



**HAL**  
open science

# Etudes comparatives de *Lactobacillus delbrueckii* sous-espèces *lactis* et *bulgaricus* : identification des déterminants du phénotype anti-inflammatoire

Hela El Kafsi

► **To cite this version:**

Hela El Kafsi. Etudes comparatives de *Lactobacillus delbrueckii* sous-espèces *lactis* et *bulgaricus* : identification des déterminants du phénotype anti-inflammatoire. Sciences agricoles. Université Paris Sud - Paris XI, 2014. Français. NNT : 2014PA112067 . tel-01297045

**HAL Id: tel-01297045**

**<https://theses.hal.science/tel-01297045>**

Submitted on 2 Apr 2016

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

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



UNIVERSITE PARIS-SUD

ECOLE DOCTORALE : *ABIES*

Discipline : Biologie

Equipe Interactions des Firmicutes avec leurs environnements

THESE DE DOCTORAT

Soutenue le 10/04/2014

Par

**Hela EL KAFSI**

Etudes comparatives de *Lactobacillus delbrueckii* sous-espèces  
*lactis* et *bulgaricus*.

Identification des déterminants du phénotype anti-inflammatoire.

Directeur de thèse : Dr. Maarten VAN DE GUCHTE, INRA Jouy en Josas

Composition du jury :

Rapporteurs

Dr. Francois LEULIER, Institut de Génomique Fonctionnelle de Lyon

Dr. Pascal LE BOURGEOIS, Université Paul Sabatier, Toulouse

Examineurs

Pr. Armel GUYONVARCH, Université Paris-Sud 11

Pr. Annick MERCENIER, Nestlé Research Center, Lausanne

Dr. Michel-Yves MISTOU, INRA, Jouy en Josas

Dr. Emmanuelle MAGUIN, INRA, Jouy en Josas





## Remerciements

Voici la partie qui sera la plus difficile à écrire.....

Si je voulais raconter ma thèse, je ne saurais pas par où commencer....

Une belle aventure humaine de plus de 3 ans.

Beaucoup d'épreuves qui ont fait de moi ce que je suis aujourd'hui.

De belles amitiés, de beaux moments avec des gens qui feront à jamais partie de moi.

Une vraie famille qui a su me soutenir dans les moments les plus durs.

Des fous rires sur le chemin de Vauboyen, à la cantine, pendant les pauses café, pour chaque pot ou pique-nique organisés.

Je ne garde que le meilleur.....

Cette thèse je la dédie aux vrais amis que j'ai pu rencontrer, sans citer de nom car la liste sera longue..... Mais je sais qu'ils se reconnaîtront.

Je la dédie à mon frère, qui, a toujours été un exemple pour moi.

Je la dédie à mes tantes et à mes oncles qui ont toujours cru en moi.

Je la dédie à tous mes cousins et cousines, un grand merci d'avoir fait les plus beaux souvenirs de ma vie.

Je la dédie à mes deux grands-mères, j'espère vous rentre fières.

Je la dédie à mes amies de Tunis, qui malgré la distance, sont toujours restées présentes.

Je la dédie à toi, Mohamed, Salma, Nouha, Zakia, Sarra, Amine, Emna, Meha, Dali, Rim, Samhour, Katka,..... Merci pour votre compréhension à chaque fois que j'ai refusé de sortir pour pouvoir rédiger, merci de m'avoir gardé comme amie malgré mon insociabilité totale depuis des mois, merci aussi pour votre soutien inconditionné, votre amour est une force.

Je la dédie à mes parents, ceux dont l'amour ma bercé..... Ce travail est pour vous, le fruit de vos sacrifices.

Cette thèse est l'occasion pour moi de vous dire tout mon amour et ma reconnaissance.

Pour finir, je dédie ce travail à mon mari : A toi Walid, ces 11 années passées à tes côtés m'ont fait grandir. Cette thèse est aussi la tienne. Merci d'avoir été le père, la mère, le frère, l'ami avant d'être le mari. Merci pour ton aide inconditionnée, merci pour ton amour.

## Table des matières

I.	Interaction microbiote-hôte dans le tractus digestif, implications pour la santé humaine.....	9
I.A	Homéostasie intestinale.....	10
I.A.1.	Rôle de la barrière intestinale dans le maintien de l'homéostasie.....	10
I.A.2.	Rôle du microbiote dans le maintien de l'homéostasie intestinale.....	11
I.A.3.	Approches d'analyse de la biodiversité microbienne du tractus digestif .....	14
I.B	Dysbiose intestinale et maladies associées.....	17
I.B.1.	Implication du microbiote dans l'obésité.....	20
I.B.2.	Implication du microbiote dans le diabète de type 1 .....	21
I.B.3.	Implication du microbiote dans le diabète de type 2 .....	23
I.B.4.	Implication du microbiote intestinal dans les maladies inflammatoires chroniques de l'intestin .....	24
I.C	Maladies inflammatoires de l'intestin.....	25
I.C.1.	Etudes de la composition du microbiote intestinal comme marqueur des MICI .....	26
I.C.2.	Etiologie des maladies inflammatoires chroniques de l'intestin .....	28
I.C.3.	Thérapies contre les maladies inflammatoires chroniques de l'intestin .....	34
I.D	Interaction des bactéries commensales intestinales avec l'hôte et effets sur la santé.....	38
II.	Bactéries probiotiques et santé humaine.....	40
II.A	Généralité.....	40
II.B	Les bactéries probiotiques des genres Lactobacillus et Bifidobacteriu.....	41
II.C	Niche écologique, lactobacilles et bifidobactéries intestinau.....	42
II.D	Propriétés intrinsèques de résistance dans le tractus digestif : Lactobacillus et Bifidobacteriu.....	43
II.E	Effets anti-inflammatoires des bactéries probiotiques, lactobacilles et bifidobactérie.....	44
II.E.1.	Etudes cliniques et mise en évidence de l'effet anti-inflammatoire de certains lactobacilles et bifidobactéries .....	44
II.E.2.	Modes d'interaction de bactéries probiotiques avec l'hôte .....	45
II.E.3.	Effecteurs bactériens à l'origine des effets anti-inflammatoires de certains probiotiques .....	47
III.	Alimentation, bactéries alimentaires et santé humaine .....	50

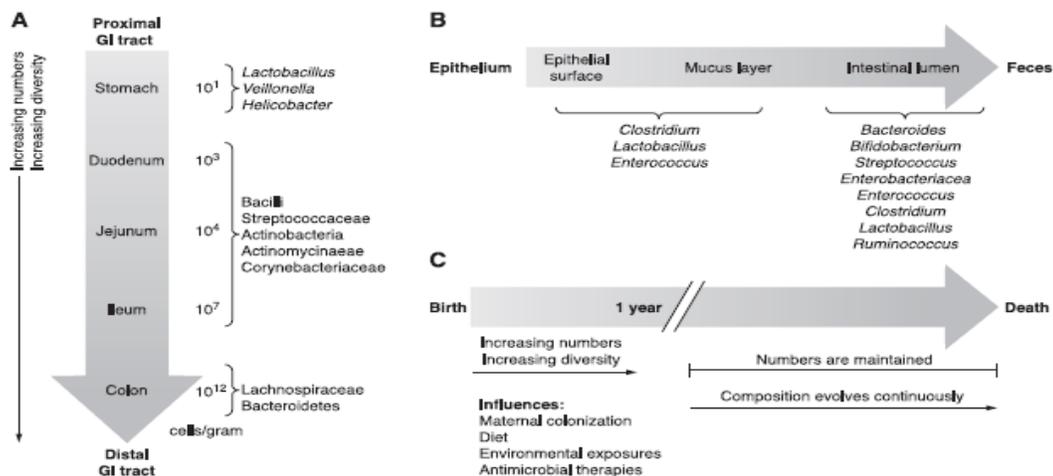
III.A	Alimentation et santé humaine.....	51
III.A.1.	Impact de l'alimentation sur le microbiote.....	51
III.A.2.	Aliments prébiotiques: impact sur les MICI .....	52
III.B	Rôle des produits fermentés et des bactéries lactiques dans la santé humain.....	53
III.B.1.	Produits fermentés du lait et apport en nutriments, l'exemple du yaourt .....	55
III.B.2.	Produits fermentés du lait et activités enzymatiques, l'exemple du yaourt .....	58
III.C	Adaptation et persistance des bactéries lactiques des produits fermentés dans le tractus digestif .....	59
III.D	Propriétés immuno-modulatoires des bactéries lactiques alimentaires.....	60
III.E	Effets anti-inflammatoires de <i>L. delbrueckii</i> .....	64
III.E.1.	Effets anti-inflammatoires de <i>L. delbrueckii</i> .....	66
IV.	Approches « omiques » pour l'identification d'effecteurs probiotiques .....	68
IV.A	Génomique comparative, corrélation entre phénotypes et présence ou absence de gènes.....	69
IV.A.1.	Génomique comparative et identification d'effecteurs bactériens potentiellement à l'origine des effets probiotiques .....	69
IV.B	Transcriptomique comparative, corrélation entre phénotype probiotique et expression de gènes .....	71
IV.B.1.	Transcriptomique comparative et identification d'effecteurs bactériens potentiellement à l'origine des effets probiotiques .....	72
IV.C	Protéomique comparative et corrélation entre phénotype probiotique et protéome.....	74
IV.C.1.	Protéomique comparative et identification d'effecteurs bactériens potentiellement à l'origine des effets probiotiques .....	74
V.	Résultats.....	77
V.A	Genome sequence of <i>Lactobacillus delbrueckii</i> ssp. <i>lactis</i> CNRZ327, a dairy bacterium with anti-inflammatory properties.....	78
V.B	<i>Lactobacillus delbrueckii</i> ssp. <i>lactis</i> and ssp. <i>bulgaricus</i> : a chronicle of evolution in action...84	
V.C	Omics approaches to characterize anti-inflammatory properties of <i>Lactobacillus delbrueckii</i> .....	131
V.D	<i>Lactobacillus delbrueckii</i> ssp. <i>bulgaricus</i> and ssp. <i>lactis</i> : two subspecies, two evolution histories.....	164
VI.	Discussion générale et perspectives .....	189
VII.	Références Bibliographiques .....	196

