BMC Genomics

RESEARCH ARTICLE Open Access

LAB-Secretome: a genome-scale comparative analysis of the predicted extracellular and surface-associated proteins of Lactic Acid Bacteria

Miaomiao Zhou[1,3*], Daniel Theunissen[2,3], Michiel Wels[1,2,3], Roland J Siezen[1,2,3]

Abstract

Background: In Lactic Acid Bacteria (LAB), the extracellular and surface-associated proteins can be involved in processes such as cell wall metabolism, degradation and uptake of nutrients, communication and binding to substrates or hosts. A genome-scale comparative study of these proteins (secretomes) can provide vast information towards the understanding of the molecular evolution, diversity, function and adaptation of LAB to their specific environmental niches.

Results: We have performed an extensive prediction and comparison of the secretomes from 26 sequenced LAB genomes. A new approach to detect homolog clusters of secretome proteins (LaCOGs) was designed by integrating protein subcellular location prediction and homology clustering methods. The initial clusters were further adjusted semi-manually based on multiple sequence alignments, domain compositions, pseudogene analysis and biological function of the proteins. Ubiquitous protein families were identified, as well as species-specific, strain-specific, and niche-specific LaCOGs. Comparative analysis of protein subfamilies has shown that the distribution and functional specificity of LaCOGs could be used to explain many niche-specific phenotypes.
A comprehensive and user-friendly database LAB-Secretome was constructed to store, visualize and update the extracellular proteins and LaCOGs http://www.cmbi.ru.nl/lab_secretome/. This database will be updated regularly when new bacterial genomes become available.

Conclusions: The LAB-Secretome database could be used to understand the evolution and adaptation of lactic acid bacteria to their environmental niches, to improve protein functional annotation and to serve as basis for targeted experimental studies.

Background

Lactic Acid Bacteria (LAB) have been used for centuries in industrial and artisanal food and feed fermentations as starter cultures and are important bacteria linked to the human gastro-intestinal (GI) tract [1-8]. Phylogenetically they form a relatively compact group of mainly Gram-positive, anaerobic, non-sporulating, low G+C content acid-tolerant bacteria [9-12]. The genera that comprise the LAB belong to the order Lactobacillales, and are primarily *Lactobacillus, Pediococcus, Lactococcus*, *Streptococcus* and *Leuconostoc*, while some peripheral genera are *Enterococcus, Oenococcus, Aerococcus*, and *Carnobacterium*. Interestingly, even within such a compact group, vastly divergent phenotypes have been reported, providing indications of high flexibility and adaptation of these species to their living environments [13-16].

Extracellular and surface-associated proteins play a most important role in many essential interactions and adaptations of LAB to their environment [17-26]. By definition these proteins are either exposed on (anchored to membrane GO:0046658, intrinsic to external side of plasma membrane GO:0031233 and the cell wall, GO:0005618) or released (extracellular milieu, GO:0005576) from the bacterial cell surface. On a genome scale these proteins form a subset of the proteome which contains both the exoproteome [27] and part of the surface proteome [28], but excluding the integral membrane

* Correspondence: m.zhou@cmbi.ru.nl
[1]Centre for Molecular and Biomolecular Informatics, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands
Full list of author information is available at the end of the article

proteins (GO: 0005887) and the proteins that are intrinsic to internal side of plasma membrane (GO:0031235). This subset of the proteome belongs to what Desvaux *et al* have defined as "secretome" [27] and is known to mainly be involved processes such as: (1) recognition, binding, degradation and uptake of extracellular complex nutrients, (2) signal transduction, (3) communication with the environment and (4) attachment of the bacterial cell to specific sites or surfaces, e.g. to intestinal mucosa cells of the host [29-37]. Hence, genome-scale comparative analysis of these secretome (surface-associated and released from the cell) proteins may provide an understanding of the molecular function, evolution, and diversity of different LAB species and their adaptation to different environments.

Here we report a comparison of the predicted secretomes of 26 sequenced genomes of LAB representing 18 different species (Table 1). The secretome clusters of orthologous protein families (LaCOGs: Lactobacillales Cluster of Ortholog Groups) were extracted by combining homology clustering methods with protein subcellular location (SCL) prediction. The comparative analysis of LaCOGs shows many niche-specific protein families that can be used as leads for future experiments.

The complete results of this study are stored in our open-source database LAB-Secretome http://www.cmbi.ru.nl/lab_secretome with a user-friendly web-interface. An automatic update scheme was constructed to be able to add information to the database on new bacterial genomes.

Results and Discussion

Construction of the secretome protein clusters (LaCOGs)

In this study we focus on those proteins that are predicted to be wholly or largely on the outside of the cell, regardless of the translocation systems. These proteins

Table 1 The predicted LAB secretomes (genomes included in the original LaCOG analysis 43 are marked by *)

LAB species and strains	Total proteins	Secretome proteins (%)							
		A	B	C	D	E	F	G	Total (%)
E.faecalis_V583	3186	2.32	1.26	3.36	0.97	0.16	1.6	0.13	9.8
L.acidophilus_NCFM	1834	2.24	0.65	4.09	0.93	0	2.45	0.05	10.41
*L.gasseri_ATCC_33323**	1733	1.85	0.69	3.92	0.52	0.12	0.69	0	7.79
*L.johnsonii_NCC_533**	1789	2.07	0.89	4.3	0.56	0.39	0.06	0	8.27
L.delbrueckii_bulgaricus _ATCC11842	1536	1.56	0.13	3.45	1.04	0.07	2.02	0	8.27
*L.delbrueckii_bulgaricus _ATCC_BAA-365**	1681	1.43	0.06	3.15	0.95	0.18	2.08	0	7.85
*L.casei_ATCC_334**	2693	1.63	0.78	3.79	0.78	0.15	1.41	0.07	8.61
L.casei_BL23	2973	1.68	0.77	3.4	0.84	0	1.35	0.13	8.17
L.salivarius_UCC118	1973	0.91	0.25	3.4	0.61	0.15	1.27	0.1	6.69
L.sakei_23K	1845	1.52	0.33	3.36	0.76	0.05	2.06	0.27	8.35
*L.plantarum_WCFS1**	2981	1.61	1.11	3.99	0.91	0.3	0.1	0	8.02
L.brevis_ATCC_367	2178	1.29	0.55	3.35	1.52	0.14	2.53	0.09	9.47
L.fermentum_IFO_3956	1826	0.66	0.22	2.96	0.55	0	1.15	0.05	5.59
L.helveticus_DPC_4571	1597	1.38	0.13	4.51	0.44	0	2.13	0	8.59
L.reuteri_F275_JGI	1881	0.74	0.21	3.67	0.85	0	1.01	0	6.48
L.reuteri_F275_Kitasato	1803	0.78	0.28	3.55	1	0	1.22	0	6.83
L._lactis_cremoris_MG1363	2393	1.46	0.46	3.01	0.79	0	1.96	0	7.68
*L.lactis_cremoris_SK11**	2459	1.38	0.41	3.17	1.02	0.12	1.67	0.08	7.85
*L.lactis_lactis_IL1403**	2284	1.4	0.61	4.29	0.74	0.04	1.62	0.18	8.88
L.citreum_KM20	1784	0.06	0.28	4.43	1.23	1.23	0	0.06	7.29
*S.thermophilus_CNRZ1066**	1872	1.28	0.05	3.47	0.53	0.27	0.43	0.05	6.08
*S.thermophilus_LMD-9**	1669	1.5	0.24	3.89	0.54	0.18	0.84	0	7.19
S.thermophilus_LMG_18311	1854	1.29	0.11	3.78	0.54	0.49	0.65	0	6.86
*L.mesenteroides_ATCC_8293**	1966	0.1	0.31	4.93	1.12	0.31	1.22	0.15	8.14
*O.oeni_PSU-1**	1664	0.12	0.06	4.33	0.9	1.56	0	0.06	7.03
*P.pentosaceus_ATCC_25745**	1727	1.1	0.17	3.88	0.35	0.17	0.98	0.12	6.77

A: Lipid anchored; B: LPxTG Cell-wall anchored; C: N-terminally anchored (No cleavage site); D: N-terminally anchored (with cleavage site); E: Secreted via minor pathways (bacteriocin) (no cleavage site); F: Extracellular (with cleavage site); G: C-terminally anchored (with cleavage site)

The SCL prediction was made by LocateP.

form a sub-proteome of what Desvaux *et.al* defined as the "secretome" [27] by excluding the translocation systems, the integral membrane proteins, and non-protein products. Although we adapt this term "secretome" to describe our protein subset of interest, we must specify that in our analysis the term "secretome" refers to only the proteins that are released from the cells to the extracellular milieu (also called exoproteome), and the proteins that remain cell-surface associated, but nothing else.

Ideally, a comparative secretome analysis should be performed on the experimentally validated sub-proteomes or on *in silico* predicted secretome proteins with the highest possible accuracies. However, it is well-known that wet-lab proteomic studies are extremely costly and can lead to many false predictions of subcellular location, while all the currently available *in silico* protein SCL predictors have only 80%-93% prediction accuracy [38-41]. Therefore, instead of clustering predicted extracellular proteins directly, we designed an alternative process which firstly groups all proteins in the sequenced LAB genomes into ortholog groups (LaCOGs) and afterwards extracts the secretome groups by using genome-scale SCL predictions (Figure 1). In this way, the wrongly predicted secretome proteins could be reduced because homologous proteins with similar functions and domains always tend to have the same SCL, and *vice versa* [39-42].

The *Lactobacillales*-specific clusters of orthologous groups of proteins (LaCOGs) previously generated by Makarova *et.al* [43] were used as the basis for protein clustering into protein families. In total 3374 (729 new and 2645 existing) LaCOGs were formed by adding 14 recently sequenced LAB genomes to the Makarova et. al. set. Subsequently, a genome-scale SCL prediction was performed on all proteins in the 26 genomes (Table 1). By combining the SCL prediction and LaCOGs, and after manual curation (see below), we defined 462 secretome LaCOGs (of which 212 are new compared to the Makarova et. al. set) composed of 3357 proteins, representing 7.4% of the complete genome dataset and 93% of all predicted secretome proteins in these 26 genomes. We defined thirteen general functional classes for these proteins, and the distribution of these clustered secretome proteins over the classes and LaCOGs is shown in Figure 2. An additional 249 putative secretome proteins could not be grouped into these LaCOGs, comprising 69 proteins that had only a distant homolog in non-LAB, and 180 proteins that had no homolog in any sequenced bacterial genomes, which we termed the extracellular "ORFans" (Table 2, Additional file 1, sheet S1).

Although the LAB genomes vary in size, the size of the secretome as a fraction of each genome was fairly consistent (6-10%), as well as the distribution of proteins over different SCLs. The N-terminally anchored proteins with no signal peptidase cleavage site are the most abundant kind among all predicted secretome proteins. A striking feature of numerous secretome proteins, and particularly surface-associated proteins, is that they are large and consist of many different domains (often in repeats), and domain compositions (see examples in Figure 3). In fact, this variation in domain composition has been used in constructing and sub-dividing the LaCOGs and separating sub-families of homologous proteins. Distinct combinations of domains provide hints for functions of these extracellular proteins in cell-wall metabolism, cell-wall binding and their communication with the environment (see below).

