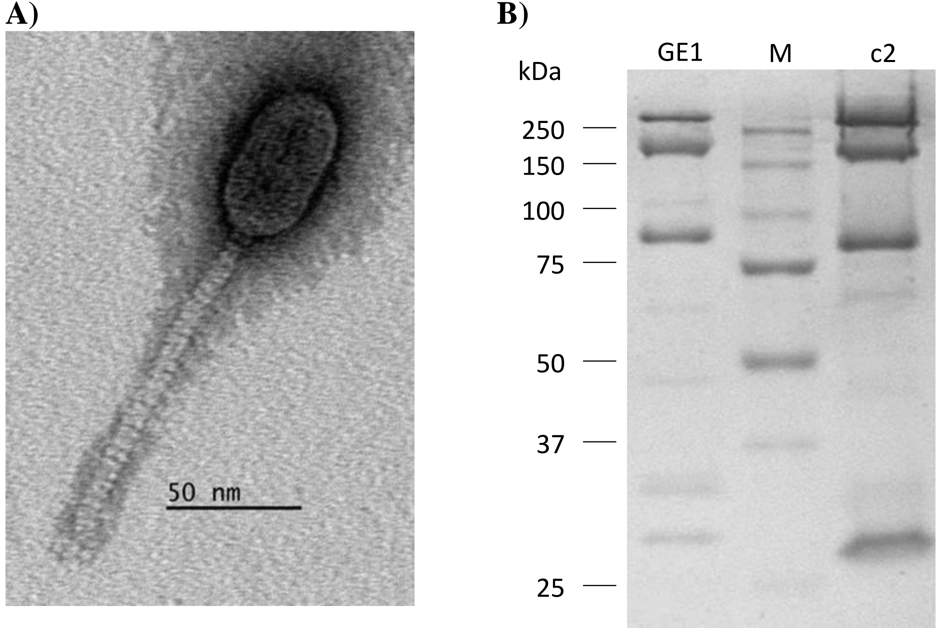


Figure 4. Morphology and structural proteome of phage GE1. Panel A) Electron micrograph of phage GE1. Panel B) Structural protein profiles of phages GE1 and c2.M, Molecular Markers (Bio-Rad). Numbers on the left indicate the molecular mass of the markers (Lane M).

Genome of phage GE1 and its *cos* site

Phage GE1 has a 24,847 bp linear double-stranded DNA genome with a GC content of 37.8%, similar to the GC content of its host genome (Ferrario et al., 2013) but slightly higher than that of *L. lactis* phage genomes (35%) (Dupuis and Moineau, 2010). Analysis of the phage GE1 genome did not identify any tRNA or recognizable virulence factors. The genome has cohesive ends consisting of single-stranded 9-bp 3' overhangs (figure 5a). Analysis of the DNA region surrounding the *cos* site revealed characteristics similar to *L. lactis* phages belonging to the c2 group (Fortier et al., 2006). This region is characterized by inverted repeat sequences as well as 22 direct repeats. Overall, this region is AT-rich (69%) compared to the rest of the GE1 genome (62%). Three G boxes and no C clusters were found near the cos site (figure 5a).
The inverted repeat found in the cos site of phage GE1 (GCAA) lacks a conserved T in one of the consensus sequences (*TCAN*(N4-7)*NACT*), typically found within a 15-bp segment spanning the cos site of many *L. lactis* phages (Perrin et al., 1997). The absence of this conserved base in the consensus cos site is reminiscent of the *cos* site of lactococcal phage Q54, which has a CCAA inverted repeat.
The *cos* region of GE1 also contains five λ-like R consensus sequences, which are usually involved in terminase recognition and binding, as well as in packaging termination (David Cue, 1993). The sequences were compared with those previously

reported for *L. lactis* phages c2, bIL67 (c2-like), and Q54. As reported for phage Q54 (Fortier et al., 2006), phage GE1's five λ-R sequences are unevenly distributed, but across a smaller genomic region (500 bp compared to 900 bp of phage Q54). This contrast with coliphage λ and lactococcal phage c2 whose λ-R sequences are regularly spaced. Alignment analysis of the R sites from GE1 (figure 5b) found bases in common involved in the packaging process, similar to those observed for c2 and Q54.

