

Lactococcus garvieae: Where Is It From? A First Approach to Explore the Evolutionary History of This Emerging Pathogen

Chiara Ferrario¹, Giovanni Ricci¹, Christian Milani², Gabriele Andrea Lugli², Marco Ventura², Giovanni Eraclio¹, Francesca Borgo¹, Maria Grazia Fortina^{1*}

¹ Department of Food, Environmental and Nutritional Sciences (DeFENS) - Division of Food Microbiology and Bioprocesses, Università degli Studi di Milano, Milan, Italy,

² Department of Life Sciences, Laboratory of Probiogenomics, Università di Parma, Parma, Italy

Abstract

The population structure and diversity of *Lactococcus garvieae*, an emerging pathogen of increasing clinical significance, was determined at both gene and genome level. Selected lactococcal isolates of various origins were analyzed by a multi locus sequence typing (MLST). This gene-based analysis was compared to genomic characteristics, estimated through the complete genome sequences available in database. The MLST identified two branches containing the majority of the strains and two branches bearing one strain each. One strain was particularly differentiated from the other *L. garvieae* strains, showing a significant genetic distance. The genomic characteristics, correlated to the MLST-based phylogeny, indicated that this “separated strain” appeared first and could be considered the evolutionary intermediate between *Lactococcus lactis* and *L. garvieae* main clusters. A preliminary genome analysis of *L. garvieae* indicated a pan-genome constituted of about 4100 genes, which included 1341 core genes and 2760 genes belonging to the dispensable genome. A total of 1491 Clusters of Orthologous Genes (COGs) were found to be specific to the 11 *L. garvieae* genomes, with the genome of the “separated strain” showing the highest presence of unique genes.

Citation: Ferrario C, Ricci G, Milani C, Lugli GA, Ventura M, et al. (2013) *Lactococcus garvieae*: Where Is It From? A First Approach to Explore the Evolutionary History of This Emerging Pathogen. PLoS ONE 8(12): e84796. doi:10.1371/journal.pone.0084796

Editor: Jose Luis Balcazar, Catalan Institute for Water Research (ICRA), Spain

Received: May 15, 2013; **Accepted:** November 19, 2013; **Published:** December 31, 2013

Copyright: © 2013 Ferrario et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was carried out with the financial support of the Lombardy Region, Italy, under the Project “Post genomica batterica per la qualità e la sicurezza degli alimenti.” The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: grazia.fortina@unimi.it

Introduction

Over the last decades, the development of efficient molecular methods has revolutionized the microbiological studies and improved the knowledge about the population structure within a single species. The analysis of polymorphisms in a bacterial population, normally subjected to complex processes of diversification, allows the reconstruction of the evolutionary history of a microbe. Various approaches have been developed to trace the history of several bacterial species, including pathogens or opportunistic pathogens. Multilocus sequence typing (MLST) [1] is currently the most widely employed approach to probe the population biology and to predict ancestral genotypes and patterns of descent within groups of related genotypes [2–7]. The recent developments in generating whole genome sequences in a short period of time allow to obtain further knowledge about genetic variability [8–12]. Today, with the increasing number of complete genome sequences for single bacterial species, that take into account the variability of the dispensable genome, it is possible to trace evolutionary events that have led to genetic changes and that leave a characteristic fingerprint.

Lactococcus garvieae (the elder synonym of *Enterococcus seriolicida*) is known as the causative agent of lactococcosis, a septicemic process, described for the first time at the end of the 50s in Japan, in the intensive production of rainbow trout [13]. Since then, *L.*

garvieae progressively spread in numerous countries and was identified as responsible for outbreaks of this disease in several fish species [14–16]. During the last decades, due to an improvement in molecular methodologies, this microorganism, phenotypically similar to the better known *Lactococcus lactis*, has also been isolated in other animal species, as cows and buffalos with mammary infection, in raw cow milk and in human clinical samples [17–20]. Genotypic studies were mainly carried out on fish isolates [21–23]. The data obtained allowed the differentiation of *L. garvieae* in relation to the host origin and, within the rainbow trout strains, to their geographical origin. More recently, studies carried out on dairy products obtained from raw milk, suggested another possible ecological niche of origin of *L. garvieae*. In some artisanal dairy products, the dominant microbial population was constituted by this species [24–25]. Further comparative studies carried out on different isolates, with the aim to identify a possible differential genetic marker, did not produce relevant results. Genes responsible for the utilization of lactose, initially considered specific for the dairy isolates [26], were also found in few strains coming from other sources [27]. The presence of a capsule, previously identified as the main virulence factor in the fish-borne strains, is characteristic to only a few strains isolated from diseased fish and it was not found in strains from other sources [28–31]. The absence of capsule was verified by the genome analysis of 11 strains of *L. garvieae*, which over the past two years have become

available in public databases, reflecting the increasing interest in the study of this species: five diseased fish isolates [29,32–34], one human clinical isolate [35], two dairy strains [34,36], one duck intestine isolate [37] and two meat isolates [38]. Recently, we investigated the genetic heterogeneity of a collection of *L. garvieae* strains originating not only from fish and dairy products, but also from food niches not yet studied for the presence of *L. garvieae*: meat products, vegetables and cereals [39]. This strain collection was subjected to typing studies and to a preliminary Multi Locus Restriction Typing analysis carried out on genes belonging to the core genome of the species. The obtained results revealed the presence of at least two genomic lineages within *L. garvieae* population, not entirely coherent with the ecological niche of origin of these strains.

In the present study, comparison among *L. garvieae* available complete genomes, together with multilocus sequence typing (MLST) experiments, were carried out with the aim to better understand the evolutionary history and the genomic complexity of this emerging zoonotic pathogen.

Results and Discussion

Multilocus Sequence Typing (MLST)

Nineteen *Lactococcus garvieae* strains were selected from a larger strain collection previously explored through different genotyping methods [39] and chosen as representative of the isolation niche and of the different individuated genomic lineages (Table 1). They were subjected to a MLST that targeted seven unlinked housekeeping genes, 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 tested strains identified 18 different Sequence Types (STs), highlighting a significant heterogeneity in this strain collection. All loci were polymorphic (Table 1). The number of alleles varied between eight in *gapC*, the most conserved locus, and 14 in *rpoC*, suggesting a different evolution rate for different loci, equally distributed along the genome sequences (the minimum distance among the loci was 18 kb).

Table 1. *L. garvieae* strains analyzed and allelic profiles.

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

ST: Sequence Type.

/: Unknown.

doi:10.1371/journal.pone.0084796.t001

Table 2. Polymorphism observed in seven housekeeping genes in *L. garvieae*.