False predictions and pseudogenes

The preliminary secretome clusters were curated manually and corrected based on expert knowledge, e.g. for false-positive and false-negative predictions, incorrect gene starts, pseudogenes, etc. Examples of proteins of known intracellular function, but with consistent false-positive extracellular SCL prediction are listed in Additional file 2, sheet S1. In most cases the mis-prediction was caused by an α-helix-like N-terminal sequence in these proteins (possibly as part of the hydrophobic core of a globular protein), leading to the prediction as a signal peptide by LocateP. A further improvement was made by finding and removing those LaCOGs that have proteins which are anchored in the cell membrane with a single N-terminal transmembrane helix, but with the rest of the protein inside the cell (so-called outside-in topology, GO:0031235) [44-53]. By aligning proteins within these LaCOGs we found that these proteins do not have positively charged residues preceding the N-terminal hydrophobic helix, but exclusively have a positively charged residue(s) immediately downstream of the transmembrane helix (examples in Additional file 2, sheet S2). Hence such features could be used for further development of a model for SCL prediction of N-terminally anchored proteins by LocateP.

Nearly 400 pseudogenes were identified, but this is probably an underestimate. In most cases this was due to gene frameshifts, and occasionally to N- or C-terminal truncation of genes. Most of these genes could be concatenated to encode larger proteins with high similarity to known proteins in the LaCOGs. Many of these pseudogenes were initially predicted to encode intracellular proteins by LocateP, but after concatenation these proteins are predicted to be extracellular and/or contain domains of extracellular functionalities. An example are the proteins encoded by adjacent genes LSA1731 and LSA1730 in *L.sakei* 23K which were annotated as hypothetical proteins. The concatenated protein

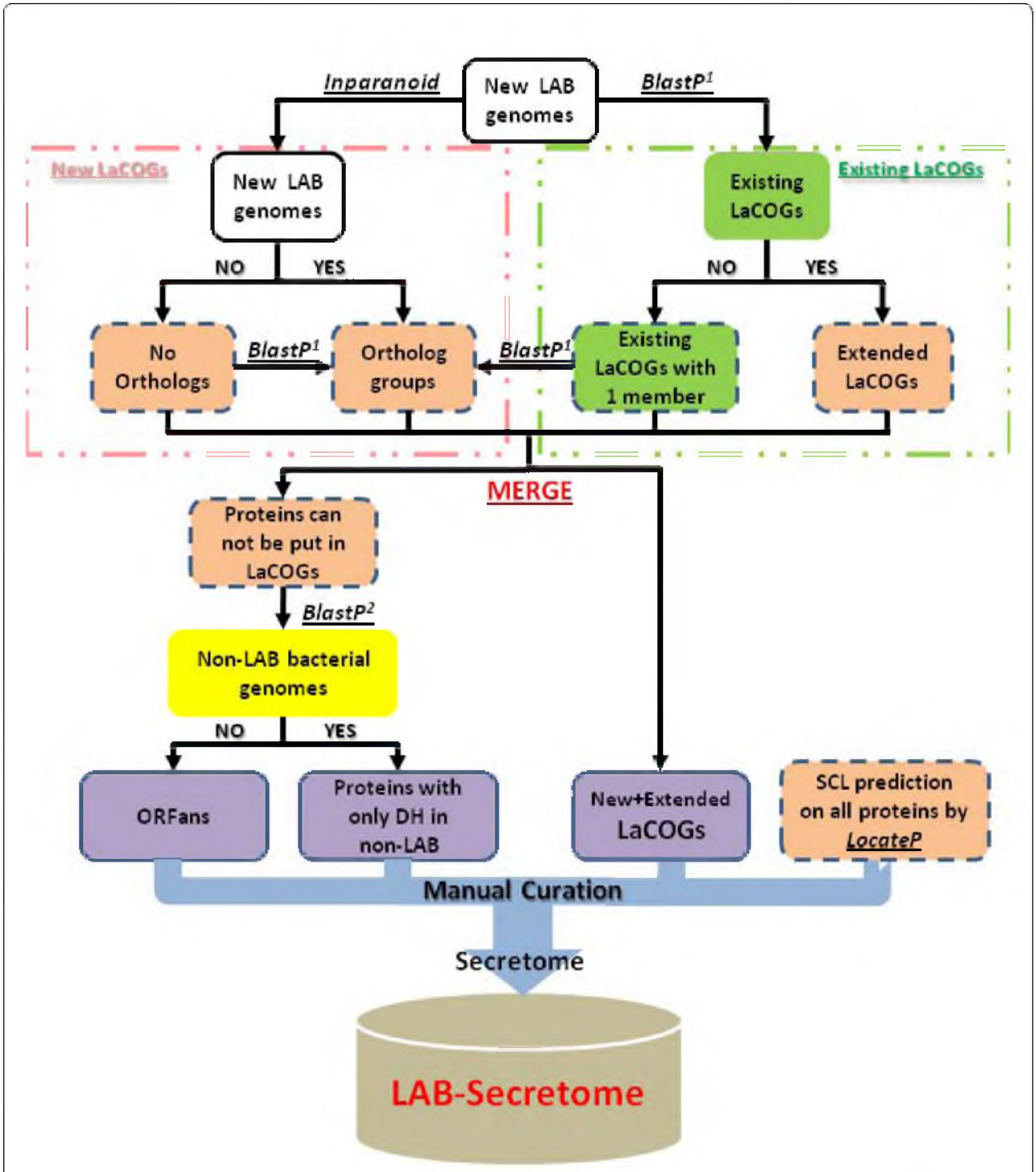

Figure 1 The flowchart for constructing the secretome LaCOGs. The completely sequenced LAB genomes are used as input data. No plasmid sequences were used for the Inparanoid search. The squares with dash-line frames are intermediate products that are not user-queryable from the LAB-Secretome interface; the squares with full-line frames are the final information stored in LAB-Secretome database. The upper left frame shows the processes that produce new LACOGs; the upper right frame shows the processes that extend existing LaCOGs. The new LaCOGs are coded starting with "9", the extended existing LaCOGs retain the original names from Makarova *et.al* [43]. BlastP[1]: the Blast results were processed by a revised criterion "uniform top 3" (see Material and Methods); BlastP[2]: the Blast results were processed by cut-off of 1e-3 and aligned sequence coverage of 60% for distant homolog identification. This work scheme can be used to update the LAB-Secretome database when new bacterial genomes are available.

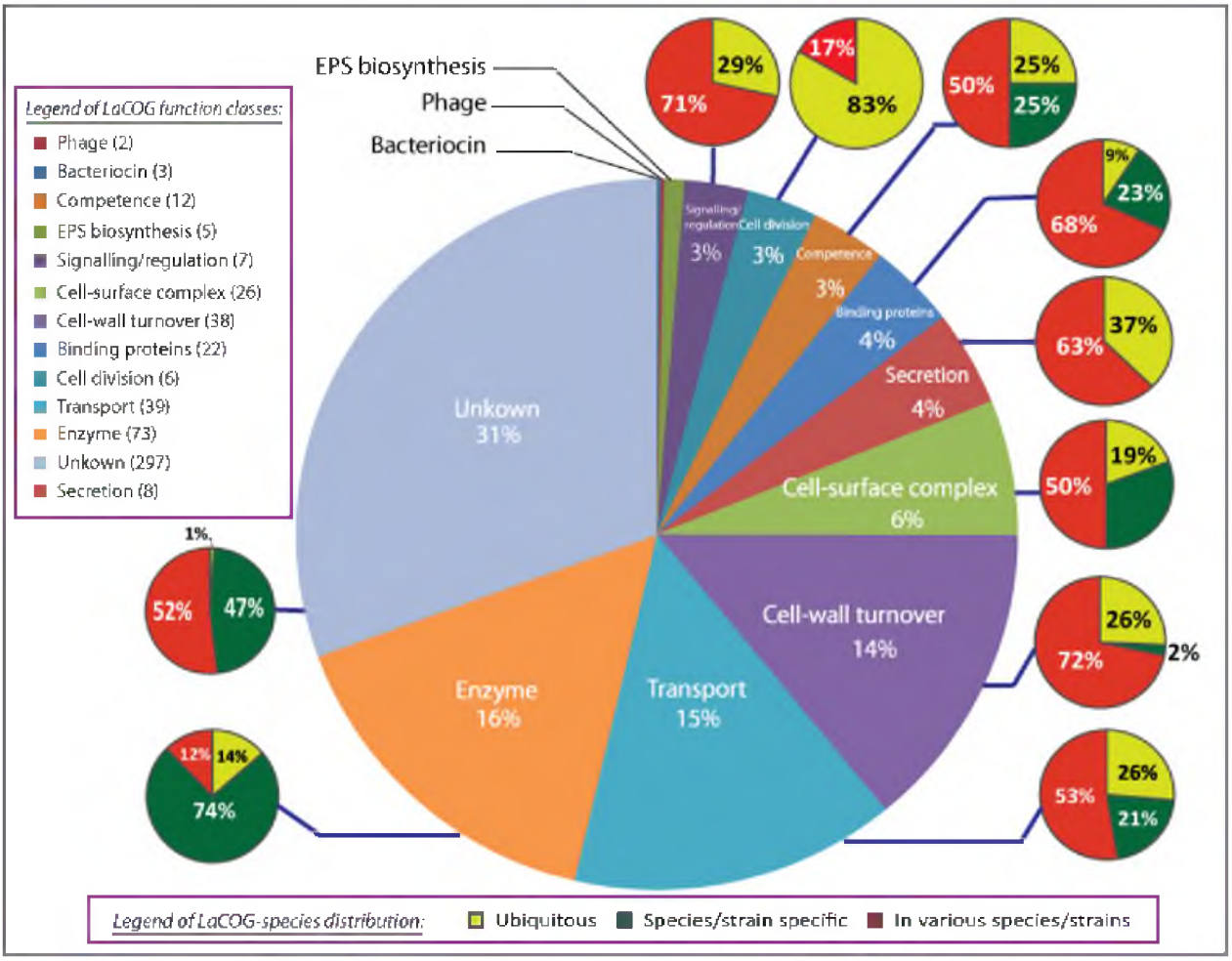

Figure 2 overview of distribution of secretome proteins in LaCOGs. The central pie depicts the distribution of secretome proteins in LaCOGs according to their functional classes. The percentage was calculated as the number of proteins in the category divided by the total of 3357 secretome proteins that were clustered into LaCOGs. The number of LaCOGs in each category is listed in the pie chart legend behind the name of the functional class. The separate yellow-red-green piecharts for each functional class represents the distribution of this LaCOG in the LAB genomes, i.e. ubiquitous, .species/strain-specific, or variable.

showed high similarity to proteins in LaCOG02935 which were exclusively cell-surface protein Csc complex family members [54]. In total 129 concatenated pseudoproteins were made with 279 protein fragments (Additional file 3, sheet S1), while 87 pseudogenes could not be combined (Additional file 3, sheet S2).

The LAB-Secretome database

The LAB-Secretome database http://www.cmbi.ru.nl/lab_secretome was constructed to store and browse all the predicted extracellular proteins and LaCOGs. An overview page summarizes all predicted secretomes, LaCOGs, distant homologs in non-LAB species and the ORFans, with hyperlinks to the corresponding HTML pages to help users to browse the whole database (Figure 4A). The LAB-Secretome database can be queried in many ways, e.g. by bacterial species, protein subcellular location, protein accession identifiers, LaCOG numbers, protein functional classes, and Pfam domain accession codes or domain functions (Figure 4B). Visualization includes a description of LaCOG members and function, protein functional domain composition, and multiple alignments with notification of corrected start codons, pseudogenes and concatenated proteins (Figure 4D). A Blast function, utilizing the BlastP [55] program, enables users to query the clustering information of their proteins of interest to the extracellular proteins and families that are already in the database (Figure 4C). An automatic updating scheme for the LaCOGs (Figure 1) was designed to ensure that the need for manual curation is minimized when adding new bacterial genomes to the database.