A)

```
ttaattggtgatgtcatcttttatataacctgctttctttagagtgttgtatatagagttaatggc

aaggaggtaatcagatgcagttaatagatacctgttagtgtaattcgttactccttttatgtcctc
                                          ORF 48
tgttattgtattcaatataatttgctcttctgttaagttgctattattcatAATATATTTCCTTTC

TATATATATAGTCTATCATAATACATTACTATTGTCAACTGTTATGTGATCATGTGTGTGCTTAGT
                                                    A/T-richregion
ACTGTGCTTACTCATCTATGTCTATGTATATCTGTTGCTTATTGATTGAATGAAATAATAAATAAA

TGTAAAGAGAAAGGTAAAGAATTAAATGAAAAAGATTTCGTGTAAATTGTAAATTTTGTCAGGAAG
                                       G box
AGTGGCGTAAGTGTACGGGGGTGTAATAATTGGGGGTAGGGGGCAAGAGTGCAACAACTCCTTCGC
                   A/T-richregion
ATTTACATTTCAATTTTCATTTTCAATTATTTTCATAATATAAAAACCTTACACAAGCTCACAGAA

TGCGTTCTAAGCTCATATAAGGTATCTTGctataatgttatcttaacttctttaaactctttctgt

aacagctctagagcatcatgcaccctttcttct//ttttctttatctaaaacattgcgtaacttag
                                                            ORF 1
ctttacaccaaactgtaccaggtttctctttcaatttctctaaccattcttctaagttcat
```

B)

```
Q54     R1   ATCTTAAACAGAAATGCAA
        R2   TCTAAAATCAGAAAGAATA
        R3   TAATAAAAAGAAATAAAC
        R4   TTTTAGACAAGAAAGGAGG
        R5   AAAGCAAGAAGAAAGAGCC
        R6   AATGGGAAAAGAAAATTTT
        R7   CTTAAGAAGCGAAACATAT
        R8   CTACACAAGAGAAACACGA
        R9   GCAATAAAACGAAAAGAAA
GE1     R1   TTGATTGAATGAAATAATA
        R2   AATGTAAAGAGAAAGGTAA
        R3   AGAATTAAATGAAAAAGAT
        R4   CAAACTCAAAGAAAGGGTC
        R5   ATGTCAGTAAGAAACGGTT
c2      R1   TAAGAAAGAAGAAAGAAAG
        R2   TAACTAAAAAGAAAGAAGA
        R3   TTTTCAAAACGAAAAATGG
        R4   TAATAAAATAGAAAAGTGA
```

Figure 5. Analysis of GE1 *cos* region. (A) Analysis of the *cos* site and the flanking regions. Direct repeats (bold arrows), inverted repeats (bold, dash arrows), A/T rich regions and G-rich segments are indicated. Details of the cohesive termini (rectangle) are also shown together with the site of cleavage (vertical arrows). Shading indicated nucleotides that belong to the consensus sequence found in all lactococcal phage for which the *cos* site has been determined. (B) Multiple alignments of λ-like R consensus sequences from Q54, GE1, and c2. Conserved nucleotides are shaded.

Analysis of phage GE1 genes

A total of 48orfs longer than 30 codons were predicted from the genome sequence (table 1). The size of the gene products range from 44 (ORF17 and ORF26) to 839 amino acids (ORF37), with an average predicted protein size of 153 amino acids. Coding regions represent approximately 88% of the genome and the longest non-coding region is 657 bp, occuring between orf27 and orf28. As for c2-like lactococcal phages, the GE1 genome is organized in two main gene clusters, grouping the presumed early- and late-expressed genes (Brussow, 2001; Lubbers et al., 1995). The late-expressed genes code for proteins involved in packaging (orf40), cell lysis (orf29 and orf38) and phage morphology (orf32, orf34, and orf37). No lysogeny module was found in the genome, confirming the lytic nature of phage GE1.
Seven orfs (1, 11, 14, 20, 24, 27, and 28) possess a typical RBS domain (AGGAGA). Of these, orf27 and orf28 are located at the ends of the 657-bp intergenic region between early- and late-expressed genes. Moreover, orf27 (early gene) is also preceded by the consensus -10 and -35 promoter sequences (TTGACA-17 bp-TATAAT), whereas upstream of orf28 (late gene) only a-10 consensus sequence was found. These characteristics suggest that the region between orf27 and orf28 could act as an origin of replication, as already reported for phage c2 (Lubbers et al., 1994).

Proteome of phage GE1

The functions of phage GE1 deduced proteins were determined by comparing (BLASTp) with the GenBank database. When low similarity was found between GE1 proteins and other phage proteins, putative functions were reinforced by identifying conserved domains and/or position-specific iterative comparisons. Only ORFs with a significant protein hit in GenBank are presented in table 1. Predicted functions were attributed to 14 ORFs (out of 48 ORFs, 29%) and only 17 ORFs (35%) matched proteins in the GenBank database, indicating that new viral genes/proteins were identified by characterizing this novel phage. Details are given below for some ORFs.
Homing endonucleases and recombination genes.Three ORFs (ORF9, ORF13 and ORF39) were predicted to be homing endonucleases. These ORFs had similarity with deduced proteins found in Lactococcus phages 712 and P745 (Mahony et al., 2013), both 936-like phages, as well as with *L.lactis* phage Q54 (Fortier et al., 2006). Since homing endonucleases promote genetic exchange and horizontal transfer (Belfort and Bonocora, 2014), this part of the phage GE1 genome may have arisen through the acquisition of mobile DNA, as already proposed for phage 712 (Mahony et al. 2006). ORF15 and ORF22 had similarity to proteins with putative recombination functions,