Locus	Size (bp)	No. alleles	No. Polymorphic sites	π	π_{MAX}	Tajima's D ^a	Fu & Li's D ^a	Fu & Li's F ^a	I _A ⁵
<i>als</i>	811	11	110	0.0396±0.0086	0.107	-1.068	-1.288	-1.407	
<i>atpA</i>	803	9	92	0.0358±0.0117	0.093	-1.133	-0.984	-1.147	
<i>tuf</i>	809	12	37	0.0122±0.0022	0.028	-1.163	-1.521	-1.626	
<i>gapC</i>	821	8	23	0.0074±0.0032	0.022	-1.6321 ^b	-1.7136 ^b	-1.886 ^b	
<i>gyrB</i>	827	11	164	0.0663±0.0159	0.151	-0.703	-0.887	-0.954	
<i>rpoC</i>	830	14	79	0.0271±0.006	0.072	-0.701	-1.067	-1.111	
<i>galP</i>	812	13	186	0.0616±0.0188	0.183	-1.413	-1.736	-1.889	
Conc	5713	18	691	0.0297±0.0068	0.091	/	/	/	0.127 ^c
Conc S _A	5713	8	87	0.0063±0.0007	0.009	/	/	/	0.162 ^c
Conc S _B	5713	8	110	0.0071±0.0008	0.011	/	/	/	0.549 ^c

Conc: concatenated sequences of seven loci.

S_A = Subgroup A, S_B = Subgroup B.

^aStatistical significance: Not significant, p>0.10.

^bStatistical significance: Not significant, 0.10<p<0.05.

^clinkage disequilibrium detected.

/: not determined.

π = defined as the average number of nucleotide differences per site for a group of DNA sequences sampled.

π_{MAX} = maximal nucleotide diversity, defined as the number of nucleotide differences per site between the two most divergent sequences within the population.

doi:10.1371/journal.pone.0084796.t002

The analysis of allelic profiles highlighted a first relationship among strains. Through the eBURST algorithm that defines Clonal Complexes (CCs) by single-locus variants, we identified three main CCs, in which 50% of all the strains were distributed (Table 1). CC1 included seven strains grouped in ST3, ST4, and ST13 sequencing types. CC2 grouped ST16 and ST17, with representative strains ATCC49156 and LG2 respectively. CC3 included four strains belonging to ST10 and ST11. Therefore, the CCs were not homogeneous with reference to the niche of isolation. The remaining 13 strains represented 11 unique STs, indicating a high genotype frequency.

In order to extend the analysis of the genetic diversity of *L. garvieae*, we calculated the average nucleotide diversity π , considering only one sample from each ST. We also measured the π_{MAX} , defined as the number of nucleotide differences per site between the two most divergent sequences within the population. This value in fact is not directly correlated to sampling size but only to the extreme values of sequence divergence [40]. The average nucleotide diversity π of *L. garvieae* generated by the analysis of the concatenated DNA sequences of all loci was 0.0297±0.0068, corresponding to 691 polymorphic sites (Table 2). This π value was significantly higher than π for similar species, like *L. lactis* (π 0.0082±0.0010) [40] that appears monophyletic, suggesting the presence of different genetic lineages. For single loci, π ranges from 0.0074±0.0032 for *gapC* to 0.0663±0.0159 for *gyrB*, and these results were also confirmed by the determination of π_{MAX} , supporting the hypothesis of different evolution rate of the considered loci.

The phylogeny of the 26 *L. garvieae* strains was analyzed by constructing a neighbor-joining tree from the 5713 bp concatenated sequence of the seven loci (Figure 1). The tree revealed the presence of two main subgroups, as highlighted in our previous work [39]. Subgroup S_A consisted of strains included in CC1 and CC2 and three strains with the unique STs. Subgroup S_B included strains of CC3 and eight strains with six different STs. Moreover, in this analysis we found that strain I113 and, particularly, strain DCC43 were the most different among all studied isolates, and clustered in independent branches. Strain DCC43 also showed the

highest proportion of unique Single Nucleotide Polymorphisms (SNPs) (data not shown). The phylogenetic tree was compared to the topologies of the seven trees constructed for each locus (data not shown). The trees obtained from the analysis of each locus were very similar to the one obtained from the analysis of the concatenated sequence of all loci.

After sequence alignment within the subgroups, the number of polymorphisms and genetic diversity within each subpopulation were reduced (Table 2). This suggests low genetic exchange between these *L. garvieae* subgroups. Moreover, the presence of strains I113 and DCC43, which were not included into any subgroup, significantly influenced the mean genetic diversity of the total population. 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 originated (Figure 2).

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 probably indicate distinct inclinations and adaptive abilities to environments of the two subgroups: S_B seems to respond to selective pressure increasing the recombination rate. It is interesting to note that the recombination events in S_B did not seem to contribute to nucleotide diversity (π for S_A and S_B are similar): recombination among members of the same subgroup did not introduce significative polymorphisms that affect nucleotide diversity. Recombination events in *L. garvieae* population were also investigated using the SplitsTree program, with the split decomposition methods on the concatenated sequence of the total population and for subgroups (Figure 3). 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