## Liste des Principales abréviations

MICI: maladies inflammatoires chroniques des intestins  
TLR: toll-like receptors  
NLR: NOD-like receptors  
NOD-2: Nucleotide-binding oligomerization domain-containing protein 2  
MRR: Microbial recognition receptors (MRR)  
MAMP: Microbe associated molecular patterns  
PAMP: Pathogen-associated molecular patterns  
PRRs: Pattern recognition receptors  
IL- : Interleukine  
RLRs: retinoic acid-inducible gene I (RIG-I) like receptors  
MC : maladie de Crohn  
CU : colite ulcéreuse  
TNF- $\alpha$ : Tumor necrosis factor  $\alpha$   
NF- $\kappa$ B: Nuclear factor kappa-light-chain-enhancer of activated B cells  
Ssp. : sous-espèce  
PCR : Polymerase Chain Reaction  
LPS: lipopolysaccharides  
TNBS: 2,4,6-trinitrobenzene sulfonic acid  
DSS: Dextran Sulfate Sodium  
PPAR- $\gamma$ : Peroxisome proliferator-activated receptor gamma  
PBMC: Peripheral Blood Mononuclear Cell  
HT-29: Human colon adenocarcinoma grade II cell line  
CWP: cell wall peptidoglycan  
LTA: lipoteichoic acid  
WTA: Wall teichoic acid  
SDP: Sortase dependant protein  
MP: Membrane protein  
MAP: Membrane associated protein  
LPP: Lipoproteins  
SLP: surface layer protein  
EPS : exopolysaccharides.  
IS : séquence d'insertion  
Mub : mucus binding protein  
MUB domain : mucus binding protein domain.  
ARNm : ARN messenger

## Préface

Le tractus digestif constitue un environnement complexe et dynamique. Sa complexité réside entre autre dans le fait qu'il renferme un « organe », dont nous ignorons presque tout. Cet organe appelé microbiote intestinal est constitué de l'ensemble des microorganismes qui habitent cet environnement spécifique [1]. La majorité du microbiote est composée de bactéries anaérobies qui dominant largement les bactéries anaérobies facultatives et les bactéries aérobies. Le microbiote intestinal varie en composition et nombre de microbes le long du tractus digestif (Figure 1 A). Cette variation dans la composition du microbiote est aussi observée sur le plan longitudinal, entre la surface des cellules épithéliales intestinales, le mucus et la lumière intestinale (Figure 1 B). La composition de cette communauté microbienne est sensible à des facteurs environnementaux tels que l'alimentation, l'état physiologiques de l'hôte et d'éventuels traitements par exemple, la consommation d'antibiotiques (Figure 1 C) [2].



**Figure 1. Les aspects spatiaux et temporels de la composition du microbiote intestinal (d'après [2]).**

A : Variation en nombre et composition du microbiote le long du tractus digestif. Un nombre relativement faible en bactéries est observé dans la partie supérieure du tractus digestif, alors que dans la partie inférieure, la population bactérienne avoisine les  $10^{11}$  -  $10^{12}$  cellules par gramme, la densité la plus élevée observée pour les habitats microbiens connus à ce jour. B : Variation longitudinale de la composition du microbiote intestinal. Les espèces bactériennes montrent une association à long terme avec un habitat particulier (espèces autochtones ou résidentes) ou sont en transit et ne colonisent pas le tractus digestif en condition normale. C : Etablissement et maintien des communautés microbiennes dans le temps, ainsi que certains facteurs influençant leur composition. Le microbiote d'un nourrisson évolue après la naissance pour atteindre après environ 1 an une composition se rapprochant de celle d'un microbiote

d'adulte. Des facteurs environnementaux sont des déterminants clés de la structure du microbiote intestinal chez l'adulte.

Le microbiote intestinal a un rôle métabolique important par exemple, le métabolisme de sucres complexes de lipides et de sels biliaires [3], la production de vitamines (vitamine K, vitamine B12 et acide folique) [2] mais il joue également un rôle important dans le développement post-natal de l'intestin, la physiologie de l'hôte [4], ainsi que dans le développement et la modulation du système immunitaire [5–7]. Par conséquent, un certain équilibre dans la composition du microbiote est important pour assurer un état d'homéostasie du système digestif au sens large en prenant en compte aliments, hôte et microbiote. Des modifications dans le microbiote intestinal sont associées à de nombreuses pathologies humaines par exemple, les maladies inflammatoires de l'intestin comme la maladie de Crohn et la colite ulcéreuse [8].

La diversité des espèces microbiennes et en conséquence des gènes microbiens au niveau intestinal dépendent en partie des aliments ingérés par l'individu [9]. De plus, certaines bactéries apportées par les aliments pourraient également avoir des effets fonctionnels. Citons pour exemple, les microorganismes impliquées dans la production d'aliments fermentés et les probiotiques dont certains possèdent des propriétés décrites dans la littérature comme inhibitrices de pathogènes (par exemple contre *Helicobacter pylori*), des effets bénéfiques sur certaines maladies du tractus gastro-intestinal (cancer du colon, maladies diarrhéiques, réactions allergiques), ou des effets anti-inflammatoires (effets observés dans le contexte de la maladie de Crohn et colite ulcéreuse) [10].

## **I. Interaction microbiote-hôte dans le tractus digestif, implications pour la santé humaine**

## ***I.A Homéostasie intestinale***

L'homéostasie intestinale est un état d'équilibre dans l'intestin qui repose sur des interactions complexes entre le microbiote et l'hôte par exemple l'épithélium intestinal et le système immunitaire. Divers mécanismes de régulation coopèrent pour maintenir l'homéostasie intestinale. La vision actuelle est que leur dérégulation précipiterait l'apparition de pathologies chroniques comme les maladies inflammatoires chroniques des intestins (MICI).

### **I.A.1. Rôle de la barrière intestinale dans le maintien de l'homéostasie**

La barrière intestinale constitue une première ligne de défense contre les pathogènes qui peuvent se présenter dans la lumière intestinale. Cette barrière compte plusieurs niveaux de défense afin de limiter la translocation de pathogènes de la lumière intestinale vers la lamina propria:

Les cellules épithéliales constituent une monocouche de cellules, qui en plus de leur fonction dans la digestion et l'absorption d'éléments nutritifs, forment une barrière physique entre le contenu de la lumière intestinale et la lamina propria. Des jonctions apicales entre les cellules dites jonctions serrées, contrôlent le passage para-cellulaire des bactéries à travers l'épithélium. A partir des cellules souches situées à la base des cryptes, 4 principaux types de cellules épithéliales intestinales se différencient. Dans l'intestin grêle, les cellules se différencient en :

- cellules absorbantes qui représentent plus de 80% des cellules épithéliales,
- cellules caliciformes (goblet cells) qui produisent du mucus,
- cellules entéro-endocrines sécrétant des hormones lors de leur migration ascendante le long de la villosité,
- cellules de Paneth qui produisent des agents antimicrobiens et des facteurs de croissance.

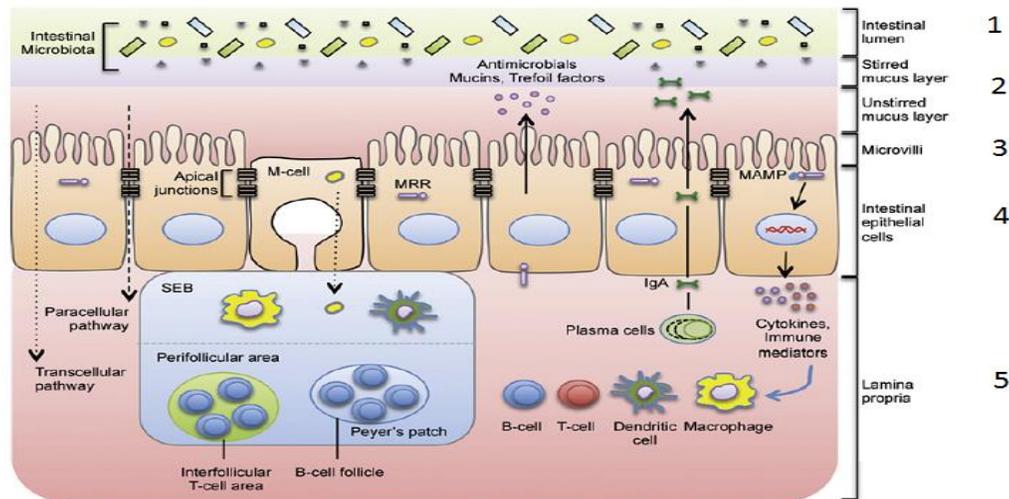
Contrairement à l'intestin grêle, le colon est constitué d'un épithélium de surface sans villosité, et ne contient pas de cellules de Paneth. Les cellules épithéliales intestinales sont aussi des cellules

présentatrices d'antigènes non-spécialisées médiant la réponse immunitaire. Pour finir, les cellules épithéliales intestinales perçoivent le contenu de la lumière intestinale, incluant les bactéries, par des récepteurs de reconnaissance de motifs moléculaires, comme les récepteurs TLR (toll-like receptors) et les récepteurs NLR (NOD-like receptors). Ces interactions induisent la sécrétion de médiateurs de la réponse immunitaire innée comme les cytokines ou les chimiokines qui agissent sur des cellules immunitaires présentes dans la lamina propria. Ces interactions activent les défenses de l'hôte en cas de danger, tout en prévenant une inflammation excessive [11].