found in *Streptococcus pseudopneumoniae* and 936-like phage jm3 (Mahony et al., 2013).
Replication. ORF12 and ORF14 have homology with a glutamyl-tRNA synthase protein found in *Lactococcus* phage c2 and with a single-stranded binding protein of lactococcal 936-like phages, such as phi7 and P113G (Mahony et al., 2013). Glutamate tRNA synthase catalyses the attachment of the amino acid glutamine to its cognate transfer RNA molecule in a highly specific two-step reaction (Freist W et al., 1997). Single stranded binding proteins are required during DNA replication, repair and recombination (Brussow, 2001; Scaltriti et al., 2009, 2013).
Morphogenesis and lysis. Phage GE1 structural proteins (ORF32/major capsid protein; ORF34/ major tail protein; ORF37/minor tail protein) and proteins involved in packaging (ORF40/terminase) and lysis (ORF29/holin; ORF38/holin) showed similarities to proteins found in other phages possessing the same morphological features, particularly c2, Q54, CB17, 5447, and 923 (Deveau et al. 2006; Rakonjac et al., 2005). ORF35 possesses an immunoglobulin-like domain but did not match with other proteins in databases. Immunoglobulin-like proteins may be involved in host recognition (Fraser et al., 2007). The structural protein profile of phage GE1 was determined by SDS-PAGE (figure 4B) and found to be highly related to phage c2 (Fortier et al. 2006).

Table 1.Features of the ORFs of *L. garvieae* phage GE1.

ORF	Strand[a]	Positions Start	Stop	Size aa	MM[b]kDa	pI	Putative RBS and start codon[c]	Predicted function[d]	Best-match BLASTp result	# aa shared with best match / total # aa in best match (% ID)	E value	Size[e] (aa)	Accession numbers
1	-	529	104	141	16.9	6.8	AGGAGActagaATG	-					
2	-	678	526	50	6.1	9.8	aAGGAGtgcaagctATG	-					
3	-	854	675	59	7.2	9.3	AAGAAAGgcaatgtagtgaaagaATG	-					
4	-	1039	842	65	7.7	9.5	AAGAAAGAggtagaaaATG	-					
5	-	1295	1029	88	10.1	9.8	AAGAAAGaaaacatctaaaATG	-					
6	-	1650	1477	57	6.8	8.7	gAAGAGGtatagaaATG	-					
7	-	1949	1653	98	11.1	4.5	AGAGGAttgagATG	-					
8	-	2256	1951	101	11.5	4.5	GAGgtGAAaagagcATG	-					
9	-	2678	2253	141	16.2	10.1	AGGAGtttatagaccaGTG	HNH endonuclease	ORF31 *Lactococcus* phage 712	66/137 (48)	3e-32	141	YP_764291.1
10	-	2947	2675	90	10.4	7.9	GAAAGGGGAtatatATG	-					
11	-	3588	3073	171	20.0	9.4	AGGAGAactttaATG	-	ORF41*Lactococcus* phage 191	54/162(33%)	5e-11	177	AFE86775.1
12	-	3770	3588	60	7.1	4.7	GAAAGGccattATG	tRNAsynthetase	ORF6*Lactococcus* phage c2	24/52(46%)	0.15	54	NP_043532.1
13	-	4179	3754	141	16.4	9.8	ctgacttgccattttaacATG	HNH endonuclease	ORF48 *L .lactis* phage P475	74/142 (52)	3e-36	145	AGI11121.1
14	-	4510	4181	109	11.9	4.7	AAGAGAatataaattATG	ssDNA-binding protein	ORF40*Lactococcus* phage 7	49/117 (42)	2e-18	119	YP_008318242.1
15	-	5031	4531	166	18.3	7.8	AGGAAAttgaattcgcATG	ERF recombinase	Rec, *S .pseudopneumoniae*	58/123 (47)	1e-29	207	WP_000163473.1
16	-	5231	5028	67	7.8	5.2	AGGAGGacaagaagaaATG	-					
17	-	5402	5268	44	5.4	6.7	atacAGAGGtggatagATG	-					
18	-	5637	5389	82	10.0	9.5	AtgAGGGAtttaagttagATG	-	ORF10 *Lactococcus* phage c2	31/77(40%)	2e-05	83	NP_043536.1
19	-	5920	5621	99	11.9	4.9	CGTAGAGGActttggcaagtATG	-					
20	-	6176	5913	87	10.3	4.6	AGGAGAgtaatATG	-	gp25*Lactococcus* phage KSY1	27/60(45%)	3e-07	99	YP_001469023.1
21	-	6531	6352	59	7.1	7.9	AGGttGAGAtattagaATG	-					
22	-	7478	6528	316	36.1	5.4	AAGGAGtAaatacaGTG	DNA polymerase	ORF42 *Lactococcus* phage jm3	166/317 (52)	6e-106	315	YP_008318192.1
23	-	7737	7465	90	10.5	8.9	GGAAAGGctttggattATG	-					
24	-	8148	7921	75	9.0	8.9	AGGAGAaaacaaaATG	-					
25	-	8355	8218	45	5.4	9.4	AAAGGAcggtgttattATG	-					
26	-	8476	8342	44	5.1	9.7	GAGAGGtttagaaaATG	-					
27	-	8651	8466	61	7.3	4.7	AGGAGActattatATG	-					
28	+	9309	9461	50	5.8	4.5	AGGAGAaaggtttATG	-					
29	+	9463	9951	162	18.2	5.2	tactcattacctatttaatATG	Holin	ORF24 *Lactococcus* phage c2	92/159(58)	1e-59	161	NP_043550.1
30	+	10496	11167	223	24.9	9.0	AAAGGAAAtaattaaTTG	-					
31	+	11177	12109	310	34.7	5.2	AAAtAGAGGtaacaaATG	-					