Table 2 Overview of the LaCOGs (genomes included in the original LaCOG analysis 43 are marked by *)

LAB species and strains	Secretome size	Proteins in LaCOG	Distant Homologs	ORFans	LaCOGs
E.faecalis V583	281	232	22	27	131
L.acidophilus NCFM	171	161	2	8	108
L.brevis ATCC 367	177	154	5	18	113
*L.casei ATCC 334 **	192	187	3	2	148
L.casei BL23	205	197	0	8	153
L.citreum KM20	112	112	0	0	93
*L.delbrueckii bulgaricus ATCC BAA-365 **	115	113	0	2	94
L.delbrueckii bulgaricus ATCC11842	87	79	3	5	68
L.fermentum IFO 3956	112	112	0	0	89
*L.gasseri ATCC 33323 **	115	113	0	2	88
L.helveticus DPC 4571	131	123	2	6	97
*L.johnsonii NCC 533 **	236	209	6	21	131
L.lactis cremoris MG1363	105	103	0	2	86
*L.lactis cremoris SK11 **	105	105	0	0	87
*L.lactis lactis IL1403 **	136	114	4	18	80
*L.mesenteroides ATCC 8293 **	112	94	5	13	77
*L.plantarum WCFS1 **	160	151	5	4	123
L.reuteri F275 JGI	159	156	1	2	124
L.reuteri F275 Kitasato	171	156	2	13	123
L.sakei 23K	114	103	4	7	80
L.salivarius UCC118	135	126	3	6	103
*O.oeni PSU-1 **	95	90	0	5	70
*P.pentosaceus ATCC 25745 **	99	89	1	9	79
*S.thermophilus CNRZ1066 **	90	90	0	0	77
*S.thermophilus LMD-9 **	97	94	1	2	84
S.thermophilus LMG 18311	94	94	0	0	81

Overview of the extracellular protein families

Ubiquitous/essential LaCOGs

Only 22 LaCOGs were found to be fully conserved among all 26 LAB secretomes, or only lacking in 1 genome (5 LaCOGs), e.g. the absence of an ATP-dependent protease from LaCOG01453 in *P. pentosaceus* (Additional file 1, sheet S3).

Most of these LaCOGs contain proteins with universal functionalities involved in cell-wall metabolism, secretion, transport and DNA uptake (Figure 2). Only one conserved family (LaCOG01219) contains proteins of as yet unknown function, but presumably essential as they are conserved in all genomes.

Most common functionalities in the secretomes of LAB

Among all 215 secretome LaCOGs with known or presumed functions, almost half of them contain proteins which are involved in cell-wall metabolism, e.g. the muramidase, lysin, lysozyme and beta-lactamase families (Figure 2). Many of these enzyme families are further subdivided into different LaCOGs based on variations in sequence homology and protein domain compositions, and some may represent species/niche-specific subfamilies. One example is the subdivision of proteins with an Nlpc/P60 family domain (e.g. gamma-D-glutamate-meso-diaminopimelate muropeptidase) into 5 separate LaCOGs (Additional file 4, sheet S1). These proteins vary in length from ~150 to ~500 amino acids, all with the Nlpc/P60 domain in the C-terminal part. In only one of these subfamilies (LaCOG90015), all 16 members have 1-3 copies of LysM domains (Pfam PF01476) in their N-terminal part, indicating extra binding functions to the cell-envelope. A similar domain architecture is found in one of the four N-acetylmuramoyl-L-alanine amidase subfamilies (LaCOG01848), which has an enzymatic C-terminal domain and 0-3 N-terminal SH3 domains (Pfam PF08239), known to bind to proline-rich regions of proteins. In the pepdidoglycan hydrolase subfamilies LaCOG00186 and LaCOG01653 the enzymatic domain is located at the N-terminus and can be followed by different kinds, combinations and numbers of binding domains such as LysM, SH3 or surface layer domain (Pfam PF03217) (Figure 5). These examples all illustrate that the many types of extracellular enzymes involved in cell-wall turnover have different mechanisms to attach to components of the cell surface.

Niche-specific LaCOG families

1/L. acidophilus complex specific The *acidophilus* "complex" including the species *L. acidophilus,*

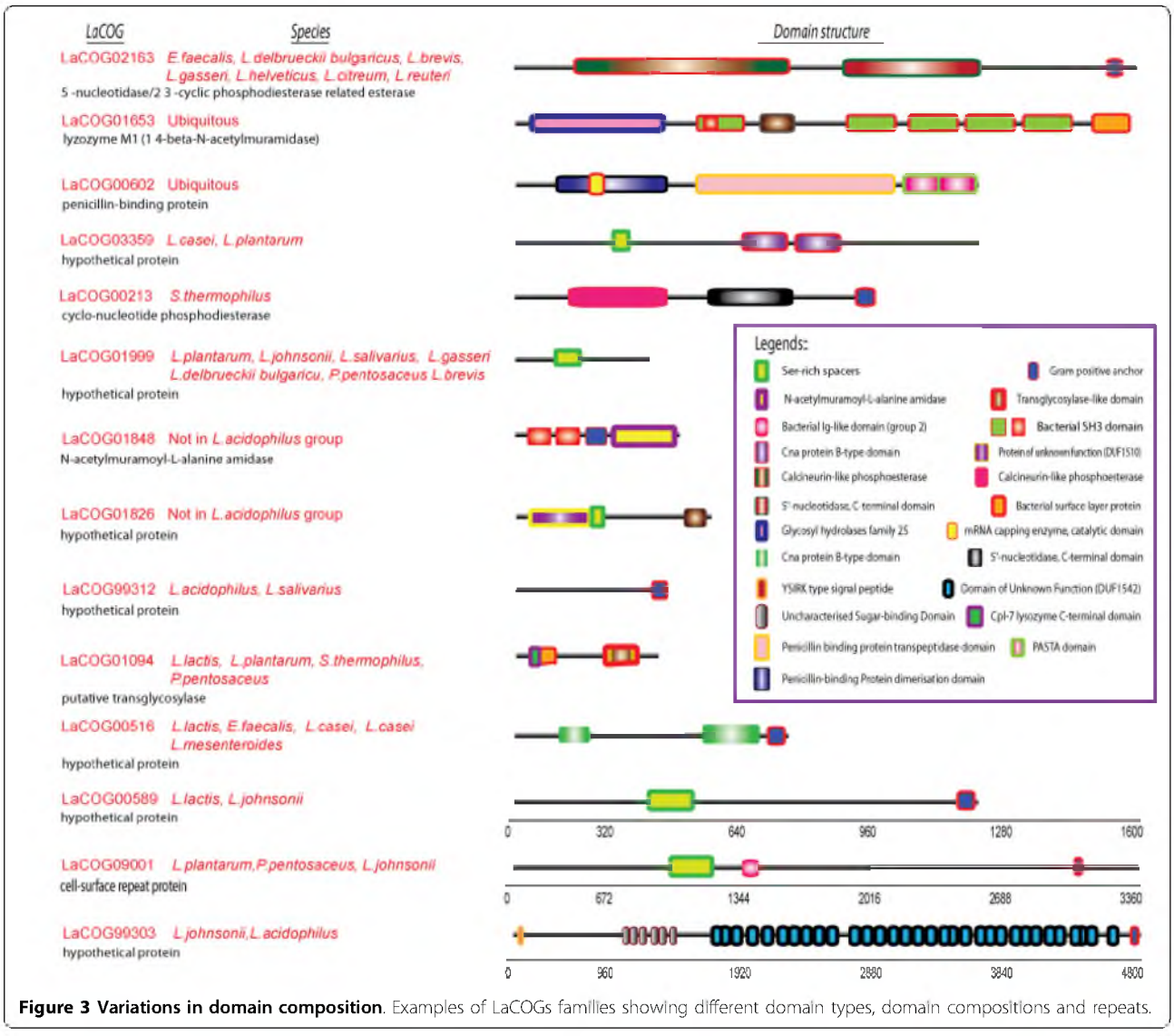

Figure 3 Variations in domain composition. Examples of LaCOGs families showing different domain types, domain compositions and repeats.

L. johnsonii, L. gasseri, L. delbrueckii ssp bulgaricus and *L. helveticus* has long been regarded as a phylogenetic subgroup [56-58]. About 30 LaCOGs appear to be specific for these species (Additional file 1, sheet S4). Their proteins include an ABC-type phosphate/phosphonate transport system (LaCOG02118), the aggregation promoting factor (LaCOG90005) [59-61], a putative competence protein (LaCOG03110) and several families of S-layer proteins, which may reflect the special binding function that these S-layer proteins generally share in these *acidophilus* complex species [62-69]. Interestingly, twenty of these *acidophilus* complex-specific LaCOGs contain only extracellular proteins of unknown function, and it should be challenging to focus on experimental determination of their function.

2/GI-tract specific If we consider the LAB species *L. acidophilus, L. johnsonii, L. gasseri, L. reuteri,* and *L. salivarius* to be specifically found in the GI-tract, then we can identify 17 LaCOGs which are not found outside of this group, of which 13 families contain only proteins of unknown function (Additional file 1, sheet S4). One mucus-binding protein family (LaCOG02280) was found to be specific for these GI-tract LAB, and contains 4 proteins from *L. acidophilus, L. gasseri* and *L. johnsonii*. All four proteins are larger than 2300 amino acids, contain a signal peptide with YSIRK domain (Pfam PF04650) and appear to be anchored to the peptidoglycan by an LPxTG cell-wall anchor (Pfam PF00746). Each protein has 5-11 copies of a mucus-binding domain, as defined by Boekhorst et al [60], showing their particular role in binding to mucus components in the GI-tract [5,70-72]. The 3 D structure of this domain of 184 residues has recently been determined and shows similarity to the functional repeat found in a family of immunoglobulin-binding proteins [73].