Les cellules épithéliales intestinales agissent aussi comme barrière chimique en produisant un gel de mucines appelé mucus qui sert de barrière physique aux bactéries, contient des immunoglobulines IgA sécrétées et constitue un réservoir de peptides antimicrobiens comme les défensines. Le mucus est constitué de deux couches. Une première couche interne (proche de l'épithélium) riche en mucines et en peptides antimicrobiens, et une deuxième couche externe (proche du lumen) qui est comme la couche interne riche en mucines et peptides anti-microbiens mais qui constitue un micro-habitat pour des bactéries du microbiote intestinal.

### **I.A.2. Rôle du microbiote dans le maintien de l'homéostasie intestinale**

Parmi les populations microbiennes habitant les différentes parties du corps humain, celle du tractus digestif est la plus dense avec plus de 70 % des microbes du corps humain localisé dans le colon [2]. La surface intestinale estimée à 200 m<sup>2</sup>, représente une surface importante pour la colonisation



**Figure 2. Maintien de l'homéostasie intestinale par la barrière intestinale** (d'après [11]).

La barrière intestinale compte plusieurs niveaux de défense permettant de limiter la translocation d'antigènes de la lumière intestinale (1) vers la lamina propria (5). Celle-ci inclut : une couche de cellules épithéliales (4), reliées entre elles par des jonctions apicales. Ces jonctions assurent la régulation du passage para-cellulaire de molécules à travers l'épithélium. Les cellules épithéliales intestinales transportent des molécules de la lumière intestinale vers la lamina propria suivant une voie transcellulaire. Des cellules épithéliales spécialisées, appelées cellules M, assurent un échantillonnage continu de la lumière intestinale. Les microvillosités (3) sont les fins prolongements cellulaires localisés sur les cellules épithéliales. Elles se situent au pôle apical de la cellule et servent à absorber des substances. Les cellules caliciformes (« goblet cells »), les cellules de Paneth et les entérocytes sécrètent la mucine et des peptides antimicrobiens qui forment une couche de mucus (2). Les cellules épithéliales sécrètent l'immunoglobuline IgA, produite initialement par les plasmocytes (« plasma cells ») dans la lamina propria. Les cellules épithéliales présentent aussi des récepteurs de reconnaissance des microbes (« microbial recognition receptors » (MRR)) comme les récepteurs TLR et NOD-like, qui reconnaissent des motifs moléculaires associés aux microbes (« microbe associated molecular patterns » (MAMP)). La reconnaissance des microorganismes intestinaux par les cellules épithéliales induit la sécrétion de cytokines et autres médiateurs de la réponse immunitaire ce qui permet d'induire une réponse de régularisation contre le microbiote intestinal et de maintenir ainsi l'homéostasie intestinale.

microbienne. Le microbiote intestinal est composé majoritairement de bactéries strictement anaérobies, dominé principalement par deux phylums : les bacteroidetes et les firmicutes. L'estimation du nombre d'espèces bactériennes dans l'intestin humain varie de 500 à 1000 espèces suivant les études [2].

En plus de la population complexe de cellules immunes distribuées dans et autour de l'épithélium, le microbiote intestinal est un acteur majeur de l'homéostasie intestinale. En effet, pour maintenir la coexistence microbiote-hôte, l'hôte doit se protéger contre les invasions microbiennes et le microbiote doit assurer sa survie en se protégeant à la fois des bactéries compétitrices pathogènes et du système immunitaire humain. La capacité à maintenir cet équilibre repose en partie au dialogue

entre l'hôte humain et le microbiote intestinal. Pour le maintien de l'homéostasie intestinale, le microbiote agit au niveau de :

1- la couche de mucus, il contribue à l'homéostasie en régulant la production par l'hôte de peptides antimicrobiens ce qui pourrait limiter le nombre de bactéries à la surface muqueuse [12].

2- la muqueuse épithéliale, les cellules épithéliales, les cellules M ainsi que les cellules dendritiques transmettent des informations sur le microbiote à d'autres sous-ensembles de cellules immunitaires ce qui permet de contrôler directement le contenu intestinal. Les récepteurs PRRs (Pattern recognition receptors) sont des récepteurs exprimés par les cellules immunitaires. Ils sont considérés comme des «microbe-sensors» qui reconnaissent les motifs moléculaires associés aux pathogènes (PAMP). La plupart des récepteurs PRRs appartiennent à 3 familles principales : les récepteurs TLRs (toll-like receptors), les récepteurs NLRs (NOD-like receptors) ainsi que les récepteurs RLRs (retinoic acid-inducible gene I (RIG-I) like receptors [13]. Enfin les cytokines et chimiokines ainsi que les métabolites produits par l'hôte et par la microflore intestinale sont des médiateurs moléculaires clés de l'homéostasie intestinale qui influencent les réponses de l'hôte et du microbiote [13]. Le microbiote est aussi impliqué dans le développement de l'épithélium lui-même : en comparant le taux de prolifération des cellules épithéliales intestinales entre souris sans microbiote et souris conventionnelles, on observe qu'il est plus faible chez les souris sans microbiote [12].

3- différents composants immunitaires de la lamina propria avec lesquels il interagit: les IgA sont la classe majeure d'immunoglobulines produites par les plasmocytes et les cellules B dans la lamina propria. Les IgA sont sécrétées en grande quantité sous forme de polymères dans la lumière intestinale. Les IgA ainsi sécrétées (SIgA) couvrent les bactéries commensales afin d'inhiber leur adhésion à l'épithélium intestinal ainsi que leur pénétration dans la lamina propria. En plus de leur

interaction avec les IgA non spécifiques, les bactéries du microbiote peuvent induire une réponse IgA spécifique. En effet, les cellules dendritiques intestinales retiennent des bactéries commensales induisant la production d'IgA spécifiques assurant ainsi une protection contre la pénétration des bactéries commensales dans la muqueuse. Le microbiote intestinal est aussi capable de moduler le développement de cellules Th17. En effet, en l'absence du microbiote, le nombre de cellules Th17 est réduit significativement. Ces cellules qui sont abondantes dans les tissus intestinaux dans l'état d'homéostasie, protègent l'hôte de façon efficace contre l'exposition aux microbes intestinaux mais peuvent aussi jouer un rôle majeur dans l'induction de certaines maladies comme les maladies inflammatoires chroniques intestinales (MICI) [12].

Pour conclure, l'homéostasie intestinale repose sur le système immunitaire mais aussi les interactions complexes entre le microbiote intestinal et son hôte. La rupture de l'homéostasie conduirait non seulement à des maladies inflammatoires intestinales telles que la maladie de Crohn et la rectocolite hémorragique, mais contribuerait également à l'obésité, des maladies auto-immunes comme le diabète de type 1 ou des maladies métaboliques comme le diabète de type 2 [14].

### **I.A.3. Approches d'analyse de la biodiversité microbienne du tractus digestif**

Les premières études sur la composition du microbiote intestinal se sont basées sur l'utilisation de techniques de culture dans des milieux sélectifs, adaptés aux spécificités métaboliques de certaines sous-populations bactériennes.

Cette approche est peu résolutive ainsi la classification au niveau espèce est difficile, voire impossible. De plus, l'étude de la communauté microbienne de l'intestin par culture est fortement biaisée puisque jusqu'à 90% des bactéries du microbiote sont estimées non-cultivables [2].

A la fin des années 90, l'amplification par PCR et séquençage de gènes bactériens codant pour l'ARN 16S ont donné accès aussi bien aux bactéries cultivables qu'aux bactéries non-cultivables. Par cette approche, trois groupes phylogéniques majeurs ont été identifiés dans les selles humaines : le groupe des *Bacteroides*, le groupe des *Clostridium coccoïdes* et le sous-groupe des *Clostridium leptum* [2]. Seulement 24 % des espèces retrouvées correspondaient à des organismes déjà identifiés et ces bactéries appartiennent à la flore dominante par exemple *Bacteroides thetaiotaomicron*, *Fusobacterium prausnitzii* et *Eubacterium rectale*..