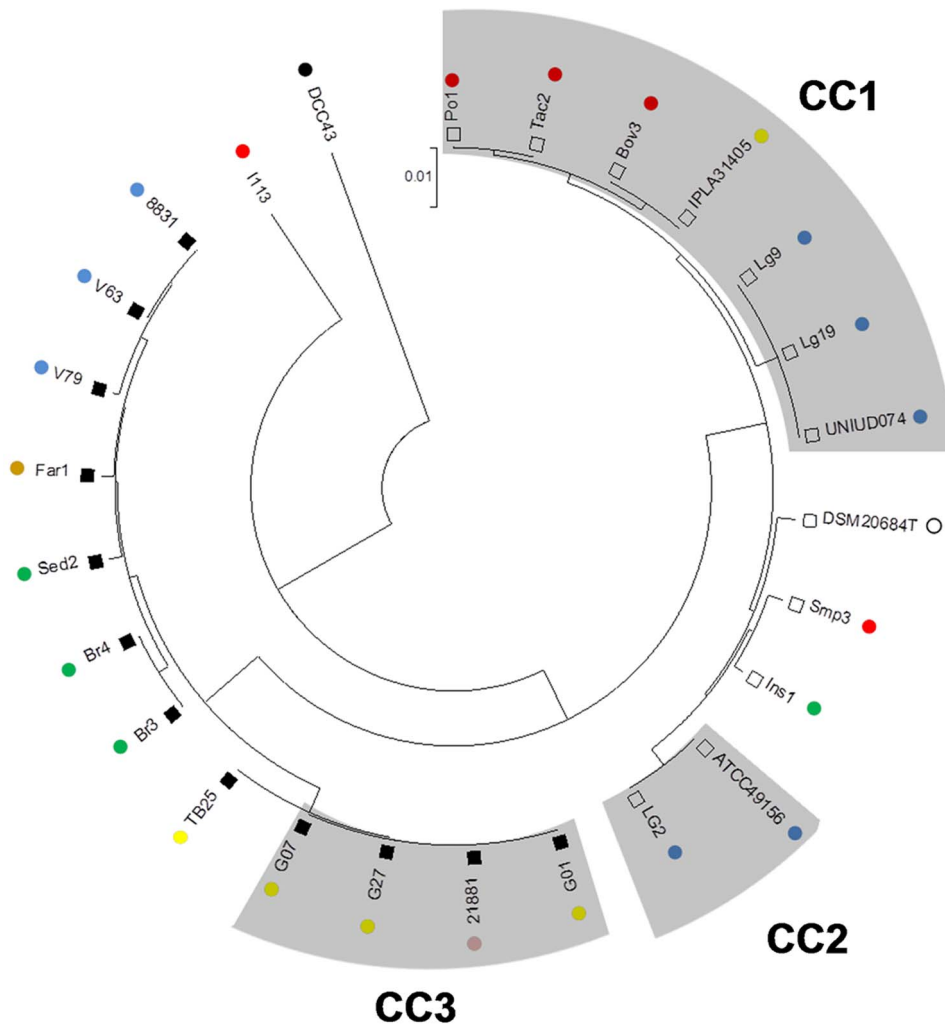


Figure 1. Phylogenetic relationships between *L. garvieae* strains. The unrooted neighbor-joining tree (bootstrap 1000, Kimura 2-parameter model) was constructed from the 5713 bp concatenated DNA sequences of the seven loci (*als*, *atpA*, *tuf*, *gapC*, *gyrB*, *rpoC* and *galP*) of *L. garvieae*. Open and closed squares correspond to subgroups S_B and S_A , 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. doi:10.1371/journal.pone.0084796.g001

also performed using phylogenetically related *L. lactis* subsp. *lactis* IL1403 (accession number AE005176) and *L. lactis* subsp. *cremoris* MG1363 (AM406671) [29]. The split graph showed the same subdivision of *L. garvieae* population, with the strain DCC43 did not belong to none of these groups. Comparison of the 16S rRNA gene showed seven SNPs in other variable regions: two of them, in position 91 and 472, common to *L. lactis*. Thus, even if the strains are closely related in respect to the 16S rRNA gene sequence homology, they are not clustered together (Figure S2), reflecting the subdivision obtained analyzing the other genes.

Tajima's D, Fu & Li's D and F tests of neutrality were used to identify the evolution model of each target gene. All three tests gave values that did not significantly deviate from 0 ($p > 0.10$; for *gapC* locus, $0.10 < p < 0.05$; Table 2), indicating that the seven loci evolved by random genetic drift. The intergenic recombination was calculated by estimating the linkage disequilibrium between loci, using the standardized index of association statistic, I_A^S . Only one sample from every ST was analyzed, to avoid introduction of linkage disequilibrium by sampling bias. Significant linkage disequilibrium was detected considering either the 18 STs of the collection (see Table 2), or the two subgroups S_A and S_B . I_A^S was not significantly different from 0, even if subgroup S_B showed a higher value, suggesting that the recombination in this cluster has experienced a recent expansion of the population size.

The sequence comparison of 16S rRNA gene, the slowest evolving molecule among housekeeping genes (Figure S2), showed a SNP in the position 203 (V2 region of *Escherichia coli*) [41], distinguishing members of the two subgroups. The strain DCC43 did not belong to none of these groups. Comparison of the 16S rRNA gene showed seven SNPs in other variable regions: two of them, in position 91 and 472, common to *L. lactis*. Thus, even if the strains are closely related in respect to the 16S rRNA gene sequence homology, they are not clustered together (Figure S2), reflecting the subdivision obtained analyzing the other genes.

Genome Comparison

General features of the *Lactococcus garvieae* genomes are reported in Table 3, 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 [36,42]. Overall, the genomic variations in size and the number of the protein-coding genes were $<15\%$ and $<21\%$ between any two strains, respectively. In comparison to