32	+	12096	13667	523	57.6	4.9	GAAGAGtgaggtagaagATG	Major capsid protein	MCP *Lactococcus* phage CB17	130/535 (24)	9e-22	489	AAZ95019.1
33	+	13684	14484	266	28.9	5.2	AAAGAGgtaacttATG	-					
34	+	14486	15100	204	22.4	4.7	AAGAAAGGtataatttaatATG	Major tail protein	ORF23 *Lactococcus* phage Q54	65/200 (33)	1e-21	212	YP_762592.1
35	+	15151	15342	63	6.7	6.0	AGGGGGAAggTTG	Ig-like domain					
36	+	15366	15617	83	9.6	5.6	AGGGGActtaatATG	-					
37	+	15798	18317	839	87.8	5.5	AGGAAAttaatTTG	Tail protein	Pl10*Lactococcus* phage 5447	123/471 (26)	3e-22	622	AAT73602.1
38	+	18331	18606	91	10.2	9.3	AGGGAAaataATG	Holin	Gl17 *Lactococcus* phage 923	54/85 (65)	1e-27	96	AAT81366.1
39	+	18684	18968	94	10.9	8.3	AGcttgtGGGGGtAtcATG	HNH endonuclese	Gp27*Lactococcus* phage Q54	44/101 (44)	4e-16	103	YP_762596.1
40	+	18968	20605	545	61.3	6.6	AtGGGGtGGtaATG	Terminase	ORF28 *Lactococcus* phage Q54	191/507 (38)	2e-96	514	YP_762597.1
41	+	20605	20901	98	11.4	9.1	AGGGAAattaatctaATG	-	ORF35 *Lactococcus* phage c2	20/65(31%)	0.005	99	NP_043561.1
42	+	20901	22100	399	46.6	5.5	AAGGAGtttatgtaATG	-					
43	+	22100	22759	219	24.9	5.8	GAAAGGttacgggtataatttaATG	-					
44	+	22759	23526	255	27.5	4.8	GAAGGGttatgttggtggattaATG	-					
45	-	23708	23568	46	4.9	9.5	AGGAGAacatcATG	-					
46	-	24237	23797	146	16.4	10.2	GAAAGGtggttatatgtaATG	-					
47	-	24410	24237	57	6.6	4.0	AAGcAGGttatataaaagATG	-					
48	-	24576	24394	60	6.8	4.7	AGGAAAtatattATG	-					

MCP, Major Capsid Protein; Rec, Recombinase; *S., Streptococcus.*

[a]Orientation of the gene in the genome.

[b]MM, molecular mass.

[c]RBS, ribosomal binding site: uppercase letters represent hypothetical RBS sequences, bold letters represent starting codons.

[d] – indicates no significant matches.

[e]Total size of the aligned proteins.

Comparison of prolate-headed phages

As demonstrated above, phage GE1 has similarities to L. lactis phages c2 and Q54, both prolate-headed phages. The genome organization and proteome of phage GE1 were specifically compared to those of L. lactis phages c2 and Q54. Figure 6 clearly confirms the relatedness of phage GE1 to c2 and Q54 phages. GE1 possesses 13 ORFs that align with those of phage c2 (amino acid identity of 25 to 58%). They are principally involved in replication (six ORFs) and morphogenesis (seven ORFs), and five of these seven proteins were also common to phage Q54 (amino acid identity of 27 to 44%). However, the genome organization of phage GE1 is closer to phage c2 than Q54, as illustrated by gene orientation (figure 6).