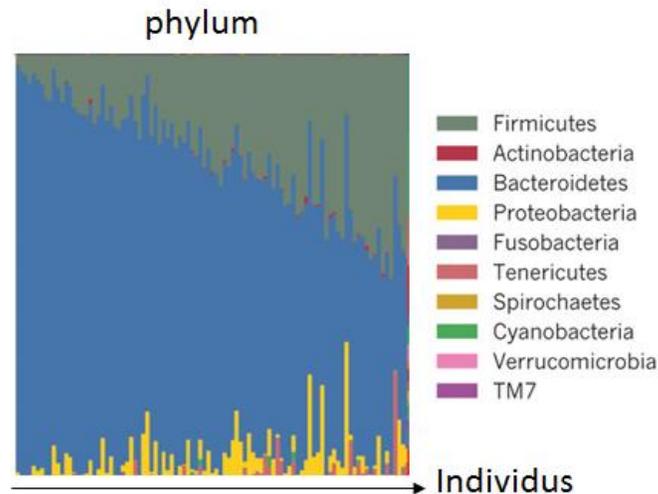
Depuis, les avancées technologiques en particulier en séquençage et bio-informatique ont permis de développer de nouvelles approches pour l'étude des  $10^{14}$  micro-organismes du tractus digestif ; chacune présente ses avantages et inconvénients (Tableau 1). Toutes ces approches ont permis d'approfondir la connaissance de la composition du microbiote habitant notre tractus digestif. Cependant, on dispose de très peu d'informations sur le rôle fonctionnel du microbiote.

Le séquençage métagénomique est un outil récent et puissant d'analyse [15–17]. Cette approche consiste en un séquençage aléatoire de l'ensemble de l'ADN d'un écosystème sans clonage préalable. Elle permet de générer des informations détaillées sur le « génome collectif » du microbiote. Ces informations peuvent être utilisées pour l'étude de la composition qualitative et quantitative du microbiote ainsi que la prédiction de ses propriétés par exemple, ses capacités métaboliques.

L'utilisation de cette technique sur un grand nombre d'échantillons fécaux humains a permis de générer un catalogue de 3.3 millions de gènes microbiens qui proviennent principalement de bactéries [1]. L'analyse qualitative du microbiote a confirmé sa complexité avec environ 160 espèces bactériennes par individu [1] celles-ci appartenant uniquement à une dizaine de phyla. La proportion de ces phyla varie considérablement suivant les individus [18].

Tableau 1. Techniques d'analyse du microbiote intestinal (d'après [2]).

Technique	16S rRNA Based?	Cost (\$)	Taxonomic Resolution/Sensitivity	Advantages	Disadvantages (Limitations)
Culture based	No	\$	Moderate	You have the organism "in hand" Functional information gained from what is known about the organism's substrate utilization and other physiological parameters	Most GI organisms cannot be cultured in current defined media Labor intensive
Full-length (Sanger) sequencing	Yes	\$\$\$\$	Very good	Sequencing the entire 16S gene maximizes the taxonomic resolution offered by the gene	Expensive (although mechanization is reducing the cost) Extensive bioinformatic analysis required
454 Pyrosequencing	Yes	\$\$\$	Good to very good	Higher throughput than Sanger sequencing (200,000 sequences versus 20,000) More sensitive (can detect less abundant organisms due to the number of reads obtained) Multiple samples can be analyzed in a single sequencing run No cloning bias introduced Less susceptible to PCR bias (shorter PCR amplicons, less influenced by G/C content)	Can't obtain the taxonomic resolution that can be achieved with full-length sequencing Shorter sequence reads (<500 bp versus 1.5 kb) Extensive bioinformatic analysis required
DGGE	Yes	\$	Poor	Rapid Fingerprints provide a good basis to compare communities from different treatment groups Bands of interest can be excised and sequenced	Shorter PCR products mean less taxonomic information can be obtained Reproducibility between gels is difficult
TRFLP	Yes	\$\$	Poor	Fingerprints provide a good basis to compare communities Multiple restriction enzymes can be used to provide greater resolution Reproducible	Limited taxonomic resolution One "phylotype" can represent more than one species
RISA	No	\$\$	Good	Greater variability between species and strains than the 16S gene When a better database has been developed, taxonomic resolution could be "Excellent"	Capillary sequencer required Limited phylogenetic data currently available; no extensive RIS database developed for gut bacteria A single bacterium can have more than one different RIS region
DNA microarrays	Yes	\$\$\$\$	Very good	Incredibly useful as a screening approach Fast, easy to use Clinical applications	Detection limited by the sequences contained on the chip (no detection of uncharacterized phylotypes) Cross-hybridization issues
FISH	Yes	\$\$\$	Good	Can target specific bacterial groups/species of interest (they must be preselected) Flexible scope: probes can be designed to target groups or individual species Direct enumeration of bacteria-16S copy number is not an issue	Can't identify novel groups of bacteria It's not a community-wide survey of "who's there" Reference strains are required to validate the results Microscope work is time-consuming (however FACS options are becoming available)
qPCR	Yes	\$\$	Good	Can target specific bacterial groups/species of interest (they must be preselected) Flexible scope: primers can be designed to target groups or individual species	Reference strains are required to validate the results Can't identify novel groups of bacteria It's not a community-wide survey of "who's there"
Metagenomics	Genome wide	\$\$\$\$\$	Good	Provides a community-wide assessment of the functional genes present 16S gene sequences provide taxonomic identification of community members Sensitivity depends on the number of sequence reads obtained	16S copy number varies between 1 and 10 Shotgun reads are mapped to reference genomes; this is limited by the number of sequenced genomes available Extensive bioinformatic analysis required Cloning biases could affect the functional gene information obtained No direct information about which genes are expressed or functioning



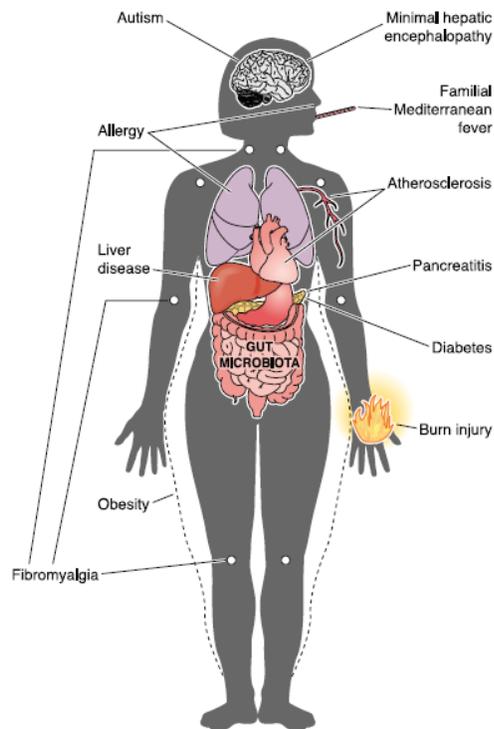
**Figure 3. Variations taxonomiques dans la composition du microbiote intestinal entre individus sains (adaptée de [18]).**

### ***I.B Dysbiose intestinale et maladies associées***

Le corps humain est désormais vu comme un super-organisme constitué de cellules humaines et de cellules microbiennes correspondant majoritairement à celles du microbiote intestinal (Figure 1). L'interaction du microbiote intestinal avec l'hôte et son environnement (généralement l'alimentation) influence la santé humaine. D'une part, le microbiote libère des toxines telles que les lipopolysaccharides et des métabolites bénéfiques comme les vitamines et les acides gras à courtes chaînes ce qui induit des effets néfastes ou bénéfiques pour la santé. D'autre part, l'hôte impose une pression sélective sur le microbiote à travers l'action de l'immunité innée ou à travers la disponibilité des nutriments. L'alimentation ainsi que les médicaments ont une importante capacité à influencer la composition ainsi que la fonction du microbiote intestinal [19].

Le rôle joué par le microbiote intestinal dans la santé humaine dépasse le cadre du tractus digestif et touche des organes plus éloignés (Figure 4). Ainsi, le microbiote intestinal a été décrit comme impliqué dans plusieurs maladies, celles qui touchent le tractus digestif comme les MICI mais aussi

celles qui touchent d'autres organes comme le diabète de type 1 et de type 2, l'obésité, l'autisme, les allergies, certaines maladies du foie, etc.



**Figure 4. Maladies associées au microbiote intestinal.** Le microbiote intestinal a été associé à plusieurs maladies touchant des organes éloignés. Les traits continus relient les organes et les maladies correspondantes. L'obésité est indiquée en pointillés. La fibromyalgie est illustrée par des points blancs correspondant aux points de pression caractéristiques utilisés pour diagnostiquer cette maladie (**d'après** [2]).

Pour pouvoir parler de corrélation entre pathologies et changements de composition du microbiote, il est essentiel de définir une composition de référence chez des individus sains. Dans ce but, des études comparatives ont été réalisées avec des groupes contrôles d'individus sains ne souffrant ni de maladie ni d'aucune forme d'infirmité afin de définir la composition du « core » commun du microbiote « sain ». Le Tableau 2 récapitule les espèces appartenant au « core » commun de plus de 50 individus sains.