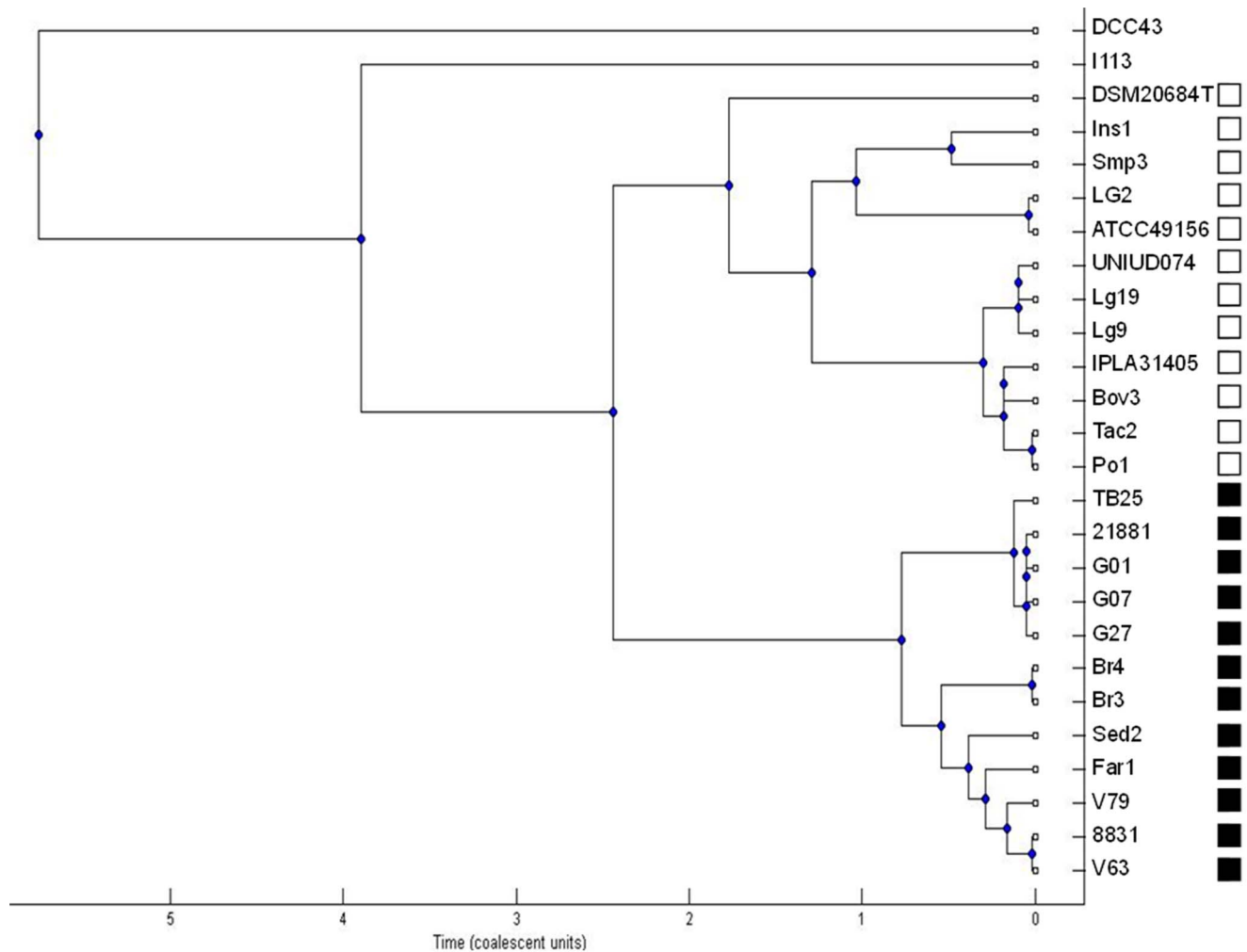


Figure 2. Major rule consensus tree based on Clonal Frame analysis of concatenated sequences of all loci, for the total population. The X-axis represents the estimated time to the most recent common ancestor of *L. garvieae*. Open and closed squares correspond to subgroups S_B and S_A , respectively.
doi:10.1371/journal.pone.0084796.g002

other *Lactococcus* species, *L. garvieae* possesses a smaller genome and a smaller number of protein-coding genes. A higher GC content, ranging from 37.70 to 38.80%, was also observed in *L. garvieae*.

To estimate the number of genes present in each *L. garvieae* strain, a pan-genome profile (a full complement of genes in a species [43–44]) and a core genome profile (the orthologous genes, OGs, that are conserved in all strains of the species) were built using all possible BLAST combinations for each sequentially added genome. We identified a total of about 4100 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. Figure 4 displays the decrease of the core genome as more genome sequences are added. It 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.

The protein coding sequences of the core were used to construct a phylogenetic tree (Figure S3), which displays the evolutionary development of the *L. garvieae* strains. The tree branching is highly similar to MLST tree generated from the seven housekeeping genes used previously, which highlighted the presence of two

clusters encompassing the majority of the *L. garvieae* strains, consisting in subgroups S_A and S_B . Remarkably, *L. garvieae* DCC43 showed a significant genetic distance from the two main subgroups.

In addition, we constructed the evolutionary relatedness between all the *L. garvieae* strains, using a matrix based on the presence/absence of OGs (Figure S4). Although this phylogenomic tree based on the matrix of the total gene presence/absence is different from the phylogenetic tree based on the core genes, the clustering of the strains largely reflects their phylogenetic relatedness.

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

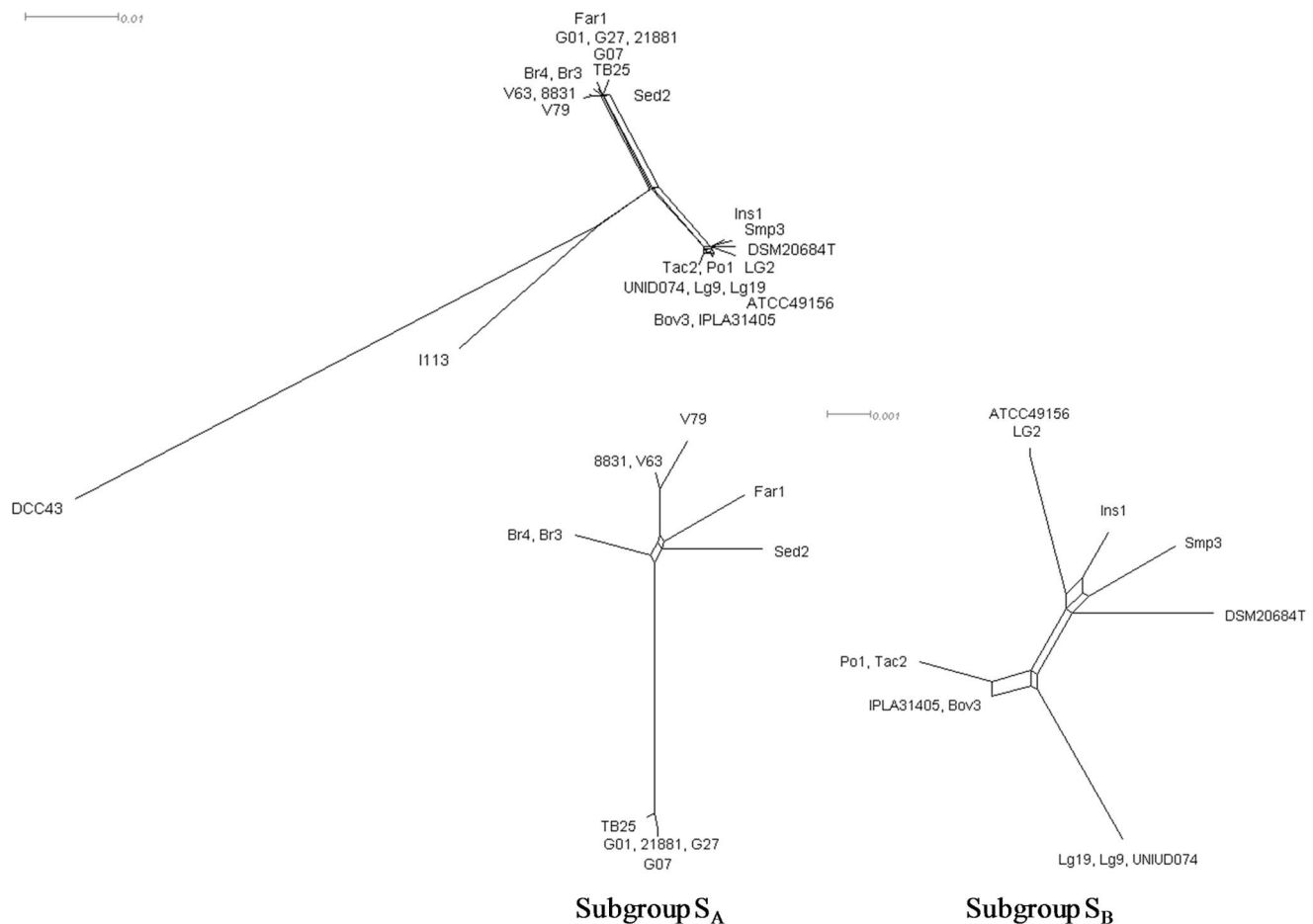


Figure 3. Splits decomposition analysis of *L. garvieae* population and subgroups. Parallelograms identify interconnected network of phylogenetic relationships between strains.
doi:10.1371/journal.pone.0084796.g003

the class of hypothetical proteins and proteins with unknown function [43].