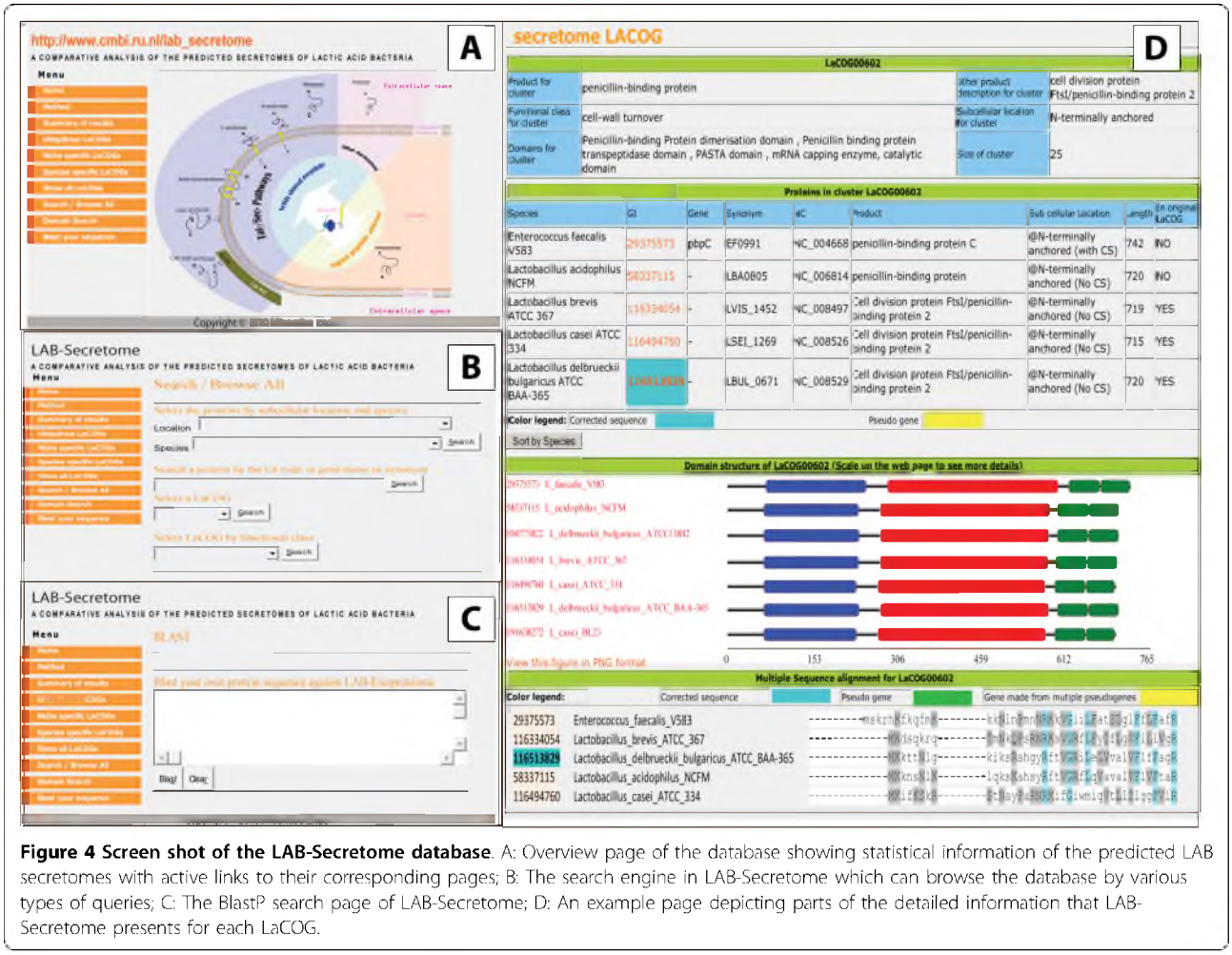

Figure 4 Screen shot of the LAB-Secretome database. A: Overview page of the database showing statistical information of the predicted LAB secretomes with active links to their corresponding pages; B: The search engine in LAB-Secretome which can browse the database by various types of queries; C: The BlastP search page of LAB-Secretome; D: An example page depicting parts of the detailed information that LAB-Secretome presents for each LaCOG.

3/Plant-associated specific Twelve LaCOGs appear to be specific for the group of plant-associated species *Leuconostoc, Oenococcus, L. plantarum, L. brevis*, and *P. pentosaceus*, of which 7 familes contain only proteins of unknown function (Additional file 1, sheet S4). One of these (LaCOG02876) includes 4 homologous proteins from *L.brevis, L.plantarum, O.oeni* and *L. citreum*, which show a high sequence similarity to each other, but the protein from *L. plantarum* has a much longer serine-rich spacer between the N- and C-terminal domains. A similar domain structure differing in a long serine-rich spacer is seen in the 2 hypothetical proteins from *L. plantarum* and *L. brevis* in LaCOG02927.

4/Dairy LAB specific A few protein families were found only to occur in the secretomes of the dairy LAB *S. thermophilus, L. lactis and E. faecalis* (Additional file 1, sheet S4). These proteins have functional properties that may be relevant to the dairy niche, e.g. LaCOG00374 contains ABC transporter substrate-binding proteins for polar amino acids, and could possibly be required for growth in milk [74-77]. The *L. lactis* strains have a single copy of this gene, while the *S. thermophilus* strains all have 3 consecutive genes encoding paralogs of this amino acid-binding protein. All dairy *Streptococcus* and *Lactococcus* strains contain a single gene encoding a beta-lactamase (LaCOG00012) which may play a role in destroying penicillin that these strains may encounter in milk [78-82]. A putative chitinase (glycosyl hydrolase family 18; LaCOG02690) is found exclusively in *E. faecalis* and in *L. lactis* strains.

Species-specific and strain-specific LaCOGs

Up to 150 LaCOGs were found to be species-specific or strain-specific (Additional file 1, sheet S5). The distinction is not so clear yet because for some species several strains were sequenced (e.g. *L. lactis, S. thermophilus*) while for many species only a single strain was sequenced to date. Most of these families are made up solely of hypothetical proteins with highly conserved sequence (Figure 2). *L. casei* and *L. lactis* have the highest number of species-specific LaCOGs, indicating that they may have more unique extracellular functions. Examples of species-specific extracellular proteins are the PrgA/PrgB/

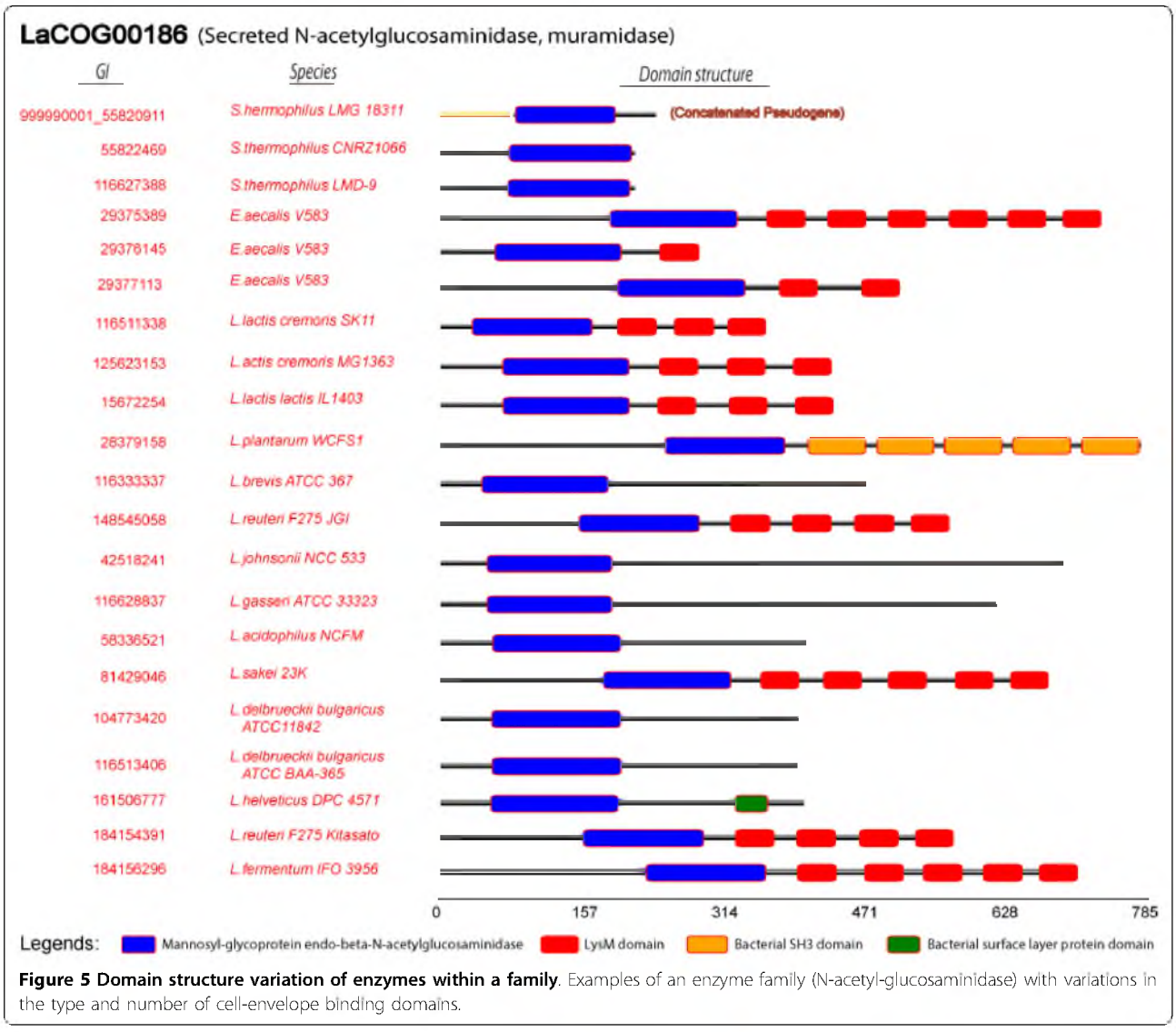

Figure 5 Domain structure variation of enzymes within a family. Examples of an enzyme family (N-acetyl-glucosaminidase) with variations in the type and number of cell-envelope binding domains.

PrgC surface proteins of *E. faecalis* [83-85], an alpha-amylase (LaCOG02644) in *L. lactis* strains, a phospholipase A2 family enzyme (LaCOG99223) in *L. casei* strains, a cyclo-nucleotide phosphodiesterase (LacOG00213) in *S. thermophilus* strains, and a mucus-binding protein (LaCOG90010) in *L. delbrueckii* strains.

Extracellular proteins not in LaCOGs: ORFans and proteins with only distant homologs in non-LAB

About 249 putative extracellular proteins could not be classified into LaCOG families, and comprise 69 proteins that have only distant homologs in non-LAB species and 180 ORFans that are species-specific (Additional file 1, sheets S6 and S7). While the ORFans are nearly all hypothetical proteins of unknown function, the distant homologs also contain proteins with a variety of known functions, such as extracellular enzymes (e.g. xylanase, pectate lyase, endo-beta-N-acetylglucosaminidase, proteases and beta-fructosidase), substrate-binding proteins of transporters, miscellaneous binding proteins and specific bacteriocins. The uniqueness of these proteins suggests that most species or strains have a few unique extracellular proteins that are not found in other sequenced LAB, and may encode unique functions that are related to their environmental niche. Quite a few of the proteins of unknown function are predicted to be lipid-anchored and therefore may represent substrate-binding proteins of uncharacterized transporters.

Specific enzyme families

LAB possess a variety of extracellular hydrolytic enzymes and transglycosylases which presumably relate to interactions with their environment, e.g. for degradation of growth substrate polymers. These enzymes have been

clustered and sub-divided into protein families (LaCOGs) based on specific domain compositions (Table 3, Additional file 4, sheet S2). For instance, the subtilisin-like serine proteases (Pfam PF00082), known to be important for growth on protein substrates [86-89], were clustered into 2 LaCOGs: the first family (LacOG02153) is composed of 7 proteins containing a protease-associated PA domain (Pfam PF02225) inserted in the catalytic domain which forms a lid structure that covers the active site, whereas the other family (LaCOG90024) was only found in *L. casei* and *L. acidophilus*, and contains subtilisin-like serine proteases without the PA domain. Putative transglycosylases, also referred to as aggregation-promoting factors [59,90-92], are divided into three subfamilies (LaCOG01580, LaCOG02932, LaCOG90005), and have a highly conserved C-terminal domain [71]. Furthermore, there are several families of hydrolases of unknown function (Table 3). The extracellular alpha/beta hydrolases with a DUF915 domain (Pfam PF06028) are subdivided into four families, two of which are highly populated (LaCOG01137 and LaCOG01138, with 46 and 30 members, respectively) and found in nearly all LAB, suggesting that they have an essential, but as yet unknown, function.