**Tableau 2. La composition du core commun du microbiote d'individus sains (d'après [20]).**

<b>Known Core Genus/Species</b>	<b>Number of Species</b>
<i>Akkermansia muciniphila</i>	1
<i>Alistipes finegoldii</i>	1
<i>Anaerotruncus colihominis</i>	1
<i>Bacteroides</i> spp.	11
<i>Bifidobacterium</i> spp.	2
<i>Clostridium</i> spp.	13
<i>Colinsella</i> spp.	2
<i>Dorea</i> spp.	2
<i>Eubacterium</i> spp.	9
<i>Faecalibacterium prausnitzii</i>	1
<i>Lactobacillus lactis</i>	1
<i>Lachnobacterium</i> spp.	1
<i>Lachnospira pectinoschiza</i>	1
<i>Parabacteroides</i> spp.	3
<i>Prevotella</i> spp.	2
<i>Roseburia intestinalis</i>	1
<i>Ruminococcus</i> spp.	13
<i>Streptococcus</i> spp.	18
<i>Subdoligranulum variabile</i>	1
<i>Sutterella wadsworthia</i>	1
Uncultured Phylotypes	387

Les changements de composition du microbiote considérés comme des dysbioses et associés à des pathologies humaines ont été caractérisés dans plusieurs études. Les 10 pathologies les plus décrites comme étant corrélées à ces modifications sont indiquées dans le Tableau 3. Parmi ces maladies, les MICI (Maladies inflammatoires chroniques des intestins), la maladie de Crohn (MC) et la colite ulcéreuse (CU) sont associées à une diminution de la diversité du microbiote avec respectivement une diminution spécifique de *Faecalibacterium prausnitzii* et d'*Akkermansia muciniphila*. Les cas d'obésité sont caractérisés par un rapport spécifique entre les genres bacteroidetes et firmicutes. Le diabète de type 1 ainsi que le diabète de type 2 sont associés à certaines caractéristiques de composition du microbiote dites signatures cependant, ces signatures diffèrent d'une étude à l'autre. Pour finir, l'infection par *Clostridium difficile* est caractérisée en plus de la présence de ce pathogène, par une très importante diminution de la diversité bactérienne [20]. L'implication du

microbiote intestinal dans certaines de ces pathologies sera décrite plus en détail dans les paragraphes suivant.

**Tableau 3. Récapitulatif des études de corrélation entre composition du microbiote intestinal et certaines pathologies (d'après [20]).**

Aberration	Most relevant observations and potential correlation	References
Crohn's disease	Diversity decrease – reduced <i>F. prausnitzii</i>	Kaser et al. 2010 <sup>51</sup> ; Sokol et al. 2009 <sup>52</sup> ; Willing et al. 2010 <sup>53</sup>
Ulcerative colitis	Diversity decrease – reduced <i>A. muciniphila</i>	Png et al. 2010 <sup>54</sup> ; Kaser et al. 2010 <sup>51</sup> ; Lepage et al. 2011 <sup>55</sup>
Irritable bowel syndrome	Global signatures – increased <i>Dorea</i> and <i>Ruminococcus</i>	Salonen et al. 2010 <sup>36</sup> ; Saulnier et al. 2011 <sup>56</sup> ; Rajilić-Stojanović et al. 2011 <sup>13</sup>
<i>Clostridium difficile</i> infection	Strong diversity decrease – presence of <i>C. difficile</i>	Grehan et al. 2010 <sup>57</sup> ; Khoruts et al. 2010 <sup>58</sup>
Colorectal cancer	Variation in <i>Bacteroides</i> spp. – increased fusobacteria	Sobhani et al. 2011 <sup>59</sup> ; Wang et al. 2012 <sup>60</sup> ; Marchesi et al. 2011 <sup>61</sup>
Allergy/atopy	Altered diversity – specific signatures	Stsepetova et al. 2007 <sup>62</sup> ; Bisgaard et al. 2011 <sup>63</sup> ; Storrø et al. 2011 <sup>64</sup>
Celiac disease	Altered composition, notably in small intestine	Nistal et al. 2012 <sup>65</sup> ; Di Cagno et al. 2011 <sup>66</sup> ; Kalliomäki et al. 2012 <sup>67</sup>
Type 1 diabetes	Signature differences	Vaarela 2011 <sup>68</sup> ; Giongo et al. 2011 <sup>69</sup> ; Brown et al. 2011 <sup>70</sup>
Type 2 diabetes	Signature differences	Larssen et al. 2010 <sup>71</sup> ; Wu et al. 2010 <sup>72</sup> ; Kootte et al. 2012 <sup>73</sup>
Obesity	Specific bacterial ratios ( <i>Bacteroidetes/Firmicutes</i> )	Ley et al. 2006 <sup>74</sup> ; Turnbaugh et al. 2009 <sup>10</sup> ; Musso et al. 2011 <sup>75</sup>

### I.B.1. Implication du microbiote dans l'obésité

L'obésité est un problème sanitaire non seulement à l'échelle individuelle mais aussi à l'échelle de la famille et de la société. De plus, elle présente les caractéristiques d'une épidémie mondiale avec 400 millions d'adultes obèses en 2005, la prédiction de plus de 700 millions de cas en 2015 et une croissance qui prolongerait au moins jusqu'en 2030 [21].

En plus de la composante génétique qui expliquerait en partie la tendance de certains individus de devenir obèses [22], plusieurs études récentes ont démontré l'implication directe du microbiote intestinal. La composition du microbiote intestinal peut d'une part influencer l'efficacité énergétique des aliments [23–25] et d'autre part, des changements d'habitudes alimentaires conditionneraient des changements de composition de la population microbienne intestinale [26,27]. Plusieurs études ont montré que le microbiote intestinal est modifié chez les individus obèses par rapport aux individus minces même si les modifications décrites ne sont pas forcément les mêmes

d'une étude à l'autre. Certaines études décrivent une augmentation des firmicutes et une diminution des bacteroidetes chez les individus obèses par rapport aux non obèses [26, 28] mais l'observation inverse a aussi été rapportée [29]. Une augmentation des actinobactéries chez les individus obèses est aussi notée [30].

Une observation similaire est faite chez la souris : l'analyse du microbiote chez des souris obèses et non-obèses a révélé un changement du rapport entre firmicutes et bacteroidetes avec 50% de diminution des bacteroidetes et 50% d'augmentation des firmicutes chez les souris obèses [31].

Ces changements dans la composition du microbiote ne seraient pas simplement des conséquences de l'obésité mais auraient un rôle dans cette pathologie puisque le phénotype obèse est transmis par l'implantation du microbiote intestinal de souris obèses chez les souris initialement non-obèses. Ce résultat suggère fortement un rôle actif de la population microbienne de l'intestin dans cette pathologie [32].

Pour finir, une étude récente portant sur 292 Danois obèses (123 individus) ou non-obèses (169 individus) a montré l'existence de deux groupes d'individus avec des microbiotes de diversités élevées ou faibles et a révélé que la proportion d'individus obèses est plus importante dans le groupe ayant des microbiotes de faibles diversités. De plus, les individus obèses avec des microbiotes peu divers reprennent plus de poids dans la phase de stabilisation post-régime [33].

### **I.B.2. Implication du microbiote dans le diabète de type 1**

Le diabète de type 1 est une forme de diabète qui apparaît le plus souvent chez l'enfant ou chez le jeune adulte. Il s'agit d'une maladie auto-immune chronique entraînant une destruction totale par les cellules T, des cellules bêta des îlots de Langerhans du pancréas [14] qui sont normalement chargées du contrôle de la glycémie par la production d'insuline. Une revue récente décrit un trio de

facteurs induisant cette maladie auto-immune : un déséquilibre dans la composition du microbiote, une muqueuse intestinale perméable, ainsi qu'une dérégulation de la réponse immunitaire [34].

Le rôle joué par le microbiote dans l'apparition du diabète de type 1 a pu être étudié à travers différents modèles murins, quelques exemples sont présentés ci-dessous.

1) La prise orale de bactéries probiotiques (micro-organismes vivants qui lorsqu'ils sont administrés en quantité suffisante, ont des effets bénéfiques sur la santé de l'hôte au-delà de leurs valeurs nutritionnelles [35]) généralement des bactéries lactiques par des souris ou des rats prédisposés au diabète permet de retarder voir de prévenir l'apparition de la maladie [36–38].

2) La comparaison de la composition du microbiote intestinal entre rats diabétiques et non diabétiques montre qu'à l'apparition du diabète, la communauté bactérienne diffère significativement entre les deux groupes de rats. Au total 24 espèces bactériennes ont une abondance différente suivant le groupe considéré. Le microbiote des rats non diabétiques contient notamment beaucoup de bactéries des genres *Lactobacillus* et *Bifidobacterium*, comportant des souches probiotiques alors que le microbiote des rats diabétiques présente un nombre plus importants en *Bacteroides*, *Eubacterium* et *Ruminococcus* [39]. *Clostridium hylemonae* est également plus abondant dans le groupe des rats non diabétiques alors que 5 autres espèces du genre *Clostridium* sont plus abondantes dans le groupe des rats diabétiques. En plus des espèces connues, des centaines de groupes bactériens non classés en termes de genre, diffèrent par leur abondance.

3) Une souche de *Lactobacillus johnsonii* a pu être isolée des selles de rats non diabétiques [39] et son administration aux rats prédisposés au diabète a permis de prévenir l'apparition de la maladie [40].

Toutes ces observations suggèrent fortement que certaines bactéries du microbiote intestinal possèdent des effets protecteurs dans les modèles de rongeurs permettant de prévenir ou retarder l'apparition du diabète. La mise en évidence de l'implication du microbiote intestinal dans le diabète de type 1 dans des modèles de rongeurs a stimulé l'étude du microbiote intestinal de cohortes composées d'humains. L'analyse du microbiote intestinal d'enfants ayant une prédisposition génétique au diabète de type 1 a révélé des bactéries dont l'abondance est corrélée positivement ou négativement au développement de cette maladie auto-immune [41] : la proportion des bactéries du phylum bacteroidetes augmente au cours du temps chez les enfants malades alors que chez les enfants non malades, c'est la proportion des firmicutes qui augmente. Cette étude a permis de conclure que les enfants sains, qui ne présentent pas de risque génétique de diabète de type 1, possèdent un microbiote intestinal diversifié et stable au cours du temps. Les enfants prédisposés au développement de cette pathologie possèdent un microbiote moins diversifié et moins stable. Les bactéries négativement corrélées avec la maladie pourraient peut-être prévenir ou limiter le développement de l'auto-immunité chez les enfants à risque [41].