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 3). 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 (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).

Conclusions

The lack of knowledge about ecological and functional role of *Lactococcus garvieae* in niches other than fish sector, makes this emerging pathogen attractive to examine in its evolutionary history and in its global complexity. Thus, selected *L. garvieae* strains of our collection coming from various food sources, as well as seven *L. garvieae* genomes of clinical and animal isolates available in databases, were characterized through MLST and whole-genome comparison analyses.

The MLST identified two branches containing the majority of the strains, grouped into two main subgroups, and other two bearing the single strain each. The obtained phylogenetic tree including strains of *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris*, indicates that *L. garvieae* “separated strains” (I113 and DCC43) appeared first and may be considered the evolutionary missing link between *L. lactis* and *L. garvieae*. It is plausible to assume that the strains belonging to the main subgroups could have emerged more recently. Our study also provides a first insight in the core and pan-genome of *L. garvieae*. The core genome consists of 1341 OGs, the dispensable gene pool is estimated to be about 2760 OGs. This accessory genome represents a large proportion of the total genes present within the *L. garvieae* genome, and could suggest the cosmopolitan lifestyle of *L. garvieae* species. Moreover, many genes were found to be specific to the 11 *L. garvieae* genomes, with DCC43 genome showing the higher portion of unique genes.

In accordance to the genetic phylogeny, the comparison of 11 complete genomes of *L. garvieae* highlighted the majority of *L. garvieae* strains to belong to two major subgroups. The obtained consensus tree also suggests the strain DCC43 as the most ancestral lineage of the *L. garvieae* species, when rooted with *L. lactis* sequences. As proposed by Gabrielsen et al. [37], this evolutionary intermediate could represent a novel *L. garvieae* sub-species.

Table 3. 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
<i>L. garvieae</i>	8831	2,085,932	38.0	draft	2016	919	88.9	11.1	NZ_AFCD000000000
	21881	2,164,301	37.9	draft	2216	858	87.9	12.1	NZ_AFCC000000000
	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_AMQS000000000
	I113	2,178,733	37.9	draft	2159	893	88.5	11.5	NZ_AMFD000000000
	IPLA 31405	2,052,312	38.5	draft	1778	939	81.4	18.6	NZ_AKFO000000000
	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_AGQY000000000
	Tac2	2,242,863	38.2	draft	2210	896	88.3	11.7	NZ_AMFE000000000
	TB25	2,014,328	38.1	draft	2024	879	88.3	11.7	NZ_AGQX000000000
	UNIUD074	2,171,472	38.7	draft	2146	871	86	14	NZ_AFHF000000000
<i>L. lactis</i> subsp. <i>cremoris</i>	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
<i>L. lactis</i> subsp. <i>lactis</i>	II1403	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
<i>L. raffinolactis</i>	4877	2,280,761	38.6	draft	2359	803	83.1	16.9	CALL010000000

doi:10.1371/journal.pone.0084796.t003

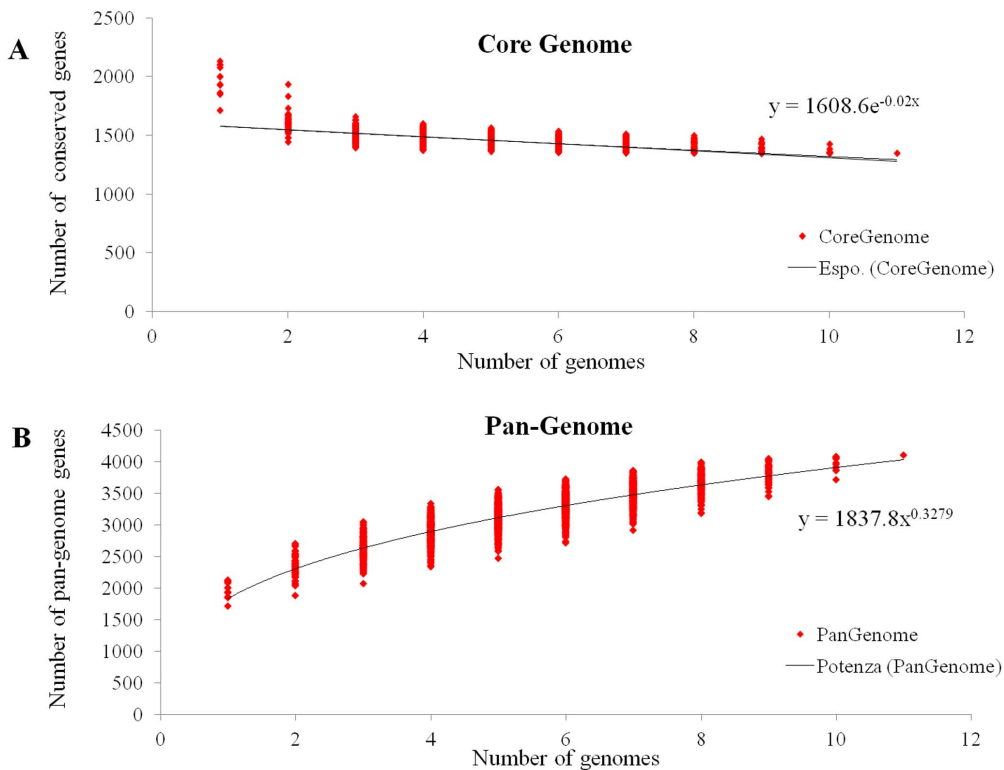


Figure 4. Pan-genome prediction. The distribution of the number of core COGs (A) and total pan-genome COGs (B) found upon sequential addition of *n* genomes. In panel A, an exponential regression to core genome data is shown as a solid curve. In panel B, power law fit to the pan-genome size is shown as solid curve.
doi:10.1371/journal.pone.0084796.g004