Specific binding-protein families

Many extracellular proteins contain known domains for binding to macromolecular substrates. In addition to domains for binding to the cell wall of the producing cell (e.g. LysM, SH3), several other domains are found which are related to binding to host macromolecules (e.g. domains annotated as mucus-binding, chitin-binding, collagen-binding, fibronectin-binding, carbohydrate-binding, etc) (Table 3). Some of these annotations derive from *in vitro* binding studies and may not reflect *in vivo* functions. In LAB, mucus-binding domains (MUB, MucBP) are found in many proteins and are thought to play a role in binding to the host GI-tract mucus layer [57,93,94]. An enormous variety is found in the size of these mucus-binding proteins and in the number of mucus-binding domains. We have made a preliminary separation into 7 different subfamilies of mucus-binding proteins based on protein size, sequence homology, domain composition and phylogeny (Table 3). The three largest subfamilies are (1) LaCOG00885 containing 11 members from different LAB but not from *L. acidophilus* group members, (2) LaCOG01470 with 28 members, found in many LAB, and (3) LaCOG03211 which includes 10 proteins. The proteins of LaCOG00885 contain solely the MucBP domains as defined by Pfam (PF00746), while the proteins of the other two LaCOGs possess multiple copies of the larger MUB domains as defined by Boekhorst *et al.* [71] (see also Figure 2 in [95]). Many mucus-binding proteins of *L. acidophilus* group members contain an N-terminal [Y/F] SIRKxxxGxxS-containing signal peptide (PF04650) which was earlier reported as a typical characteristic of the *L. acidophilus* MUB proteins [94,96], and may relate to a specific function in sorting or folding [97,98]. Furthermore, it is striking that many large genes encoding mucus-binding proteins are pseudogenes (e.g. in LaCOG01470, LaCOG03211 and LaCOG99309). While it is unlikely that these are all due to sequencing errors, it is not clear yet whether these are truly pseudogenes, or possibly may encode functional proteins after transcription with strand-slipping [5,71].

Conclusions

Lactic Acid Bacteria (LAB) occur naturally in many different fermentation environments such as plant, meat, dairy and cereal. Overall similarities have been identified among the genomes of many LAB species [61,99-105]. However, bio-diversity has also been reported frequently, showing that subtle variations in presence or absence of proteins and functional domain composition might lead to important traits during bacterial adaptation to their living environments [106-113]. Our comparative research on extracellular and surface-associated protein families has provided a more solid basis for this hypothesis. Universal families have been identified which are apparently essential for survival of all LAB, but also species-specific protein families. Besides the clustered proteins with known functions, many families of hypothetical proteins and unique proteins (ORFans and proteins with only distant homologs in non-LAB) were found.

Protein clustering supports niche-dependent features of specific subgroups of LAB (e.g. the *L. acidophilus* group) and could aid in linking bacterial phenotypes to genotypes. The distinct sub-families of the different LaCOGs have provided clues for adaptation of the bacterial cells to their living environment, such as the GI-tract. The result of this study can be used as leads for experimental work on the molecular evolution, diversity, function and adaptation of bacteria to specific environments.

Our clustering methods and database structure were designed in a way that allows adoption to other groups of bacteria than LAB. The analysis results are stored in a queryable database which provides vivid browsing functions for users, and will be updated regularly to guarantee the continuation of the service to the biology community. Our clustering information into families could definitely help in checking the quality of newly sequenced genomes and for genome (re-) annotation.

Table 3 Examples of specific enzyme and binding-protein sub-families

Product	LaCOG	Functional domain	Distribution	Special features
Specific enzyme families				
Subtilisin-like serine protease	LaCOG02153	Subtilase family	*L. casei, L. delbrueckii bulgaricus, L. johnsonii, L. lactis, S. thermophilus*	PA domain (PF02225) inserted in the subtilase family domain
	LaCOG90024	Subtilase family	*L. acidophilus, L. casei*	no PA insert domain
Trans-glycosylase	LaCOG01094	Transglycosylase-like domain,	mainly in *L.plantarum, L.lactis, S.thermophilus*	different domains for PG binding
	LaCOG01589	aggregation promoting factor related surface protein	not in *L.acidophilus* group	PG bound by LysM domain; highly conserved C-terminal domain ending in GWY
	LaCOG02932	aggregation promoting factor related surface protein	only in *L.delbrueckii bulgaricus, L.plantarum, L.acidophilus* group	highly conserved C-terminal domain ending in WY
	LaCOG90005	aggregation promoting factor related surface protein	only in *L.acidophilus* group	highly conserved C-terminal domain ending in GWY
Dextran sucrase	LaCOG90016	glycosyl hydrolase family 70	only in *Leuconostoc, L. reuteri, O. oeni*	
Cell-surface hydrolases				
alpha/beta hydrolase	LaCOG01137	alpha/beta hydrolase of unknown function (DUF915)	ubiquitous	
	LaCOG01138	alpha/beta hydrolase (DUF915)	Ubiquitous	
	LacOG01920	alpha/beta hydrolase (DUF915)	only in *L. delbrueckii bulgaricus, L.plantarum , L.casei*	
	LaCOG02785	alpha/beta hydrolase (DUF915)	only in *L.plantarum , L.casei , L.sakei*	
lipase/Acyl-hydrolase	LaCOG00342	GDSL-like Lipase/Acylhydrolase	not in *L.acidophilus* group	with GDSL-like motif
general cell surface hydrolase	LacOG02019	cell surface hydrolase membrane-bound (putative)	only in *L.delbrueckii bulgaricus, L.plantarum,L.casei , L.fermentum*	
	LaCOG01618	cell-surface hydrolase;	only *in L.plantarum , L.delbrueckii bulgaricus, P.pentosaceus*	
Binding proteins				
mannose-specific adhesion	LaCOG01741	MUB domain, Gram positive anchor	only in *L.plantarum, L.delbrueckii bulgaricus, P.pentosaceus, L.acidophilus* group	
collagen-binding protein	LaCOG00092	Collagen binding domain, Gram positive anchor	not in *L.acidophilus* group	
mucus-binding protein	LaCOG00885	MucBP domain (Classical), Gram positive anchor	not in *L.acidophilus* group	Leucine Rich Repeat, PT repeat
	LaCOG01470	MUB domain, Gram positive anchor		many pseudogenes, most *L.acidophilus* group proteins have YSIRK-type signal peptide
	LacOG02280	MUB domain, Gram positive anchor	only in *L.acidophilus* group	very large, YSIRK-type signal peptide
	LaCOG03211	MUB domain, Gram positive anchor		5 of 10 are pseudogenes; YSIRK SP in *L. acidophilus* group members
	LacOG99309	MUB domain, Gram positive anchor	only in *L.acidophilus* group	all pseudogenes; YSIRK type signal peptide
chitin-binding protein	LaCOG01300	Chitin binding domain	*E.faecalis, L.plantarum, L.sakei, L.lactis*	maybe related to niche
adherence protein	LaCOG01366	von Willebrand factor type A domain, Cna protein B-type domain	only in *L.lactis, E.faecalis, L.citreum, L.casei*	

Methods

Genome sequences and bioinformatics tools used in this research

The genome sequences of 26 selected representative lactic acid bacteria, including the protein functional annotation and the gene contexts, were obtained from the NCBI bacterial genome database (version 15 Aug., 2008) [114].

BlastP (default cutoff values of E < 1, low-complexity filter disabled) [55] and Inparanoid [115] were used for sequence homology and orthology searches, respectively. Protein subcellular location (SCL) was predicted by LocateP [38]. Multiple sequence alignments were constructed using Muscle [116]. Motif searches were performed using MEME and MAST [117]. Protein domains (version Dec. 2008) [118] originating from the Pfam database [119-121] and additional HMMs reported in other studies [54,71,96,122-124] were searched using HMMER [125] with the respective cut-off of each model. The domain functions were obtained from the GO database [126] using the PFAM2GO dataset [126].

The LAB-Secretome database was created in MySQL and the database interface was written in PHP (version5.2.7). Visualization of the protein domain composition was made using scalable vector graphics (SVG).

Protein clustering into orthology groups (LaCOGs)

First, the 22,191 proteins in 3195 LaCOGs generated by Makarova *et.al* [1] from 12 LAB genomes were used as the basis for protein clustering. All protein sequences from 14 newly sequenced LAB genomes were searched against the Makarova LaCOG set using BlastP. The proteins that have high homology to the existing LaCOGs were then selected using a revised criterion based on the well-known COG extension rule "uniform top 3" [127]: if all the top 3 (in case of LaCOG size of 2, the top 2 hits were taken) BlastP hits of a query protein belong to the same LaCOG (LaCOG size bigger than or equals to 2), then the query protein is added to this LaCOG.

Since the above-mentioned extension was purely based on the homologs of proteins that were already included in the LaCOGs by Makarova *et al.*, the specific proteins from newly sequenced species, e.g. *L. reuteri*, were not added due to the absence of the "seeding sequences" for BlastP. In order to cluster all proteins that originated from the newly sequenced genomes, a complete all-to-all Inparanoid [115] search was performed in a parallel fashion with the proteins encoded in the 14 new genomes to identify orthologous proteins. Cut-off settings of bit score 50 and sequence overlap of 50% were used. The proteins with all-to-all bidirectional-best-hit (BBH) relationship [128,129] were clustered into groups, meaning that in any such group, each member is the BBH of another member. This stringent criterion generates new cores of orthologous proteins.

Using the core ortholog clusters and the extended LaCOGs made in step one, the proteins that were not previously included in any clusters, including those proteins from Makarova LaCOGs containing only 1 member, were Blasted as queries. In this step, the revised criterion "uniform top 3" was used and new LaCOGs were made.

The newly made LaCOGs were merged with the extended Makarova LaCOGs, and the newly made ones were assigned coding numbers starting with "9" in their names, e.g. LaCOG90001, to distinguish them from the extended Makarova LaCOGs.

LaCOG quality control

In order to check the quality of the merged LaCOGs, an iterative BlastP search was performed using the clustered proteins as queries against all the proteins that were not included in any constructed LaCOGs, using the criteria of 1E-3 and query-hit protein length ratio of 0.6, which has been tested by *Boekhorst et. al.* [130] for distant homolog identification. This iterative search found that only 13 non-clustered proteins (mostly hypothetical proteins) had a distant homolog in 11 different LaCOGs, indicating that our clustering methods have extensively included most of the proteins into possible homologous clusters.

ORFans and proteins with only non-LAB distant homologs

The LAB proteins that could not be clustered into LaCOGs by the previously described procedures were then collected and Blasted against all completely sequences non-LAB bacterial genomes (both Gram- and Gram+ species). The same criterion of distant homolog identification [130] was utilized. Proteins that had no homologs in any other species were named "ORFans".

Secretome LaCOG extraction

The clustering information of merged LaCOGs, proteins that have only distant homologs in non-LAB species and the ORFans was then combined with the SCL prediction made by LocateP (Table 1). Initially, only the LaCOGs that had at least half of the members with a predicted secretome SCL corresponding to (1) lipid-anchored; (2) N-/C-terminally anchored; (3) secreted by Tat- or Sec- pathway; (4) secreted via non-classical pathways, or (5) cell-wall anchored were identified as the secretome LaCOGs. Later, all other LaCOGs were manually inspected, and a few families were identified with a mixture of secretome and intracellular proteins; only the secretome proteins were added to the database. The same classification was applied to the secretome ORFans and proteins that have only distant homologs in

non-LAB species. The resulting clusters of secretome proteins, the "secretome", can be further extended by similar processes when new (LAB) genome sequences become available.

Proteins that are exported by unknown mechanisms and so-called "moon-lighting" proteins (known intracellular function, but often also found on the outside of the cell) [131] were not considered as their extracellular SCL cannot be predicted.

Manual curation

In order to obtain as accurate as possible prediction of secretome proteins and their classification into LaCOGs, we performed a throughout manual inspection on all the secretome proteins, including the ORFans and the ones included in LaCOGs. All proteins were double checked for the ORF-calling quality by the criteria combining protein length, possible alternative start (end) codon, multiple sequence alignments, protein domain composition and SCL prediction consistency.