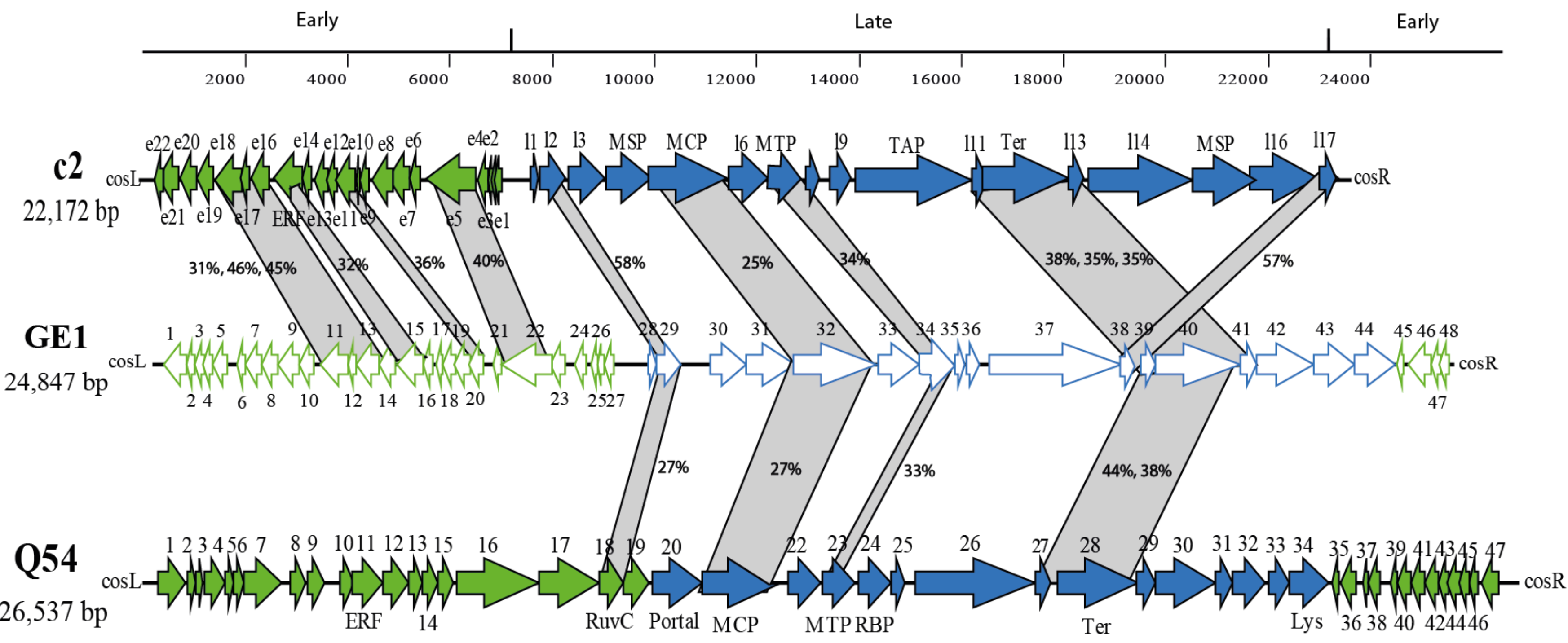


Figure 6. Genomic comparison of GE1, Q54 and c2 phages. ORFs from GE1 showing similarity with other phages are linked by gray shading and the percent amino acid identity indicated is representative of the aligned region.

Phylogenetic studies of *L. garvieae* phages

Finally, a phylogenetic analysis was performed using *L.garvieae* phages (GE1 and WP-2) as well as members of the 10 currently recognized groups of *L. lactis* phages to confirm the relationships described above. Figure 7 shows a proteomic phylogenetic tree constructed using MEGA 5 software and the neighbor-joining method. This global comparison revealed that the two *L. garvieae* phages are distributed into *L. lactis* phages, indicating that they probably evolved from a common ancestor. The genomic comparison also demonstrated that phage GE1 originates at the same point as other prolate-headed phages, diverging, together with phage Q54, in a separate branch.