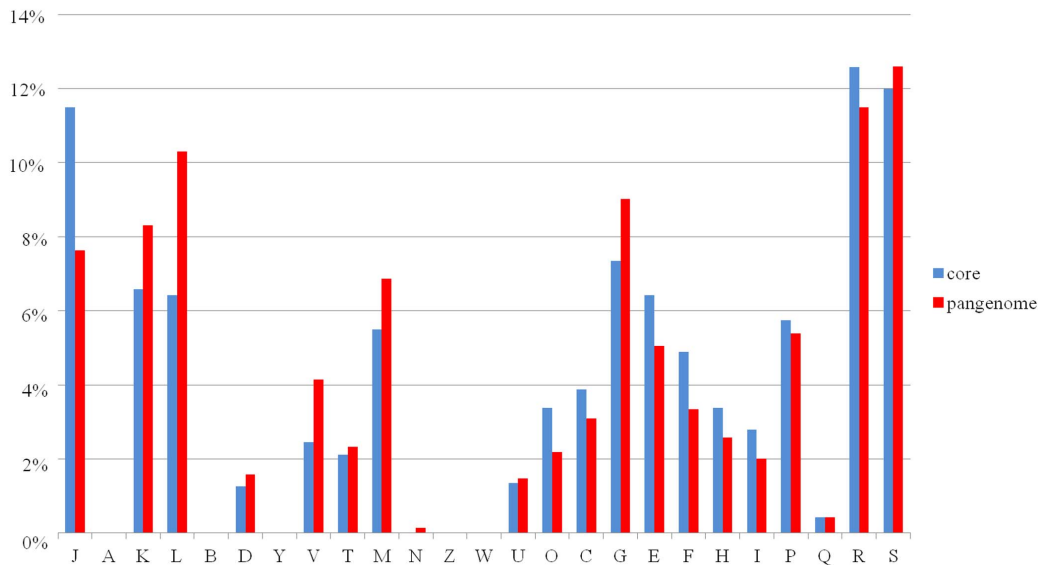


Figure 5. COG families of *L. garvieae*. Bar chart showing a representation of COG families annotation of core COGs and whole pan-genome COGs. Each COG family is identified by a one-letter abbreviation: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and 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, post-translational 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.
doi:10.1371/journal.pone.0084796.g005

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_AGQX000000000; Lg9 - NZ_AGQY000000000; I113 - NZ_AMFD000000000; Tac2 - NZ_AMFE000000000) [34,38]. 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 and 16S rRNA Sequencing

DNA was extracted as previously described [39], 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). 16S rRNA amplifications were performed as previously reported [24]. Nucleo Spin Extract II (Macherey-Nagel GmbH, Düren, Germany) was used to purify PCR products that were sequenced using the dideoxy chain-termination principle [45], employing ABI Prism Big Dye Terminator Kit (Applied Biosystems, Foster City, CA). The reaction products were analyzed with the ABI PrismTM310 DNA Sequencer. The database searches were performed by using the basic local alignment tool (BLAST) programs [46] from the National Center for Biotechnology Information website. The phylogenetic tree was constructed using the UPGMA method [47].

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 Table S1, with relevant references. Amplicons were gel purified, sequenced and analyzed 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_AFCD000000000; 21881 - NZ_AFCE000000000; ATCC 49156 - NC_015930; LG2 - NC_017490; UNIUD074 - NZ_AFHF000000000; IPLA 31405 - NZ_AKFO000000000; DCC43 - NZ_AMQS000000000). Selecting the most polymorphic regions of 800–850 bp, these were analyzed using MEGA v5 [48]. Isolate dataset creation and allele assignment 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. The number of segregating or polymorphic site (S), nucleotide diversity (π), Tajima's D, Fu & Li's D and F were calculated using DnaSP v5.10 [49]. π_{MAX} values were extracted from the squared similarity matrix calculated with DNADIST program (D option set to "similarity table") in the PHYLIP v.3.69 package [50]. For phylogenetic analysis, concatenated sequences were aligned and analyzed with MEGA v5. Genetic distances were computed by the Kimura two-parameter model, and the phylogenetic tree was constructed using the neighbor-joining method. Strains relationships were analyzed using eBURST [51] to identify potential Clonal Complexes (CCs), with the default stringent (conservative) definition. To investigate the population structure, the Clonal