Incorrectly chosen start codons in the original annotations were corrected based on sequence alignment with protein family members, position of putative ribosome-binding sites, and known features of signal peptides. Pseudogenes were initially identified when BLASTP analysis of the encoded proteins showed that they belong to extracellular protein families in LaCOGs, but that they represented only a fragment of the protein. By analysis of the coding region of these pseudogenes with their adjacent nucleotide sequences we could generally identify frameshifts, such that the missing protein part(s) were found to be encoded in a different reading frame. In these cases, the entire opening-reading frames were translated into protein fragments, regardless of the absence of start codons, and these protein fragments were concatenated to form new protein sequences that share high similarity to other known full-length proteins. In a few cases, ORFans were also identified as pseudogenes when they lacked a signal peptide, but otherwise contained protein domains typical of extracellular proteins.

Generally, we expected the ORFans to be real genes that represent unique functionality to the specific LAB in which they occur. However, because the average size of these hypothetical ORFs was below 100 amino acids, it is possible that some small ORFans could as well be wrongly predicted ORFs or pseudogenes. Proteins smaller than 80 amino acids containing only a Sec-type N-terminal signal sequence were removed from the set of predicted extracellular proteins, since their C-terminal part is generally too small to represent an extracellular domain. Moreover, many of such small proteins with a single predicted TM helix are now increasingly considered as small integral membrane proteins [132].

Additional material

Additional file 1: The overview of LAB-Secretome. Sheet S1: an overview of secretomes included in this research; sheet S2: the presence and absence patterns of the LaCOGs in 26 LAB genomes; sheet S3: the ubiquitous LaCOGs; sheet S4: the niche-specific LaCOGs; sheet S5: the species-specific LaCOGs; sheet S6: the ORFans; S7: the proteins with only distant homologs.

Additional file 2: False-positive SCL predictions. The false-positive SCL predictions that were corrected using domain composition and homolog information of LaCOGs. Sheet S1: the intracellular proteins that had been wrongly predicted to be extracellular; sheet S2: the N-terminally anchored LaCOGs with C-terminal inside topology.

Additional file 3: The extracellular pseudogenes. The secretome pseudogenes. The pseudogenes with wrongly annotated start/end codons were corrected and concatenated with corresponding gene neighbors. The resulting proteins seem to have homologs in various LaCOGs. The concatenated protein sequences are listed in the last column, with an "x" showing the conjunction site of each sequence.

Additional file 4: Interesting cases of extracellular protein families. The distribution of binding protein families: sheet S1: Nlpc-P60 families; sheet S2: Cell surface hydrolase; sheet S3: Binding proteins.

Acknowledgements

M Zhou was funded by the BioRange programme of the Netherlands Bioinformatics Centre (NBIC) which is supported by a BSIK grant through the Netherlands Genomics Initiative (NGI). DT, MW and RS are partly supported by the TI Food and Nutrition and the Kluyver Centre for Genomics of Industrial Fermentation.

Author details

[1]Centre for Molecular and Biomolecular Informatics, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands. [2]TI Food and Nutrition, Wageningen, The Netherlands. [3]NIZO food research, Ede, The Netherlands.

Authors' contributions

MZ, DT and MW carried out the LaCOGs and database construction. RS performed the manual curation of the clustered proteins and DT and MM refined the LAB-Secretome database. MM and DT drafted the manuscript. Both MW and RS participated in its coordination and helped to draft and finalize the manuscript. All authors read and approved the final manuscript.

Received: 5 August 2010 Accepted: 23 November 2010
Published: 23 November 2010

References

1. Goh YJ, Klaenhammer TR: **Genomic features of Lactobacillus species.** *Front Biosci* 2009, **14**:1362-1386.
2. Johannessen EA, Wang L, Wyse C, Cumming DR, Cooper JM: **Biocompatibility of a lab-on-a-pill sensor in artificial gastrointestinal environments.** *IEEE Trans Biomed Eng* 2006, **53(11)**:2333-2340.
3. Ljungh A, Wadstrom T: **Lactic acid bacteria as probiotics.** *Curr Issues Intest Microbiol* 2006, **7(2)**:73-89.
4. Vaughan EE, de Vries MC, Zoetendal EG, Ben-Amor K, Akkermans AD, de Vos WM: **The intestinal LABs.** *Antonie Van Leeuwenhoek* 2002, **82(1-4)**:341-352.
5. Klaenhammer TR, Altermann E, Pfeiler E, Buck BL, Goh YJ, O'Flaherty S, Barrangou R, Duong T: **Functional genomics of probiotic Lactobacilli.** *J Clin Gastroenterol* 2008, **42(Suppl 3 Pt 2)**:S160-162.
6. Settanni L, Corsetti A: **Application of bacteriocins in vegetable food biopreservation.** *Int J Food Microbiol* 2008, **121(2)**:123-138.
7. Reddy G, Altaf M, Naveena BJ, Venkateshwar M, Kumar EV: **Amylolytic bacterial lactic acid fermentation - a review.** *Biotechnol Adv* 2008, **26(1)**:22-34.

8. Lindgren SE, Dobrogosz WJ: **Antagonistic activities of lactic acid bacteria in food and feed fermentations.** *FEMS Microbiol Rev* 1990, **7(1-2)**:149-163.
9. Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J: **Polyphasic taxonomy, a consensus approach to bacterial systematics.** *Microbiol Rev* 1996, **60(2)**:407-438.
10. Stiles ME, Holzapfel WH, Holzapfel WH: **Lactic acid bacteria of foods and their current taxonomy.** *Int J Food Microbiol* 1997, **36(1)**:1-29.
11. Wood BJB, Holzapfel WH, (eds): **The genera of lactic acid bacteria.** Glasgow, United Kingdom.: Blackie Academic and Professional, 1 1995.
12. Salminen S, von Wright A, Ouwehand AC, Lahtinen S, (eds): **Lactic Acid Bacteria: Microbiological and Functional Aspects.** New York: Marcel Dekker, Inc, Third 2004.
13. Callon C, Millet L, Montel MC: **Diversity of lactic acid bacteria isolated from AOC Salers cheese.** *J Dairy Res* 2004, **71(2)**:231-244.
14. Ennahar S, Cai Y, Fujita Y: **Phylogenetic diversity of lactic acid bacteria associated with paddy rice silage as determined by 16 S ribosomal DNA analysis.** *Appl Environ Microbiol* 2003, **69(1)**:444-451.
15. Michel C, Pelletier C, Boussaha M, Douet DG, Lautraite A, Tailliez P: **Diversity of lactic acid bacteria associated with fish and the fish farm environment, established by amplified rRNA gene restriction analysis.** *Appl Environ Microbiol* 2007, **73(9)**:2947-2955.
16. Ouoba LI, Nyanga-Koumou CA, Parkouda C, Sawadogo H, Kobawila SC, Keleke S, Diawara B, Louembe D, Sutherland JP: **Genotypic diversity of lactic acid bacteria isolated from African traditional alkaline-fermented foods.** *J Appl Microbiol* 2010, **108(6)**:2019-2029.
17. Lebeer S, Vanderleyden J, De Keersmaecker SC: **Genes and molecules of lactobacilli supporting probiotic action.** *Microbiol Mol Biol Rev* 2008, **72(4)**:728-764, Table of Contents.
18. Kleerebezem M, Vaughan EE: **Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity.** *Annu Rev Microbiol* 2009, **63**:269-290.
19. van der Flier M, Chhun N, Wizemann TM, Min J, McCarthy JB, Tuomanen EI: **Adherence of Streptococcus pneumoniae to immobilized fibronectin.** *Infect Immun* 1995, **63(11)**:4317-4322.
20. Nitsche-Schmitz DP, Rohde M, Chhatwal GS: **Invasion mechanisms of Gram-positive pathogenic cocci.** *Thromb Haemost* 2007, **98(3)**:488-496.
21. Chaussee MS, Somerville GA, Reitzer L, Musser JM: **Rgg coordinates virulence factor synthesis and metabolism in Streptococcus pyogenes.** *J Bacteriol* 2003, **185(20)**:6016-6024.
22. Del Nobile MA, Altieri C, Corbo MR, Sinigaglia M, La Notte E: **Development of a structured model for batch cultures of lactic acid bacteria.** *J Ind Microbiol Biotechnol* 2003, **30(7)**:421-426.
23. Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, Leer R, Tarchini R, Peters SA, Sandbrink HM, Fiers MW, *et al*: **Complete genome sequence of Lactobacillus plantarum WCFS1.** *Proc Natl Acad Sci USA* 2003, **100(4)**:1990-1995.
24. Guzzo J, Jobin MP, Delmas F, Fortier LC, Garmyn D, Tourdot-Marechal R, Lee B, Divies C: **Regulation of stress response in Oenococcus oeni as a function of environmental changes and growth phase.** *Int J Food Microbiol* 2000, **55(1-3)**:27-31.
25. Ma Y, Curran TM, Marquis RE: **Rapid procedure for acid adaptation of oral lactic-acid bacteria and further characterization of the response.** *Can J Microbiol* 1997, **43(2)**:143-148.
26. de Vos WM, Vaughan EE: **Genetics of lactose utilization in lactic acid bacteria.** *FEMS Microbiol Rev* 1994, **15(2-3)**:217-237.
27. Desvaux M, Hebraud M, Talon R, Henderson IR: **Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue.** *Trends Microbiol* 2009, **17(4)**:139-145.
28. Rodriguez-Ortega MJ, Norais N, Bensi G, Liberatori S, Capo S, Mora M, Scarselli M, Doro F, Ferrari G, Garaguso I, *et al*: **Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome.** *Nat Biotechnol* 2006, **24(2)**:191-197.
29. Toomey N, Monaghan A, Fanning S, Bolton D: **Transfer of antibiotic resistance marker genes between lactic acid bacteria in model rumen and plant environments.** *Appl Environ Microbiol* 2009, **75(10)**:3146-3152.
30. Buck BL, Altermann E, Svingerud T, Klaenhammer TR: **Functional Analysis of Putative Adhesion Factors in Lactobacillus acidophilus NCFM.** *Appl Environ Microbiol* 2005, **71(12)**:8344-8351.
31. Kawai R, Igarashi K, Samejima M: **Gene Cloning and Heterologous Expression of Glycoside Hydrolase Family 55 β-1,3-Glucanase from the Basidiomycete Phanerochaete Chrysosporium.** *Biotechnology Letters* 2006, **28(6)**:365-371.
32. Roos S, Jonsson H: **A high-molecular-mass cell-surface protein from Lactobacillus reuteri 1063 adheres to mucus components.** *Microbiology* 2002, **148(Pt 2)**:433-442.
33. Quadri LE: **Regulation of antimicrobial peptide production by autoinducer-mediated quorum sensing in lactic acid bacteria.** *Antonie Van Leeuwenhoek* 2002, **82(1-4)**:133-145.
34. Raimann E, Schmid B, Stephan R, Tasara T: **The alternative sigma factor sigma(L) of L. monocytogenes promotes growth under diverse environmental stresses.** *Foodborne Pathog Dis* 2009, **6(5)**:583-591.
35. Vesterlund S, Karp M, Salminen S, Ouwehand AC: **Staphylococcus aureus adheres to human intestinal mucus but can be displaced by certain lactic acid bacteria.** *Microbiology* 2006, **152(Pt 6)**:1819-1826.
36. Vesterlund S, Paltta J, Karp M, Ouwehand AC: **Adhesion of bacteria to resected human colonic tissue: quantitative analysis of bacterial adhesion and viability.** *Res Microbiol* 2005, **156(2)**:238-244.
37. Kovalenko NK, Podgorskii VS, Kasumova SA: **[Adhesion of lactic acid bacteria to epithelium of different cavities of the human organism].** *Mikrobiol Z* 2004, **66(4)**:62-68.
38. Zhou M, Boekhorst J, Francke C, Siezen R: **LocateP: Genome-scale subcellular-location predictor for bacterial proteins.** *BMC Bioinformatics* 2008, **9(1)**:173.
39. Sprenger J, Fink JL, Teasdale RD: **Evaluation and comparison of mammalian subcellular localization prediction methods.** *BMC Bioinformatics* 2006, **7(Suppl 5)**:S3.
40. Nakai K, Horton P: **Computational prediction of subcellular localization.** *Methods Mol Biol* 2007, **390**:429-466.
41. Chou KC, Shen HB: **Large-scale predictions of gram-negative bacterial protein subcellular locations.** *J Proteome Res* 2006, **5(12)**:3420-3428.
42. Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA: **Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates.** *Genome Res* 2009, **19(8)**:1404-1418.
43. Makarova KS, Koonin EV: **Evolutionary genomics of lactic acid bacteria.** *J Bacteriol* 2007, **189(4)**:1199-1208.
44. Masson S, Kern T, Le Gouellec A, Giustini C, Simorre JP, Callow P, Vernet T, Gabel F, Zapun A: **Central domain of DivIB caps the C-terminal regions of the FtsL/DivIC coiled-coil rod.** *J Biol Chem* 2009, **284(40)**:27687-27700.
45. Le Gouellec A, Roux L, Fadda D, Massidda O, Vernet T, Zapun A: **Roles of pneumococcal DivIB in cell division.** *J Bacteriol* 2008, **190(13)**:4501-4511.
46. Bennett JA, Aimino RM, McCormick JR: **Streptomyces coelicolor genes ftsL and divIC play a role in cell division but are dispensable for colony formation.** *J Bacteriol* 2007, **189(24)**:8982-8992.
47. Noirclerc-Savoye M, Le Gouellec A, Morlot C, Dideberg O, Vernet T, Zapun A: **In vitro reconstitution of a trimeric complex of DivIB, DivIC and FtsL, and their transient co-localization at the division site in Streptococcus pneumoniae.** *Mol Microbiol* 2005, **55(2)**:413-424.
48. Robson SA, Michie KA, Mackay JP, Harry E, King GF: **The Bacillus subtilis cell division proteins FtsL and DivIC are intrinsically unstable and do not interact with one another in the absence of other septasomal components.** *Mol Microbiol* 2002, **44(3)**:663-674.
49. Sievers J, Errington J: **The Bacillus subtilis cell division protein FtsL localizes to sites of septation and interacts with DivIC.** *Mol Microbiol* 2000, **36(4)**:846-855.
50. Katis VL, Wake RG: **Membrane-bound division proteins DivIB and DivIC of Bacillus subtilis function solely through their external domains in both vegetative and sporulation division.** *J Bacteriol* 1999, **181(9)**:2710-2718.
51. Daniel RA, Harry EJ, Katis VL, Wake RG, Errington J: **Characterization of the essential cell division gene ftsL(yIID) of Bacillus subtilis and its role in the assembly of the division apparatus.** *Mol Microbiol* 1998, **29(2)**:593-604.
52. Katis VL, Harry EJ, Wake RG: **The Bacillus subtilis division protein DivIC is a highly abundant membrane-bound protein that localizes to the division site.** *Mol Microbiol* 1997, **26(5)**:1047-1055.
53. Levin PA, Losick R: **Characterization of a cell division gene from Bacillus subtilis that is required for vegetative and sporulation septum formation.** *J Bacteriol* 1994, **176(5)**:1451-1459.
54. Siezen R, Boekhorst J, Muscariello L, Molenaar D, Renckens B, Kleerebezem M: **Lactobacillus plantarum gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria.** *BMC Genomics* 2006, **7**:126.

55. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
56. Berger B, Pridmore RD, Barretto C, Delmas-Julien F, Schreiber K, Arigoni F, Brüssow H: **Similarity and Differences in the Lactobacillus acidophilus Group Identified by Polyphasic Analysis and Comparative Genomics⊠, †.** *J Bacteriol* 2007, **189(4)**:1311-1321.
57. Canchaya C, Claesson MJ, Fitzgerald GF, van Sinderen D, O'Toole PW: **Diversity of the genus Lactobacillus revealed by comparative genomics of five species.** *Microbiology* 2006, **152(Pt 11)**:3185-3196.
58. Claesson MJ, van Sinderen D, O'Toole PW: **The genus Lactobacillus–a genomic basis for understanding its diversity.** *FEMS Microbiol Lett* 2007, **269(1)**:22-28.
59. Ventura M, Jankovic I, Walker DC, Pridmore RD, Zink R: **Identification and Characterization of Novel Surface Proteins in Lactobacillus johnsonii and Lactobacillus gasseri.** *Appl Environ Microbiol* 2002, **68(12)**:6172-6181.
60. Callanan M, Kaleta P, O'Callaghan J, O'Sullivan O, Jordan K, McAuliffe O, Sangrador-Vegas A, Slattery L, Fitzgerald GF, Beresford T, *et al*: **Genome sequence of Lactobacillus helveticus, an organism distinguished by selective gene loss and insertion sequence element expansion.** *J Bacteriol* 2008, **190(2)**:727-735.
61. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, *et al*: **Comparative genomics of the lactic acid bacteria.** *Proc Natl Acad Sci USA* 2006, **103(42)**:15611-15616.
62. Khaleghi M, Kermanshahi RK, Yaghoobi MM, Zarkesh-Esfahani SH, Baghizadeh A: **Assessment of Bile Salt Effects on S-Layer Production, slp Gene Expression and Some Physicochemical Properties of Lactobacillus acidophilus ATCC 4356.** *J Microbiol Biotechnol* 2010, **20(4)**:749-756.
63. Goh YJ, Azcarate-Peril MA, O'Flaherty S, Durmaz E, Valence F, Jardin J, Lortal S, Klaenhammer TR: **Development and application of a upp-based counterselective gene replacement system for the study of the S-layer protein SlpX of Lactobacillus acidophilus NCFM.** *Appl Environ Microbiol* 2009, **75(10)**:3093-3105.
64. Buck BL, Altermann E, Svingerud T, Klaenhammer TR: **Functional analysis of putative adhesion factors in Lactobacillus acidophilus NCFM.** *Appl Environ Microbiol* 2005, **71(12)**:8344-8351.
65. Avall-Jaaskelainen S, Hynonen U, Ilk N, Pum D, Sleytr UB, Palva A: **Identification and characterization of domains responsible for self-assembly and cell wall binding of the surface layer protein of Lactobacillus brevis ATCC 8287.** *BMC Microbiol* 2008, **8**:165.
66. Prado Acosta M, Mercedes Palomino M, Allievi MC, Sanchez Rivas C, Ruzal SM: **Murein hydrolase activity in the surface layer of Lactobacillus acidophilus ATCC 4356.** *Appl Environ Microbiol* 2008, **74(24)**:7824-7827.
67. Konstantinov SR, Smidt H, de Vos WM, Bruijns SC, Singh SK, Valence F, Molle D, Lortal S, Altermann E, Klaenhammer TR, *et al*: **S layer protein A of Lactobacillus acidophilus NCFM regulates immature dendritic cell and T cell functions.** *Proc Natl Acad Sci USA* 2008, **105(49)**:19474-19479.
68. Chen X, Xu J, Shuai J, Chen J, Zhang Z, Fang W: **The S-layer proteins of Lactobacillus crispatus strain ZJ001 is responsible for competitive exclusion against Escherichia coli O157:H7 and Salmonella typhimurium.** *Int J Food Microbiol* 2007, **115(3)**:307-312.
69. Sanders ME, Klaenhammer TR: **Invited review: the scientific basis of Lactobacillus acidophilus NCFM functionality as a probiotic.** *J Dairy Sci* 2001, **84(2)**:319-331.
70. Azcarate-Peril MA, Altermann E, Goh YJ, Tallon R, Sanozky-Dawes RB, Pfeiler EA, O'Flaherty S, Buck BL, Dobson A, Duong T, *et al*: **Analysis of the genome sequence of Lactobacillus gasseri ATCC 33323 reveals the molecular basis of an autochthonous intestinal organism.** *Appl Environ Microbiol* 2008, **74(15)**:4610-4625.
71. Boekhorst J, Helmer Q, Kleerebezem M, Siezen RJ: **Comparative analysis of proteins with a mucus-binding domain found exclusively in lactic acid bacteria.** *Microbiology* 2006, **152(Pt 1)**:273-280.
72. van Pijkeren JP, Canchaya C, Ryan KA, Li Y, Claesson MJ, Sheil B, Steidler L, O'Mahony L, Fitzgerald GF, van Sinderen D, *et al*: **Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of Lactobacillus salivarius UCC118.** *Appl Environ Microbiol* 2006, **72(6)**:4143-4153.
73. MacKenzie DA, Tailford LE, Hemmings AM, Juge N: **Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity.** *J Biol Chem* 2009, **284(47)**:32444-32453.
74. Gendrot F, Foucaud-Scheunemann C, Ferchichi M, Hemme D: **Characterization of amino acid transport in the dairy strain Leuconostoc mesenteroides subsp. mesenteroides CNRZ 1273.** *Lett Appl Microbiol* 2002, **35(4)**:291-295.
75. Dudley E, Steele J: **Lactococcus lactis LM0230 contains a single aminotransferase involved in aspartate biosynthesis, which is essential for growth in milk.** *Microbiology* 2001, **147(Pt 1)**:215-224.
76. Juillard V, Guillot A, Le Bars D, Gripon JC: **Specificity of milk peptide utilization by Lactococcus lactis.** *Appl Environ Microbiol* 1998, **64(4)**:1230-1236.
77. Mierau I, Kunji ER, Leenhouts KJ, Hellendoorn MA, Haandrikman AJ, Poolman B, Konings WN, Venema G, Kok J: **Multiple-peptidase mutants of Lactococcus lactis are severely impaired in their ability to grow in milk.** *J Bacteriol* 1996, **178(10)**:2794-2803.
78. Florez AB, Delgado S, Mayo B: **Antimicrobial susceptibility of lactic acid bacteria isolated from a cheese environment.** *Can J Microbiol* 2005, **51(1)**:51-58.
79. Chirica LC, Guray T, Gurakan GC, Bozoglu TF: **Characterization of extracellular beta-lactamases from penicillin G-resistant cells of Streptococcus thermophilus.** *J Food Prot* 1998, **61(7)**:896-898.
80. Coque JJ, Liras P, Martin JF: **Genes for a beta-lactamase, a penicillin-binding protein and a transmembrane protein are clustered with the cephamycin biosynthetic genes in Nocardia lactamdurans.** *Embo J* 1993, **12(2)**:631-639.
81. Yamamura A, Okada A, Kameda Y, Ohtsuka J, Nakagawa N, Ebihara A, Nagata K, Tanokura M: **Structure of TTHA1623, a novel metallo-beta-lactamase superfamily protein from Thermus thermophilus HB8.** *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2009, **65(Pt 5)**:455-459.
82. Korycka-Dahl M, Richardson T, Bradley RL Jr: **Use of microbial beta-lactamase to destroy penicillin added to milk.** *J Dairy Sci* 1985, **68(8)**:1910-1916.
83. Kao SM, Olmsted SB, Viksnins AS, Gallo JC, Dunny GM: **Molecular and genetic analysis of a region of plasmid pCF10 containing positive control genes and structural genes encoding surface proteins involved in pheromone-inducible conjugation in Enterococcus faecalis.** *J Bacteriol* 1991, **173(23)**:7650-7664.
84. Ruhfel RE, Manias DA, Dunny GM: **Cloning and characterization of a region of the Enterococcus faecalis conjugative plasmid, pCF10, encoding a sex pheromone-binding function.** *J Bacteriol* 1993, **175(16)**:5253-5259.
85. Hirt H, Manias DA, Bryan EM, Klein JR, Marklund JK, Staddon JH, Paustian ML, Kapur V, Dunny GM: **Characterization of the pheromone response of the Enterococcus faecalis conjugative plasmid pCF10: complete sequence and comparative analysis of the transcriptional and phenotypic responses of pCF10-containing cells to pheromone induction.** *J Bacteriol* 2005, **187(3)**:1044-1054.
86. Kirillova Iu M, Mikhailova EO, Balaban NP, Mardanova AM, Rudenskaia GN, Kostrov SV, Sharipova MR: **[Growth conditions and production of the Bacillus intermedius subtilisin-like serine proteinase by the recombinant Bacillus subtilis strain].** *Mikrobiologiia* 2006, **75(2)**:172-178.
87. Mikhailova EO, Mardanova AM, Balaban NP, Rudenskaya GN, Ilyinskaya ON, Sharipova MR: **Biochemical properties of Bacillus intermedius subtilisin-like proteinase secreted by a Bacillus subtilis recombinant strain in its stationary phase of growth.** *Biochemistry (Mosc)* 2009, **74(3)**:308-315.
88. Siezen RJ, Bruinenberg PG, Vos P, van Alen-Boerrigter I, Nijhuis M, Alting AC, Exterkate FA, de Vos WM: **Engineering of the substrate-binding region of the subtilisin-like, cell-envelope proteinase of Lactococcus lactis.** *Protein Eng* 1993, **6(8)**:927-937.
89. Srivastava R, Liu JX, Howell SH: **Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in Arabidopsis.** *Plant J* 2008, **56(2)**:219-227.
90. Goh YJ, Klaenhammer TR: **Functional roles of aggregation-promoting-like factor in stress tolerance and adherence of Lactobacillus acidophilus NCFM.** *Appl Environ Microbiol* 2010.
91. Jankovic I, Ventura M, Meylan V, Rouvet M, Elli M, Zink R: **Contribution of aggregation-promoting factor to maintenance of cell shape in Lactobacillus gasseri 4B2.** *J Bacteriol* 2003, **185(11)**:3288-3296.
92. Marcotte H, Ferrari S, Cesena C, Hammarstrom L, Morelli L, Pozzi G, Oggioni MR: **The aggregation-promoting factor of Lactobacillus crispatus M247 and its genetic locus.** *J Appl Microbiol* 2004, **97(4)**:749-756.
93. Schneewind O, Jones KF, Fischetti VA: **Sequence and structural characteristics of the trypsin-resistant T6 surface protein of group A streptococci.** *J Bacteriol* 1990, **172(6)**:3310-3317.