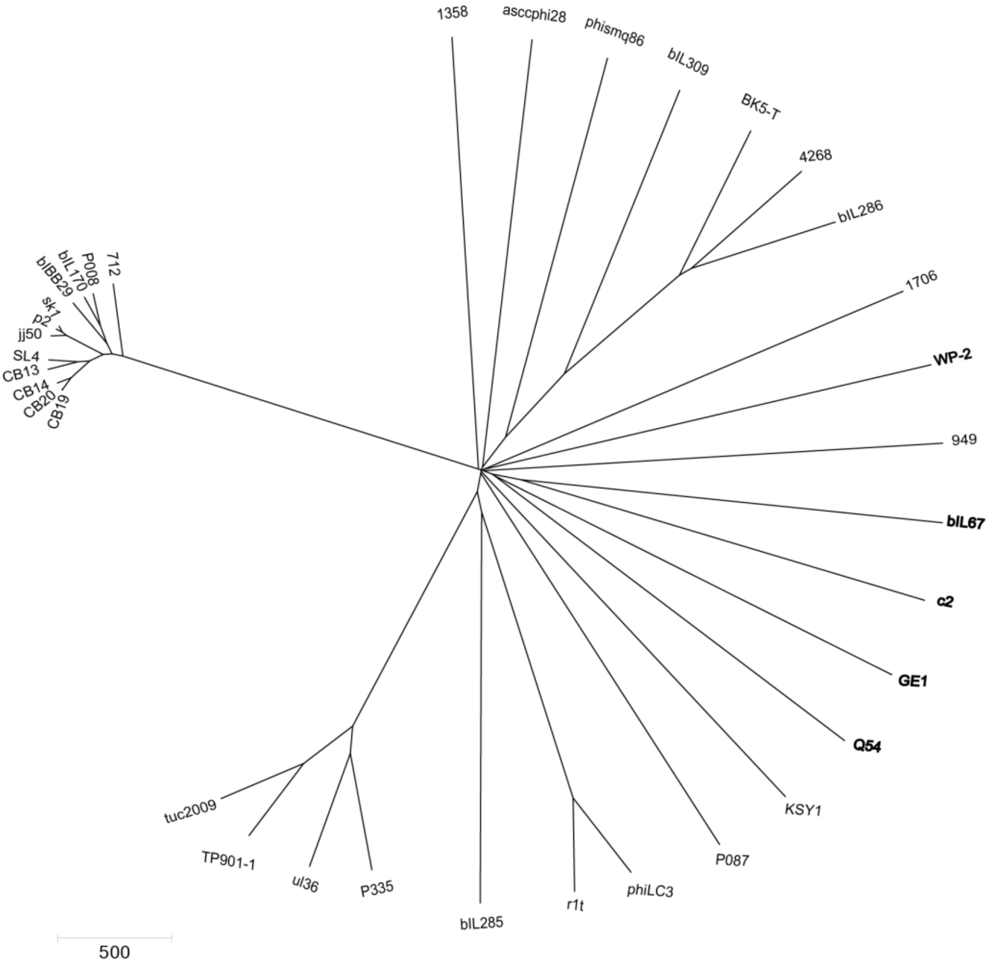


Figure 7. Phylogenetic tree of different lactococcal phage proteomes determined using the neighbor-joining algorithm. The bar shows the difference between the phages in amino acids.

CONCLUSIONS

The main conclusions of this chapter are:

- isolation of lytic phages against *L. garvieae* strains;
- their narrow lytic spectrum requires the design of phage cocktail to increase efficiency of phage therapy;
- genomic characterization of a lytic phage infecting *L. garvieae* revealed homology with other *L. lactis* phages highlighting the presence of a common ancestor.

Related publication: Eraclio G., Tremblay DM, Lacelle-Côté A, Labrie SJ, Fortina MG, Moineau S. (2015). *A new virulent phage infecting* Lactococcus garvieae*, with homology to* Lactococcus lactis *phages.* Applied and Environmental Microbiology. doi: 10.1128/AEM.02603-15.

REFERENCES

1. Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman DJ. (1997). *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Res. 25:3389-3402.

2. Aziz R.K., Bartels D., Best A.A., DeJongh M. Disz T., Edwards R.A., Formsma K., Gerdes S., Glass E.M., Kubal M., Meyer F., Olsen G.J., Olson R., Osterman A.L., Overbeek R.A., McNeil L.K., Paarmann D., Paczian T., Parrello B., Pusch G.D., Reich C., Stevens R., Vassieva O., Vonstein V., Wilke A., Zagnitko O. (2008). *The RAST server: rapid annotations using subsystems technology*. BMC Genomics. 9:75.

3. Belfort M., Bonocora R.P. 2014. *Homing endonucleases: from genetic anomalies to programmable genomic clippers*. Methods Mol Biol. 1123:1–26.

4. Boisvert S, Laviolette F, Corbeil J. (2010). *Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies*. J Comput Biol. 17:1519–1533.

5. Brüssow H. (2001). *Phages of dairy bacteria*. Annu Rev Microbiol. 55:283–303.
6. Chakraborty A., Ghosh S., Chowdhary G., Maulik U., Chakrabarti S. (2012). *DBETH: a database of bacterial exotoxins for human*. Nucleic Acids Res. 40:D615-620.

6.Chakraborty A., Ghosh S., Chowdhary G., Maulik U, Chakrabarti S. (2012). *DBETH: database of bacterial exotoxins for human.* Nucleic Acids Res. 40:615-620.

7. Chen L., Xiong Z., Sun L., Yang J., Jin Q. (2012). *VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors*. Nucleic Acids Res. 40:D641-645.

8. David Cue M.F. (1993). *The role of* cosB, *the binding site for terminase, the DNA packaging enzyme of bacteriophage lambda, in the nicking reaction*. J Mol Biol. 234:594–609.