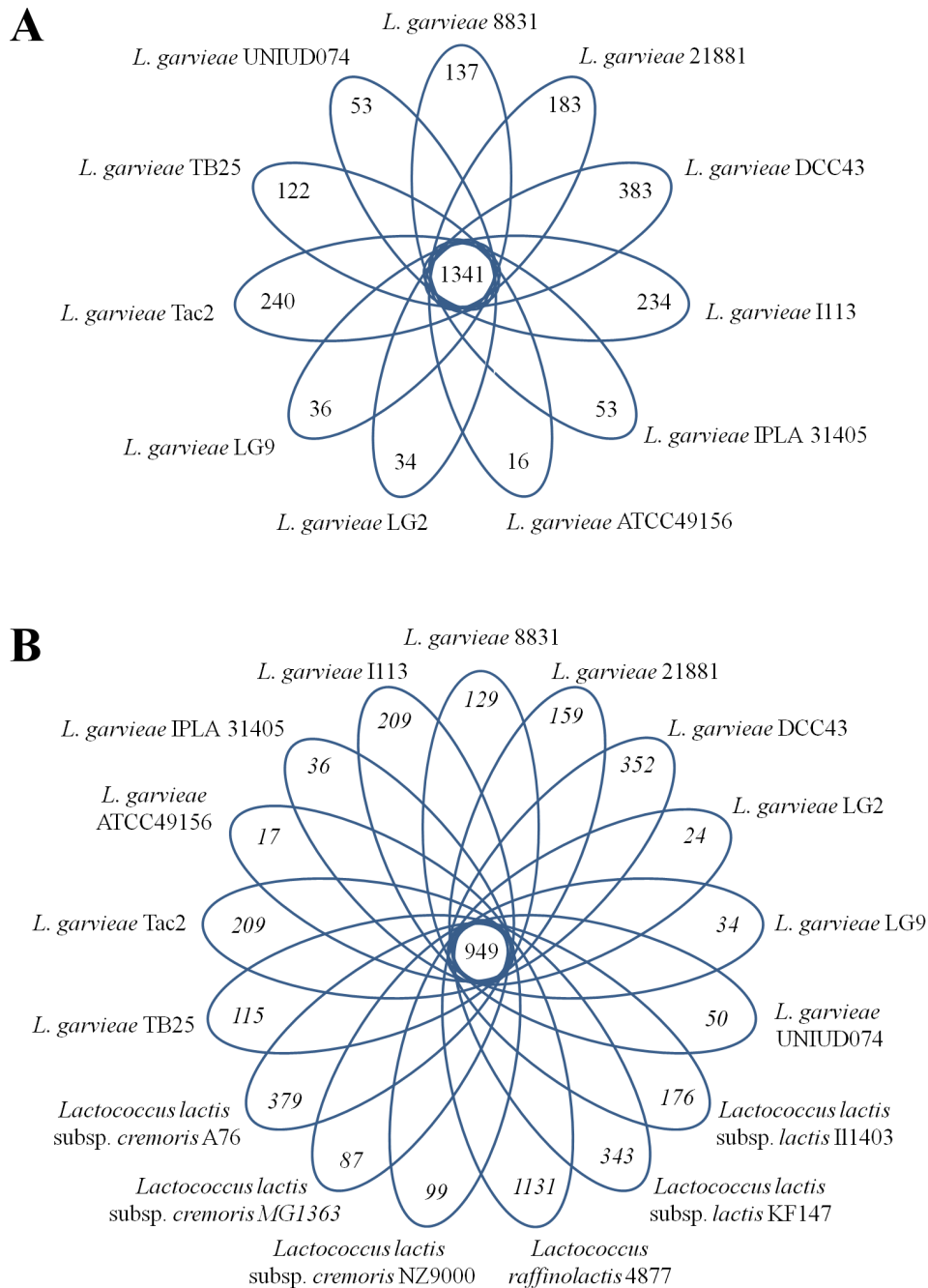


Figure 6. Genomic diversity of the *Lactococcus* species. Venn diagram of core COGs (Clusters of Orthologous Genes) shared between all the strain analyzed and COGs unique to each single strain.
doi:10.1371/journal.pone.0084796.g006

Frame method was used [52]. The recombination to mutation ratio (r/m) was calculated as reported by Vos and Didelot [52]. 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 1000 bootstraps replicates based on parsimony splits as implemented in SplitsTree v4.1 [53]. The standardized Index of Association (I_A^S) was calculated with LIAN 3.5 (<http://guanine.evolbio.mpg.de/cgi-bin/lian/lian.cgi.pl>) [54], using a Monte Carlo randomization test with 1000 resamplings.

Genome Analysis and Comparison

Each predicted proteome of the analyzed strains (Table 3), was searched for orthologues against the total proteome, where orthology between two proteins was defined as the best bidirectional FASTA hits [55]. 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 [46] (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) [56].

Following this, the unique COGs were classified by selecting the clusters with members from only one of the *Lactococcus* genomes analyzed. 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 [57], and phylogenetic trees were constructed using the Neighbor joining in PhyML [58]. 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 [59]; 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/>).

Supporting Information

Figure S1 Splits decomposition analysis of lactococcal strains. 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, c) detail of interconnection among between DCC43 and *L. lactis*.

(TIF)

Figure S2 Multiple alignment of polymorphic sites of the *L. garvieae* 16S rRNA gene sequences. SNPs were reported according to *Escherichia coli* numbering of variable regions (V1–V6) of 16S rRNA gene (Baker et al. 2003). In the UPGMA tree, stratification in subgroup is reported.

(TIF)

Figure S3 Genome phylogeny of *L. garvieae*. Phylogenetic supertree based on the aligned sequences of core proteins shared by all the analyzed *Lactococcus* genomes.

(TIF)

Figure S4 Pan-genome phylogenomic tree built on the presence/absence information of each gene of the pan-genome in each *Lactococcus* genome.

(TIF)

Table S1 Primers used for MLST study.

(DOC)

Acknowledgments

We thank Dr Milda Stuknyte for a critical reading of the manuscript and for her useful suggestions.

Author Contributions

Conceived and designed the experiments: CF MGF. Performed the experiments: CF GR GE FB CM GAL. Analyzed the data: CF GR MGF CM GAL MV. Wrote the paper: CF MGF MV.

References

- Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95: 3140–3145.
- Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, et al. (2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc Natl Acad Sci U S A* 99: 7687–7692.
- Didelot X, Falush D (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* 175: 1251–1266.
- McQuiston JR, Herrera-Leon S, Wertheim BC, Doyle J, Fields PI, et al. (2008) Molecular phylogeny of the salmonellae: relationships among *Salmonella* species and subspecies determined from four housekeeping genes and evidence of lateral gene transfer events. *J Bacteriol* 190: 7060–7067.
- Chen PE, Cook C, Stewart AC, Nagarajan N, Sommer DD, et al. (2010) Genomic characterization of the *Yersinia* genus. *Genome Biol* 11: R1.
- Stabler RA, Dawson LF, Valiente E, Cairns MD, Martin MJ, et al. (2012) Macro and micro diversity of *Clostridium difficile* isolates from diverse sources and geographical locations. *PLoS One* 7: e31559.
- Pérez-Losada M, Cabezas P, Castro-Nallar E, Crandall KA (2013) Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. *Infect Genet Evol* 16C: 38–53.
- Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, et al. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci U S A* 102: 13950–13955.
- Brochet M, Rusniok C, Couvé E, Dramsi S, Poyart C, et al. (2008) Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* 105: 15961–15966.
- Sheppard SK, McCarthy ND, Falush D, Maiden MC (2008) Convergence of *Campylobacter* species: implications for bacterial evolution. *Science* 320: 237–239.
- Kitichotirat W, Bumgarner RE, Asikainen S, Chen C (2011) Identification of the pan-genome and its components in 14 distinct *Aggregatibacter actinomycetemcomitans* strains by comparative genomic analysis. *PLoS One* 6: e22420.
- Donkor ES, Stabler RA, Hinds J, Adegbola RA, Antonio M, et al. (2012) Comparative phylogenomics of *Streptococcus pneumoniae* isolated from invasive disease and nasopharyngeal carriage from West Africans. *BMC Genomics* 13: 569.
- Vendrell D, Balcázar JL, Ruiz-Zarzuela I, de Blas I, Gironés O, et al. (2006) *Lactococcus garvieae* in fish: a review. *Comp Immunol Microbiol Infect Dis* 29: 177–198.
- Chen SC, Lin YD, Liaw LL, Wang PC (2001) *Lactococcus garvieae* infection in the giant freshwater prawn *Macrobrachium rosenbergii* confirmed by polymerase chain reaction and 16S rDNA sequencing. *Dis Aquat Organ* 45: 45–52.
- Baeck GW, Kim JH, Gomez DK, Park SC (2006) Isolation and characterization of *Streptococcus* sp. from diseased flounder (*Paralichthys olivaceus*) in Jeju Island. *J Vet Sci* 7: 53–58.
- Evans JJ, Klesius PH, Shoemaker CA (2009) First isolation and characterization of *Lactococcus garvieae* from Brazilian Nile tilapia, *Oreochromis niloticus* (L.), and pintado, *Pseudoplatystoma corruscans* (Spix & Agassiz). *J Fish Dis* 32: 943–951.
- Devriese LA, Hommez J, Laevens H, Pot B, Vandamme P, et al. (1999) Identification of aesculin-hydrolyzing streptococci, lactococci, aerococci and enterococci from subclinical intramammary infections in dairy cows. *Vet Microbiol* 70: 87–94.
- Fihnan V, Raskine L, Barrou Z, Kiffel C, Riahi J, et al. (2006) *Lactococcus garvieae* endocarditis: identification by 16S rRNA and *sodA* sequence analysis. *J Infect* 52: 3–6.
- Aubin GG, Bémer P, Guillouzoic A, Crémét L, Touchais S, et al. (2011) First report of a hip prosthetic and joint infection caused by *Lactococcus garvieae* in a woman fishmonger. *J Clin Microbiol* 49: 2074–2076.
- Russo G, Iannetta M, D’Abramo A, Mascellino MT, Pantosti A, et al. (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.
- Eldar A, Gorla M, Ghittino C, Zlotkin A, Bercovier H (1999) Biodiversity of *Lactococcus garvieae* strains isolated from fish in Europe, Asia, and Australia. *Appl Environ Microbiol* 65: 1005–1008.
- Ravelo C, Magariños B, López-Romalde S, Toranzo AE, Romalde JL (2003) Molecular fingerprinting of fish-pathogenic *Lactococcus garvieae* strains by random amplified polymorphic DNA analysis. *J Clin Microbiol* 41: 751–756.
- Schmidtke LM, Carson J (2003) *Lactococcus garvieae* strains isolated from rainbow trout and yellowtail in Australia, South Africa and Japan differentiated by repetitive sequence markers. *Bull Eur Ass Fish Pathol* 23: 206–212.
- Fortina MG, Ricci G, Acquati A, Zeppa G, Gandini L, et al. (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.