94. Fischetti VA, Pancholi V, Schneewind O: **Conservation of a hexapeptide sequence in the anchor region of surface proteins from gram-positive cocci.** *Mol Microbiol* 1990, **4(9)**:1603-1605.
95. Kleerebezem M, Hols P, Bernard E, Rolain T, Zhou M, Siezen RJ, Bron PA: **The extracellular biology of the lactobacilli.** *FEMS Microbiol Rev* 2010, **34(2)**:199-230.
96. Boekhorst J, de Been MW, Kleerebezem M, Siezen RJ: **Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs.** *J Bacteriol* 2005, **187(14)**:4928-4934.
97. Bae T, Schneewind O: **The YSIRK-G/S motif of staphylococcal protein A and its role in efficiency of signal peptide processing.** *J Bacteriol* 2003, **185(9)**:2910-2919.
98. DeDent A, Bae T, Missiakas DM, Schneewind O: **Signal peptides direct surface proteins to two distinct envelope locations of Staphylococcus aureus.** *Embo J* 2008, **27(20)**:2656-2668.
99. Zhang Z, Liu C, Zhu Y, Zhong Y, Zhu Y, Zheng H, Zhao GP, Wang SY, Guo X: **Complete genome sequence of Lactobacillus plantarum JDM1.** *J Bacteriol* 2009.
100. Boekhorst J, Siezen RJ, Zwahlen MC, Vilanova D, Pridmore RD, Mercenier A, Kleerebezem M, de Vos WM, Brussow H, Desiere F: **The complete genomes of Lactobacillus plantarum and Lactobacillus johnsonii reveal extensive differences in chromosome organization and gene content.** *Microbiology* 2004, **150(Pt 11)**:3601-3611.
101. Savijoki K, Kahala M, Palva A: **High level heterologous protein production in Lactococcus and Lactobacillus using a new secretion system based on the Lactobacillus brevis S-layer signals.** *Gene* 1997, **186(2)**:255-262.
102. De Angelis M, de Candia S, Calasso MP, Faccia M, Guinee TP, Simonetti MC, Gobbetti M: **Selection and use of autochthonous multiple strain cultures for the manufacture of high-moisture traditional Mozzarella cheese.** *Int J Food Microbiol* 2008, **125(2)**:123-132.
103. Pastink MI, Teusink B, Hols P, Visser S, de Vos WM, Hugenholtz J: **Genome-scale model of Streptococcus thermophilus LMG18311 for metabolic comparison of lactic acid bacteria.** *Appl Environ Microbiol* 2009, **75(11)**:3627-3633.
104. Wegmann U, O'Connell-Motherway M, Zomer A, Buist G, Shearman C, Canchaya C, Ventura M, Goesmann A, Gasson MJ, Kuipers OP, *et al*: **Complete genome sequence of the prototype lactic acid bacterium Lactococcus lactis subsp. cremoris MG1363.** *J Bacteriol* 2007, **189(8)**:3256-3270.
105. Le Bourgeois P, Mata M, Ritzenthaler P: **Genome comparison of Lactococcus strains by pulsed-field gel electrophoresis.** *FEMS Microbiol Lett* 1989, **50(1-2)**:65-69.
106. Petrova P, Emanuilova M, Petrov K: **Amylolytic Lactobacillus strains from Bulgarian fermented beverage boza.** *Z Naturforsch C* 2010, **65(3-4)**:218-224.
107. Liu M, Bayjanov JR, Renckens B, Nauta A, Siezen RJ: **The proteolytic system of lactic acid bacteria revisited: a genomic comparison.** *BMC Genomics* 2010, **11**:36.
108. Bounaix MS, Gabriel V, Morel S, Robert H, Rabier P, Remaud-Simeon M, Gabriel B, Fontagne-Faucher C: **Biodiversity of exopolysaccharides produced from sucrose by sourdough lactic acid bacteria.** *J Agric Food Chem* 2009, **57(22)**:10889-10897.
109. Cobo Molinos A, Abriouel H, Ben Omar N, Lopez RL, Galvez A: **Microbial diversity changes in soybean sprouts treated with enterocin AS-48.** *Food Microbiol* 2009, **26(8)**:922-926.
110. De Vuyst L, Vrancken G, Ravyts F, Rimaux T, Weckx S: **Biodiversity, ecological determinants, and metabolic exploitation of sourdough microbiota.** *Food Microbiol* 2009, **26(7)**:666-675.
111. Chao SH, Wu RJ, Watanabe K, Tsai YC: **Diversity of lactic acid bacteria in suan-tsai and fu-tsai, traditional fermented mustard products of Taiwan.** *Int J Food Microbiol* 2009, **135(3)**:203-210.
112. Roh SW, Kim KH, Nam YD, Chang HW, Park EJ, Bae JW: **Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing.** *ISME J* 2010, **4(1)**:1-16.
113. Ouadghiri M, Vancanneyt M, Vandamme P, Naser S, Gevers D, Lefebvre K, Swings J, Amar M: **Identification of lactic acid bacteria in Moroccan raw milk and traditionally fermented skimmed milk 'Iben'.** *J Appl Microbiol* 2009, **106(2)**:486-495.
114. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, *et al*: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2008, , **36 Database**: D13-21.
115. O'Brien KP, Remm M, Sonnhammer EL: **Inparanoid: a comprehensive database of eukaryotic orthologs.** *Nucleic Acids Res* 2005, , **33 Database**: D476-480.
116. Edgar R: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC Bioinformatics* 2004, **5(1)**:113.
117. Bailey TL, Williams N, Misleh C, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs.** *Nucl Acids Res* 2006, **34(suppl_2)**: W369-373.
118. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, *et al*: **Pfam: clans, web tools and services.** *Nucleic Acids Res* 2006, , **34 Database**: D247-251.
119. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, ollich V, Lassmann T, Moxon S, Marshall M, Khanna A, *et al*: **Pfam: clans, web tools and services.** *Nucl Acids Res* 2006, **34(suppl_1)**:D247-251.
120. Sonnhammer E, Eddy S, Birney E, Bateman A, Durbin R: **Pfam: multiple sequence alignments and HMM-profiles of protein domains.** *Nucl Acids Res* 1998, **26(1)**:320-322.
121. Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer EL: **Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins.** *Nucleic Acids Res* 1999, **27(1)**:260-262.
122. Boekhorst J, Wels M, Kleerebezem M, Siezen RJ: **The predicted secretome of Lactobacillus plantarum WCFS1 sheds light on interactions with its environment.** *Microbiology* 2006, **152(Pt 11)**:3175-3183.
123. Siezen R, Boekhorst J, Muscariello L, Molenaar D, Renckens B, Kleerebezem M: **Lactobacillus plantarum gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria.** *BMC Genomics* 2006, **7(1)**:126.
124. Kerkhoven R, van Enckevort FH, Boekhorst J, Molenaar D, Siezen RJ: **Visualization for genomics: the Microbial Genome Viewer.** *Bioinformatics* 2004, **20(11)**:1812-1814.
125. Eddy SR: **A Probabilistic Model of Local Sequence Alignment That Simplifies Statistical Significance Estimation.** *PLoS Comput Biol* 2008, **4(5)**: e1000069.
126. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25(1)**:25-29.
127. Snel B, Bork P, Huynen MA: **The identification of functional modules from the genomic association of genes.** *Proc Natl Acad Sci USA* 2002, **99(9)**:5890-5895.
128. Tatusov RL, Natale D, A n, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, *et al*: **The COG database: new developments in phylogenetic classification of proteins from complete genomes.** *Nucl Acids Res* 2001, **29(1)**:22-28.
129. Tatusov RL, Koonin EV, Lipman DJ: **A Genomic Perspective on Protein Families.** *Science* 1997, **278(5338)**:631-637.
130. Boekhorst J, Snel B: **Identification of homologs in insignificant blast hits by exploiting extrinsic gene properties.** *BMC Bioinformatics* 2007, **8**:356.
131. Bendtsen JD, Kiemer L, Fausboll A, Brunak S: **Non-classical protein secretion in bacteria.** *BMC Microbiol* 2005, **5**:58.
132. Hubert P, Sawma P, Duneau JP, Khao J, Henin J, Bagnard D, Sturgis J: **Single-spanning transmembrane domains in cell growth and cell-cell interactions: More than meets the eye?** *Cell Adh Migr* 2010, **4(2)**.

doi:10.1186/1471-2164-11-651
Cite this article as: Zhou *et al.*: LAB-Secretome: a genome-scale comparative analysis of the predicted extracellular and surface-associated proteins of Lactic Acid Bacteria. *BMC Genomics* 2010 **11**:651.