9. Deveau H., Van Calsteren M.R., Moineau S. (2002). *Effect of exopolysaccharides on phage-host interactions in* Lactococcus lactis. Appl Environ Microbiol. 68:4364–4369.

10. Deveau H., Labrie S.J., Chopin M.C., Moineau S. (2006). *Biodiversity and classification of lactococcal phages*. Appl Environ Microbiol.72:4338–4346.

11. Dupuis M.E., Moineau S. (2010). Genome *organization and characterization of the virulent lactococcal phage 1358 and its similarities to* Listeria *phages*. Appl Environ Microbiol. 76:1623–1632.

12. Émond É., Holler B.J., Boucher I., Vandenbergh P.A., Vedamuthu E.R., Kondo J.K., Moineau S. (1997). *Phenotypic and genetic characterization of the bacteriophage abortive infection mechanism AbiK from* Lactococcus lactis. Appl Environ Microbiol. 63:1274–1283.

13. Ferrario C., Ricci G., Milani C., Lugli G.A., Ventura M., Eraclio G., Borgo F., Fortina M.G. (2013). Lactococcus garvieae*: where is it from? A first approach to explore the evolutionary history of this emerging pathogen*. PLoS One 8. doi: 10.1371/journal.pone.0084796.

14. Ferrario C., Ricci G., Borgo F., Rollando A., Fortina M.G. (2012). *Genetic investigation within* Lactococcus garvieae *revealed two genomic lineages*. FEMS Microbiol Lett. 332:153-161.

15. Fortier L.C., Bransi A., Moineau S. (2006). *Genome sequence and global gene expression of Q54, a new phage species linking the 936 and c2 phage species of* Lactococcus lactis. J Bacteriol. 188:6101–6114.

16. Fraser J.S., Maxwell K.L., Davidson A.R. (2007). *Immunoglobulin-like domains on bacteriophage: weapons of modest damage?* Curr Opin Microbiol. 10:382-387.

17. Freist W., Gauss D.H., Ibba M.S.D. (1997). *Glutaminyl-tRNA synthetase*. Biol Chem. 378:1313–1329.

18. Ghasemi S.M., Bouzari M., Yoon B.H., Chang H.I. (2014). C*omparative genomic analysis of* Lactococcus garvieae *phage WP-2, a new member of* Picovirinae *subfamily of* Podoviridae. Gene. 551:222–229.

19. El Haddad L., Moineau S. (2013). *Characterization of a novel panton-valentine leukocidin (PVL)-encoding staphylococcal phage and its naturally PVL-lacking variant*. Appl Environ Microbiol. 79:2828–2832.

20. Labrie S., Moineau S. (2002). *Complete genomic sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of Lactococcal phages*. Virology. 296(2):308–320.

21. Laslett D., Canback B. (2004). *ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences*. Nucleic Acids Res. 32:11–16.

22. Lowe T.M., Eddy S.R. (1997). *tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence*. Nucleic Acids Res. 25:955–964.

23. Lubbers M.W., Waterfield N.R., Beresford T.P.J., Le Page R.W.F., Jarvis A.W. (1995). *Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2*

genome and identification of the structural genes. Appl Environ Microbiol. 61:4348–4356.

24. Lubbers M.W., Ward L.J., Beresford T.P., Jarvis B.D., Jarvis A.W. (1994). *Sequencing and analysis of the* cos *region of the lactococcal bacteriophage c2*. Mol Gen Genet. 245:160–166.

25. Mahony J., Kot W., Murphy J., Ainsworth S., Neve H., Hansen L.H., Heller K.J., Sørensen S.J., Hammer K., Cambillau C., Vogensen F.K., Van Sinderen D. (2013). *Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny*. Appl Environ Microbiol. 79:4385–4392.

26. Mahony J., Deveau H., Mc Grath S., Ventura M., Canchaya C., Moineau S., Fitzgerald G.F., Van Sinderen D. (2006). *Sequence and comparative genomic analysis of lactococcal bacteriophages jj50, 712 and P008: evolutionary insights into the 936 phage species*. FEMS Microbiol Lett. 261:253–261.

27. Moineau S., Pandian S.,. Klaenhammer T. R. (1994). *Evolution of a lytic bacteriophage via DNA acquisition from the* Lactococcus lactis *chromosome*. Applied and Environmental Microbiology. 60:1832–1841.

28. Moineau S., Durmaz E., Pandian S., Klaenhammer T.R. (1993). *Differentiation of two abortive mechanisms by using monoclonal antibodies directed toward lactococcal bacteriophage capsid proteins.* Appl Environ Microbiol. 59:208–212.

29. Perrin R., Billard P., Branlant C. (1997). *Comparative analysis of the genomic DNA terminal regions of the lactococcal bacteriophages from species c2*. Res Microbiol. 148:573–583.