### **I.B.3. Implication du microbiote dans le diabète de type 2**

Le diabète de type 2 est une maladie métabolique dont la cause principale est la résistance à l'insuline reliée à l'obésité. Cette maladie qui survient généralement à l'âge adulte, est caractérisée par une faible inflammation chronique associée à la production anormale de plusieurs médiateurs inflammatoires comme le TNF- $\alpha$  (Tumor necrosis factor) [42].

Plusieurs études ont mis en évidence une relation entre la composition du microbiote intestinal et le diabète de type 2. A titre d'exemple, le niveau de *Bifidobacterium* est positivement corrélé avec l'amélioration de la tolérance au glucose et la baisse de l'inflammation chez la souris [43]. Chez l'homme, la comparaison du microbiote intestinal de 16 patients atteints de diabètes de type 2 et de

12 individus sains montre qu'il n'existe pas de différence significative en termes de diversité bactérienne entre les deux groupes d'individus [44]. Cependant, la similarité des populations bactériennes est plus importante entre les individus du même groupe qu'entre individus des deux groupes (sains et malades). Cette observation suggère que la composition du microbiote intestinal pourrait être corrélée avec l'état de maladie mais que les effectifs dans chaque groupe seraient trop faibles pour assurer la robustesse de l'analyse statistique. Parmi les différences notables entre microbiotes d'individus sains et malades figure la sous-représentation de *Bacteroides vulgatus* et du genre *Bifidobacterium* dans le microbiote du groupe de diabétiques [44]. L'étude récente menée sur 292 Danois montre que les individus dont le microbiote a une faible diversité bactérienne présentent des perturbations métaboliques associées à une augmentation du risque de pré-diabète et de diabète de type-2 [33].

Pour conclure, la diversité du microbiote et sa composition diffèrent entre individus tout venant et patients atteints de diabète de type 2. La dysbiose aurait un rôle dans l'apparition et/ou l'évolution de la pathologie.

#### **I.B.4. Implication du microbiote intestinal dans les maladies inflammatoires chroniques de l'intestin**

L'impact du microbiote sur la santé humaine est très étudié dans le contexte de maladies inflammatoires chroniques de l'intestin (MICI) telles que la maladie de Crohn et la colite ulcéreuse. Vu l'importance de ces pathologies en constante augmentation au niveau mondial et des résultats obtenus à travers les nombreuses études réalisées sur ce sujet, nous y consacrerons la section suivante.

## ***I.C Maladies inflammatoires de l'intestin***

Le terme « Maladies inflammatoires chroniques de l'intestin » est utilisé pour désigner certains états d'inflammation chronique à médiation immunitaire du tractus digestif. Parmi les maladies inflammatoires chroniques, les plus étudiées sont la maladie de Crohn (MC) et la colite ulcéreuse (CU), deux maladies dont les symptômes sont similaires mais qui affectent le tractus digestif de façon différente. La colite ulcéreuse induit une inflammation du colon et du rectum mais n'induit pas d'inflammation dans la partie supérieure du tractus gastro-intestinal alors que la maladie de Crohn peut aussi affecter la partie supérieure du tractus digestif ce qui peut induire une malabsorption des macro- et micro-nutriments. Ces deux maladies sont associées à un risque élevé de cancer du côlon [8].

Une muqueuse saine représente un premier mécanisme de défense contre la microflore intestinale et les pathogènes invasifs. Dans le cas d'une maladie inflammatoire chronique de l'intestin, la muqueuse endommagée permet le passage d'antigènes. La fonction protectrice de la muqueuse est réduite suite à un changement de composition et d'épaisseur de la couche de mucus ainsi qu'une modification de la glycosylation des glycoprotéines du mucus [45].

Ces changements de la muqueuse intestinale sont aussi associés à une baisse de sécrétion des IgA et une augmentation de celle des IgG qui conduit à une diminution de la protection vis-à-vis du microbiote et à l'établissement d'une réponse pro-inflammatoire. L'augmentation d'adhésion de bactéries pathogènes à la muqueuse endommagée entretient l'inflammation et augmente les dommages causés à la muqueuse intestinale [45]. La MC et la CU présentent aussi des variations caractéristiques dans la composition du microbiote [46]. Cependant, il n'est pas établi si ces variations sont une cause ou une conséquence de la maladie.

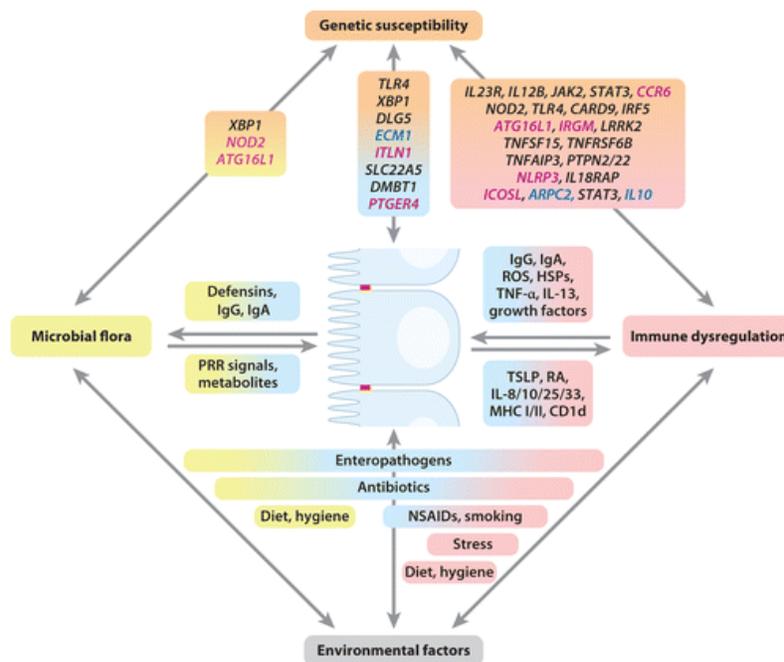
### **I.C.1. Etudes de la composition du microbiote intestinal comme marqueur des MICI**

La composition microbienne du tractus intestinal est restée longtemps inconnue. En effet, mise à part une partie des bactéries intestinales identifiées par culture à partir de selles, jusqu'à 90% des bactéries, ne pouvant pas être cultivées en laboratoire, sont restées inconnues. Afin de mieux explorer le microbiote intestinal, des outils moléculaires à haut débit ont été développés ce qui a permis de dépasser les limites des cultures bactériennes. Ces nouveaux outils ont permis de confirmer que dès la naissance les humains sont colonisés par un microbiote intestinal qui évolue au cours du temps et qui, chez l'adulte, est individu dépendant, stable dans le temps et similaire entre jumeaux monozygotes [20]. Ci-dessous nous nous intéresserons à la composition du microbiote intestinal d'individus sains et atteints de MICI et verrons comme le microbiote pourrait être un marqueur de ces pathologies.

#### *Composition du microbiote intestinal chez les individus sains et atteints de MICI*

Les corrélations entre composition du microbiote et l'état de santé d'individus atteints de MICI a fait l'objet de nombreuses études. Une première analyse de l'ARN 16S bactérien dans les selles de 6 individus atteints de la maladie de Crohn et de 6 individus sains a permis de détecter 88 espèces différentes pour les individus sains pour seulement 54 espèces pour le groupe de malades, établissant une baisse de diversité bactérienne au cours de la maladie [47]. La diversité au sein du phylum des firmicutes était particulièrement réduite : 43 espèces de firmicutes sont détectées dans les selles des individus sains pour seulement 13 espèces chez les malades. Parmi les Firmicutes, la diversité des Clostridiaceae était clairement affectée notamment l'espèce *Clostridium leptum* sous-représentée chez les malades par rapport au groupe contrôle. L'analyse de l'ARN 16S dans les selles de 16 patients atteints de colite ulcéreuse et de 8 individus sains, montre aussi une diminution de la

diversité du microbiote intestinal chez les malades. Le microbiote des patients atteints de colite ulcéreuse est également moins stable au cours du temps par rapport à celui des individus sains [48]. En plus des études basées sur le séquençage de l'ARN 16S, les études métagénomiques ont aussi mis en évidence une corrélation entre composition du microbiote intestinal humain et MICI. L'approche métagénomique a permis d'observer une diminution de la diversité du microbiote chez les individus atteints de MC ou de CU par rapport aux individus sains ainsi qu'une différence significative entre la composition des microbiotes de patients atteints de CU et ceux atteints de MC [108, 116]. Le microbiote des individus atteints de MC présente une diminution plus importante de *Faecalibacterium prausnitzii* [50] que celle observée chez les individus atteints de CU alors qu'*Akkermansia muciniphila* est 10 fois moins abondante chez les malades atteints de CU comparé aux individus atteints de MC [20].