25. Fernández E, Alegria A, Delgado S, Mayo B (2010) Phenotypic, genetic and technological characterization of *Lactococcus garvieae* strains isolated from a raw milk cheese. *Int Dairy J* 20: 142–148.
26. Fortina MG, Ricci G, 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.
27. Aguado-Urda M, Cutuli MT, Blanco MM, Aspiroz C, Tejedor JL, et al. (2010) Utilization of lactose and presence of the phospho- β -galactosidase (*lacG*) gene in *Lactococcus garvieae* isolates from different sources. *Int Microbiol* 13: 189–193.
28. Barnes AC, Guyot C, Hanse BG, Mackenzie K, Horn MT, et al. (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.
29. Morita H, Toh H, Oshima K, Yoshizaki M, Kawanishi M, et al. (2011) Complete genome sequence and comparative analysis of the fish pathogen *Lactococcus garvieae*. *PLoS One* 6: e23184.
30. Reimundo P, Rivas AJ, Osorio CR, Méndez J, Pérez-Pascual D, et al. (2011) Application of suppressive subtractive hybridization to the identification of genetic differences between two *Lactococcus garvieae* strains showing distinct differences in virulence for Rainbow trout and mouse. *Microbiology* 157: 2106–2119.
31. Miyauchi E, Toh H, Nakano A, Tanabe S, Morita H (2012) Comparative genomic analysis of *Lactococcus garvieae* strains isolated from different sources reveals candidate virulence genes. *Int J Microbiol* 2012: 728276.
32. Aguado-Urda M, López-Campos GH, Gibello A, Cutuli MT, López-Alonso V, et al. (2011) Genome sequence of *Lactococcus garvieae* 8831, isolated from rainbow trout lactococcosis outbreaks in Spain. *J Bacteriol* 193: 4263–4264.
33. Reimundo P, Pignatelli M, Alcaraz LD, D'Auria G, Moya A, et al. (2011) Genome sequence of *Lactococcus garvieae* UNIUD074, isolated in Italy from a lactococcosis outbreak. *J Bacteriol* 193: 3684–3685.
34. Ricci G, Ferrario C, Borgo F, Rollando A, Fortina MG (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.
35. Aguado-Urda M, López-Campos GH, Blanco MM, Fernández-Garayzábal JF, Cutuli MT, et al. (2011) Genome sequence of *Lactococcus garvieae* 21881, isolated from a case of human septicaemia. *J Bacteriol* 193: 4033–4034.
36. Flórez AB, Reimundo P, Delgado S, Fernández E, Alegria A, et al. (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.
37. Gabrielsen C, Brede DA, Hernández PE, Nes IF, Diep DB (2012) Genome sequence of the bacteriocin-producing strain *Lactococcus garvieae* DCC43. *J Bacteriol* 194: 6976–6977.
38. Ricci G, Ferrario C, Borgo F, Eraclio G, Fortina MG (2013) Genome sequences of two *Lactococcus garvieae* strains isolated from meat. *Genome Announc* 1: e00018-12.
39. Ferrario C, Ricci G, Borgo F, Rollando A, Fortina MG (2012) Genetic investigation within *Lactococcus garvieae* revealed two genomic lineages. *FEMS Microbiol Lett* 332: 153–161.
40. Passerini D, Beltramo C, Coddeville M, Quentin Y, Ritzenthaler P, et al. (2010) Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. *PLoS One* 5: e15306.
41. Baker GC, Smith JJ, Cowan DA (2006) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55: 541–555.
42. Aguado-Urda M, Gibello A, Blanco MM, López-Campos GH, Cutuli MT, et al. (2012) Characterization of plasmids in a human clinical strain of *Lactococcus garvieae*. *PLoS One* 7: e40119.
43. Mira A, Martín-Cuadrado AB, D'Auria G, Rodríguez-Valera F (2010) The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol* 13: 45–57.
44. Milani C, Duranti S, Lugli GA, Bottacini F, Strati F, et al. (2013) Comparative genomics of *Bifidobacterium animalis* subsp. *lactis* reveals a strict monophyletic bifidobacterial taxon. *Appl Environ Microbiol* 79: 4304–4315.
45. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463–5467.
46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
47. Sneath PHA, Sokal RR (1973) *Numerical Taxonomy*. Freeman, San Francisco.
48. Takamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
49. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
50. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164–166.
51. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 186: 1518–1530.
52. Vos M, Didelot X (2009) A comparison of homologous recombination rates in bacteria and archaea. *ISME J* 3: 199–208.
53. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23: 254–267.
54. Haubold H, Hudson RR (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* 16: 847–848.
55. Pearson WR (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol* 132: 185–219.
56. van Dongen S (2000) Graph clustering by flow simulation. PhD thesis. University of Utrecht, Utrecht, The Netherlands.
57. Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7: 539.
58. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
59. Zhao Y, Wu J, Yang J, Sun S, Xiao J, et al. (2012) PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28: 416–418.