30. Rakonjac J., O'Toole P.W., Lubbers M. (2005). *Isolation of lactococcal prolate phage-phage recombinants by an enrichment strategy reveals two novel host range determinants*. J Bacteriol. 187:3110–3121.

31. Samson J.E., Moineau S. (2010). *Characterization of L*actococcus lactis *phage 949 and comparison with other lactococcal phages*. Appl Environ Microbiol. 76:6843–6852.

32. Scaltriti E., Polverini E., Grolli S., Eufemi E., Moineau S., Cambillau C., Ramoni R. (2013). *The DNA binding mechanism of a SSB protein from* Lactococcus lactis *siphophage p2.* Biochim Biophys Acta. 1834:1070–1076.

33. Scaltriti E., Tegoni M., Rivetti C., Launay H., Masson J.Y., Magadan A.H., Tremblay D., Moineau S., Ramoni R., Lichière J., Campanacci V., Cambillau C.,

Ortiz-Lombardía M. (2009). *Structure and function of phage p2 ORF34(p2), a new type of single-stranded DNA binding protein*. Mol Microbiol. 73:1156–1170.

34. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

ACKNOWLEDGEMENTS

This thesis work was carried out between the years 2012-2015 at the Department of Food, Environmental and Nutritional Sciences of the University of Milan, Italy. Financial support from the Graduate School in Molecular Sciences and Plant, Food and Environmental Biotechnology is gratefully acknowledged.

First of all I would like to thank prof. Maria Grazia Fortina for her technical and financial support during these years.
Thanks to Giovanni Ricci, Mattia Quattrini, Gaia Serra, Ilaria Vecchietti and Paolo D'Incecco for the accurate support and help during the experiments. Thanks to Prof. Paolo Moroni for providing American strains, Dr. Stefania Arioli and Dr. Maria Chiara Remagni for the precious advice.
Secondly, a special thank goes to prof. Sylvain Moineau for his technical and financial support; many thanks to Denise Tremblay for electronic micrograph, and also to all the people of the lab: your continuous availability made possible a part of this thesis.

Finally, a thank goes to my family: my parents Giacomo and Giulia for everything; my brother Andrea, his girlfriend Nadia and their baby Giacomo: I wish you all the best.

LIST OF PAPER, ORAL COMMUNICATIONS AND POSTERS

Peer-reviewed publications

Paper I Eraclio G., Ricci G., Quattrini M., Moroni P, Fortina M.G. (2015). *Virulence-related genes in* Lactococcus garvieae *coming from different ecological niches and geographical areas*. Submitted FEMS Microbiology Letters.

Paper II Eraclio G., Fortina M.G., Tremblay D.M., D'Incecco P., Pelegrino L., Moineau S. (2015). *Primary characterization of temperate bacteriophages infecting* Lactococcus garvieae *reveals new insight into the evolution of the species.* Submitted Applied and Environmental Microbiology.

Paper III Eraclio G., Tremblay D.M., Lacelle-Côté A., Labrie S.J., Fortina M.G., Moineau S. (2015*). A new virulent phage infecting* Lactococcus garvieae, *with homology to* Lactococcus lactis *phages*. Applied and Environmental Microbiology. doi: 10.1128/AEM.02603-15.

Paper IV Eraclio G., Ricci G., Fortina M.G. (2014). *Insertion sequence elements in* Lactococcus garvieae. Gene 555(2):291-296.

Paper V Ferrario C., Ricci G., Milani C., Lugli G.A., Ventura M., Eraclio G., Borgo F., Fortina M.G. (2013). Lactococcus garvieae: where is it from? *A first approach to explore the evolutionary history of this emerging pathogen*. PLoS One. 8(12):e84796.

Paper VI Ricci G., Ferrario C., Borgo F., Eraclio G., Fortina M.G. (2013). *Genome sequences of two* Lactococcus garvieae *strains isolated from meat.* Genome Announc. e00018-12.

Posters and oral communications in scientific congress

I Eraclio G., Tremblay D., Moineau S., Fortina M.G. (2015). Lactococcus garvieae*'s viruses: an insight into temperate and virulent bacteriophages.* Oral presentation in 6th Congress of European Microbiologists (FEMS), Maastricht (The Netherlands) 7-11 June 2015.

II Eraclio G. (2015). *Polyphasic approach to study the evolutionary history of* Lactococcus garvieae. Oral communication in 20th Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology, Perugia (Italy), 23-25 September 2015.

III Eraclio G. (2014). *Insertion sequences and temperate phages along the genome of the opportunistic pathogen* Lactococcus garvieae. Poster in 19th Workshop on the Developments in the Italian PhD Research on Food Science, Technology and Biotechnology, Bari (Italy), 24-26 September 2014.

IV Eraclio G. (2013). *Emerging pathogens in food: new approaches for their control*. Poster in 18th Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology, Conegliano (Italy), 25-27 September 2013.