**Figure 5. Etiologie des maladies inflammatoires chroniques de l'intestin (d'après [8]).** Le développement des MICI dépend de plusieurs facteurs : la susceptibilité génétique de l'hôte, le microbiote intestinal, des facteurs environnementaux ainsi que le système immunitaire de l'hôte. Ces facteurs interagissent comme indiqué par les flèches. Les gènes de l'hôte associés aux MICI qui codent pour des composés qui interviennent dans la même voie cellulaire sont disposés sur la même ligne. Les gènes spécifiquement impliqués dans la maladie de Crohn sont indiqués en magenta, alors que ceux spécifiques à la colite ulcéreuse sont en bleu. Les gènes associés aux deux maladies sont

indiqués en noir. Abbreviations : HSPs, heat shock proteins ; MHC, major histocompatibility complex ; NSAIDs, nonsteroidal anti-inflammatory drugs ; PRR, pattern-recognition receptor ; RA, retinoic acid ; ROS, reactive oxygen species ; TSLP, thymic stromal lymphopoietin.

### **I.C.2. Etiologie des maladies inflammatoires chroniques de l'intestin**

L'étiologie des maladies inflammatoires chroniques de l'intestin reste inconnue. Elles sont multifactorielles [8] et les interactions entre le microbiote intestinal humain et le système immunitaire mucoale ainsi que l'influence des facteurs environnementaux sur ces interactions, semblent particulièrement impliqués dans le développement des MICI (Figure 5).

#### *Prédisposition génétique aux maladies inflammatoires de l'intestin*

Les MICI sont considérées comme des maladies polygéniques (où interviennent différents gènes), familiales dans 5-10% des cas et sporadiques dans les autres cas [51]. Les jumeaux monozygotes présentent une concordance phénotypique chez 50-75% des patients atteints de MC. Chez les patients atteints de CU, la concordance phénotypique chez les jumeaux monozygotes est moins importante (10-20%) [51].

Les études sur les facteurs génétiques associés aux MICI ont permis d'identifier plus de 30 loci à risque [8]. Bien que ces études aient révélé un important recouvrement des gènes à risque entre MC et CU, certains facteurs génétiques semblent associés à seulement l'une ou l'autre de ces maladies (Figure 7) [8]. La prévalence des loci à risque change suivant les populations considérées ainsi le récepteur NOD2 et les gènes impliqués dans l'autophagie sont des gènes de risque majoritaires dans la population caucasienne alors que pour la population Asiatique, ils ne constituent pas des facteurs de susceptibilité [8].

Voyons maintenant à titre d'exemple le cas du gène NOD-2 qui est un des loci à risque pour les MICI. Ce gène spécifie un récepteur intracellulaire aux lipopolysaccharides (LPS) et aux peptidoglycanes par l'intermédiaire de sa région terminale riche en leucines (LRR). Il intervient

dans l'immunité innée notamment par l'activation cellulaire via le facteur transcriptionnel NF- $\kappa$ B. Son expression dans les cellules immunocompétentes est limitée aux monocytes et aux macrophages. Les patients atteints de MC montrent un polymorphisme génétique avec trois mutations identifiées dans NOD-2 : deux substitutions nucléotidiques entraînant des substitutions d'acides aminés et une insertion entraînant un décalage du cadre de lecture et la synthèse d'une protéine tronquée. Une première étude [52] a montré au moins une de ces trois mutations chez 29% des patients atteints de MC dans une cohorte mixte de cas familiaux et sporadiques par rapport à une fréquence de 7% chez les individus contrôles. Une étude complémentaire [53] a montré que l'une des deux mutations induisant une substitution d'acide aminé était plus souvent observée dans les formes familiales de MC par rapport aux formes sporadiques de MC. Pour les cas de MC sporadiques, deux études [37, 38] ont montré une corrélation positive entre la mutation conduisant à une protéine NOD-2 tronquée et la maladie. Dans le cas de cette mutation, une explication qui ne repose pas sur l'effet de NOD-2 sur l'activation de NF- $\kappa$ B a été proposée. La forme tronquée de NOD-2 conduit à une diminution de la résistance aux bactéries invasives [56]. Cette propriété explique la présence de bactéries intra-épithéliales qui pourraient stimuler l'inflammation chronique caractéristique des patients MC.

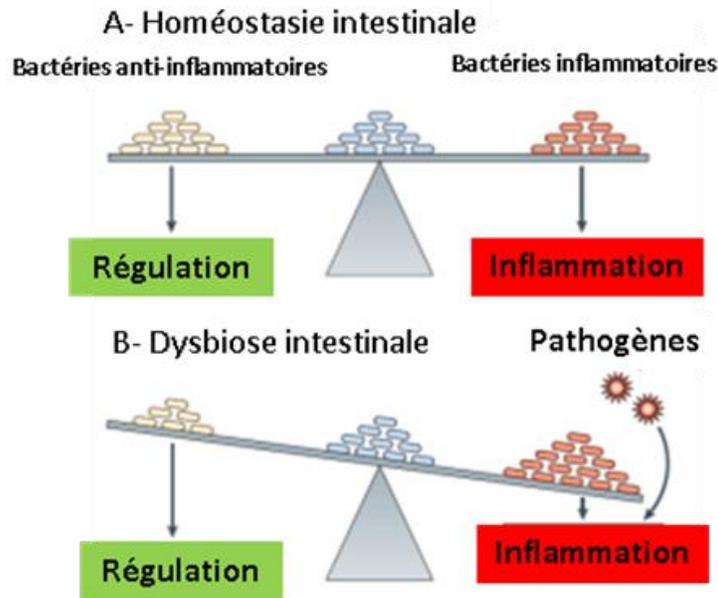


Figure 6. Dérégulation immunitaire associée à l'état de dysbiose du microbiote (d'après [57])

### Rôle des bactéries commensales du microbiote intestinal dans les MICI

Le microbiote intestinal est un facteur important influençant les MICI. La composition du microbiote intestinal assure l'équilibre entre bactéries à effets anti- et pro- inflammatoires ce qui contribue à la santé digestive [57] (

Figure 6). La dysbiose du microbiote intestinal serait ainsi impliquée dans les MICI [4] comme le suggèrent plusieurs études résumées ci-dessous. Nous avons vu précédemment que la maladie de Crohn et la colite ulcéreuse sont toutes deux associées à une diminution de la diversité du microbiote ainsi qu'à une réduction spécifique de certaines espèces telles que *Faecalibacterium prausnitzii*, une espèce anti-inflammatoire dans le cas des MC [8,50,58] et d'*Akkermansia muciniphila* dans le cas des CU [8,58,59]. Par rapport aux individus sains, des patients MC montrent également de plus faibles concentrations (i) du genre *Lactobacillus* au niveau du colon ainsi que (ii) des genres *Lactobacillus* et *Bifidobacterium* dans les selles [10]. La diminution des

genres et espèces bactériennes exerçant des effets anti-inflammatoires associée à la dybiose serait un facteur clef dans la persistance de l'inflammation dans les MICI [10]. En accord avec cette hypothèse, l'administration de *F. prausnitzii* dans un modèle murin de colite induite par le TNBS (acide trinitrobenzène sulfonique) permet d'améliorer les symptômes [60].

La compréhension des facteurs à l'origine des MICI a conduit à mettre l'accent sur l'étude de la relation entre le système immunitaire et le microbiote intestinal. En effet un rôle néfaste du microbiote dans les MICI a été démontré par plusieurs études utilisant des modèles animaux et murins [8]. Une des observations les plus importantes est l'absence de développement de maladie chez des modèles murins d'inflammation intestinale après l'élimination de toute la flore intestinale [61]. De plus, la surface la plus inflammée chez les patients atteints de MICI est celle qui présente la densité la plus importante de bactéries du microbiote [46]. Ces observations suggèrent fortement que les MICI pourraient constituer une réponse immunitaire inappropriée aux bactéries commensales du microbiote chez un hôte génétiquement prédisposé. Toutes ces observations appuient le rôle central joué par le microbiote intestinal dans le développement des maladies inflammatoires chroniques des intestins.

#### Rôle des bactéries pathogènes dans l'induction des MICI

Le rôle joué par certaines bactéries pathogènes dans l'induction des MICI a été évoqué dans plusieurs études présentées ci-dessous. En 1984, Chiodini et al [62] ont montré une forte association entre *Mycobacterium avium* ssp. *paratuberculosis* (MAP) et la MC. Cependant cette observation n'a été confirmée que par un petit nombre d'études et il s'est avéré qu'un traitement antibiotique anti-tuberculeux n'est pas efficace pour le traitement de patients atteints de MC. Ces résultats ont renforcé le scepticisme concernant l'hypothèse d'un rôle de MAP dans les MICI [63,64]. Par la suite, plusieurs bactéries potentiellement pathogènes ont été associées aux MICI : (i)

des souches adhéro-invasives d'*Escherichia coli* ont été isolées à partir de biopsie intestinale de patients atteints de MC [65], (ii) *Yersinia* et *Pseudomonas* sont aussi susceptibles d'être des déclencheurs de MC [66], (iii) *Fusobacterium* a été localisée dans le côlon de patients atteints de CU et cette bactérie induit une CU lorsqu'elle est injectée chez les souris [67,68] et enfin, (iv) une surabondance d'*E. coli* a été observée chez des patients atteints de CU ce qui suggère un rôle de cette bactérie dans la genèse et/ou le maintien de la maladie [69].

### Facteurs environnementaux et MICI

Certains facteurs environnementaux ont été associés au développement des maladies inflammatoires chroniques de l'intestin (MICI) probablement par leurs effets sur le microbiote intestinal ainsi que sur le système immunitaire [8]. Les facteurs environnementaux qui pourraient être associés aux MICI sont entre autres l'hygiène, l'alimentation, le stress ainsi que certains traitements médicamenteux (Figure 5).

L'hygiène est un facteur déterminant pour la sensibilité de certaines populations aux MICI : des études épidémiologiques montrent que les MICI sont inégalement réparties dans le monde avec les taux les plus élevés recensés dans les pays industrialisés [70, 71]. Une des hypothèses pour expliquer ce phénomène est l'hygiène. En début de vie, les habitants des pays industrialisés sont exposés à une diversité microbienne moins importante que d'autres populations. Cette sous-exposition conduirait à l'âge adulte, à un système immunitaire moins tolérant aux micro-organismes et à une réponse immunitaire inappropriée qui augmenterait le risque de développer une MICI [9].

L'alimentation est un facteur environnemental qui influence la composition du microbiote intestinal, identifié comme facteur de risque pour les MICI. Plusieurs études se sont intéressées au rôle de l'alimentation dans les MICI en recherchant des associations entre les régimes alimentaires et l'incidence de ces maladies [9]. Par exemple, une étude réalisée par Hou et al en 2011 [72] a

montré qu'une forte consommation de graisses, d'acides gras polyinsaturés, d'oméga-6 et de viande est associée à une augmentation du risque de MC et de CU. A l'inverse une consommation importante de fibres et de fruits est associée à une baisse du risque de MC et celle de légumes est associée à une baisse de risque de CU [72].

Le stress est un facteur favorable aux MICI potentiellement via l'influence qu'il a sur le microbiote intestinal [73].

Une importante prise d'antibiotique pourrait être un facteur favorable aux MICI. En effet, ce genre de traitement qui ne distingue pas entre les micro-organismes pathogènes et symbiotiques, peut altérer la composition du microbiote et par conséquent favoriser l'apparition des MICI [57].

En résumé tous ces facteurs environnementaux altèrent la composition du microbiote ou dérèglent le système immunitaire. Ils pourraient donc influencer indirectement la susceptibilité aux MICI de certains individus.

#### *Déficiences du système immunitaire et MICI*

La barrière épithéliale intestinale est soutenue dans son rôle de défense par le système immunitaire inné de la lamina propria. En effet comme nous l'avons vu précédemment, le système immunitaire inné comporte des cellules défendant l'hôte face aux intrusions de pathogènes qui maintiennent un état de réponse basale en présence des bactéries commensales. Les cellules dendritiques, les macrophages, les cellules lymphoïdes innées ainsi que les neutrophiles jouent un rôle central parmi les acteurs de la réponse immunitaire innée. Plusieurs études mettent en évidence une relation entre déficiences du système immunitaire inné et MICI.

Les patients atteints de déficiences du système immunitaire inné, comme la granulomatose chronique qui est associée à une réponse défectueuse à des motifs dans l'ADN bactérien (les

oligonucléotides CpG), tendent à développer des MICI [74]. de façon complémentaire, il a été observé que les individus atteints de la MC présentent une réponse immunitaire innée défectueuse avec une diminution de l'activité des macrophages *in vitro* ainsi qu'une altération du recrutement des neutrophiles *in vivo* [75]. De façon similaire chez des patients atteints de CU, la sécrétion de la lipoxine A<sub>4</sub> (médiateur anti-inflammatoire synthétisé par les cellules neutrophiles du système immunitaire inné) au niveau de la muqueuse est faible ce qui est cohérent encore une fois avec un rôle des cellules du système immunitaire inné dans les MICI [74]. Les individus atteints de MICI souffrent de la diminution de la fonction protectrice de la muqueuse intestinale (changement de la composition et de l'épaisseur du mucus et altération de l'état de glycosylation des mucines). ces modifications du mucus permettent une augmentation de l'adhésion des bactéries pathogènes à la muqueuse ce qui favorise le maintien de l'état inflammatoire [45]. Certaines cytokines pro-inflammatoires ont été décrites comme ayant un rôle central dans la pathologie de la maladie de Crohn. Ainsi la baisse de production de la cytokine pro-inflammatoire TNF- $\alpha$  ou encore la diminution de son effet (où les deux à la fois) chez les patients atteints de MC est considérée comme bénéfique [63]. L'IL-22, une cytokine majeure de l'homéostasie intestinale, possède une activité protectrice dans différents modèles de colites par la stimulation de processus antimicrobiens et réparateurs. Des études réalisées chez des patients atteints de MC révèlent une diminution des cellules sécrétrices de IL-22 dans la Lamina propria ce qui est cohérent avec le rôle supposé de ces cytokines (ainsi que le rôle des cellules sécrétrices de ces cytokines) dans les MICI [74].

### **I.C.3. Thérapies contre les maladies inflammatoires chroniques de l'intestin**

Grâce à l'accumulation des connaissances sur les cascades réactionnelles qui gèrent le déclenchement de l'état inflammatoire, plusieurs approches thérapeutiques anti-inflammatoires ont été développées. Les stratégies employées pour atténuer l'inflammation dans le cas de MICI

peuvent cibler plusieurs niveaux de l'inflammation. Citons par exemple l'interruption de certaines voies de signalisation médiées par des cytokines, l'inhibition du facteur transcriptionnel NF- $\kappa$ B (facteur contrôlant la transcription d'un ensemble de cytokines pro-inflammatoires, parmi lesquelles le TNF- $\alpha$ ), l'inhibition du recrutement de cellules inflammatoires dans la muqueuse ainsi que l'antagonisme de certains médiateurs inflammatoires comme le TNF- $\alpha$  sous l'action de l'anticorps monoclonal infliximab [76].

Certaines cytokines anti-inflammatoires peuvent aussi être utilisées pour le traitement de l'inflammation associée aux MICI. A titre d'exemple, il est maintenant bien établi que l'IL-10 conduit à une diminution à la fois de la synthèse des cytokines inflammatoires et de la présentation d'antigènes. Pour cette raison, cette cytokine a été proposée comme immuno-modulateur pour le traitement de la MC. L'apport d'IL-10 par la bactérie *Lactococcus lactis* au niveau de l'épithélium intestinal a montré des résultats prometteurs dans deux modèles de MICI chez la souris [77]. Comme autres thérapies immunosuppressives et immunomodulatrices, les stéroïdes utilisés depuis 1960, sont devenus de plus en plus utilisés dans les maladies inflammatoires chroniques de l'intestin. Bien qu'efficace ce genre de médicament augmente la prévalence d'infections opportunistes [78]. Ce genre de traitement peut de plus induire une hépatotoxicité, une fibrose, un lymphome ainsi qu'une suppression pathologique de la fonction de la moelle osseuse[10].

Les effets d'une nutrition exclusivement entérale (substitution de l'alimentation orale par une sonde introduite dans le tube digestif par voie nasale) sur la rémission des patients atteints de MC ont aussi été largement étudiés. Ces régimes alimentaires contrôlés sont considérés comme une approche thérapeutique dans certains pays [79, 80], ils maintiennent l'état de rémission [81] et sont une alternative aux traitements pharmacologiques souvent accompagnés d'effets secondaires.

Mentionnons qu'une efficacité similaire dans le traitement de la MC a été observée pour des régimes avec différents degrés d'hydrolyse des protéines [82] ; la composition du régime délivré par

voie entérale aurait donc peu d'effet. Même si la thérapie par l'alimentation semble être efficace contre la MC, le mécanisme d'action n'est pas encore établi. Soulignons cependant que les études disponibles suggèrent qu'il n'y a pas d'importants changements dans le microbiote fécal en fonction de l'approche thérapeutique par la nutrition exclusivement entérale [83,84].

Suite aux études montrant une dysbiose chez les patients atteints de MICI, des bactériothérapies c'est-à-dire l'ingestion de bactéries ont été proposées comme traitement des MICI. De nombreuses études ont été réalisées en utilisant des modèles animaux et avec des patients atteints de MICI. Deux essais cliniques suggèrent qu'une préparation d'*Escherichia coli* probiotique est aussi efficace que l'utilisation de mesalazine (l'acide 5-aminosalicylique, un anti-inflammatoire utilisé comme traitement standard pour les MICI) dans la rémission d'une CU [85,86]. A titre de deuxième exemple, le mélange probiotique VSL3# a été utilisé pour le traitement de MICI [87]. Ce mélange bactérien contient 4 souches de lactobacilles (*L. casei*, *L. plantarum*, *L. acidophilus*, *L. bulgaricus*), 3 souches de bifidobacterium (*B. longum*, *B. breve*, *B. infantis*) et une souche de *Streptococcus thermophilus*. VSL3# s'est avéré plus efficace que le placebo pour réduire le risque de pouchite (inflammation de la poche constituée après une anastomose iléo-anale) [88, 89]. Un troisième exemple est l'utilisation de *Lactobacillus casei* pour restaurer la réponse immunitaire IgA chez des individus atteints de MC [90]. Dans une étude précédente du même laboratoire, l'administration orale de *Lactobacillus casei* à des patients atteints de gastro-entérite virale stimulait la réponse IgA et raccourcissait la durée de la diarrhée [91]. Une récente revue sur les thérapies à base de probiotiques pour l'induction et le maintien de la rémission des MICI permet de conclure que la sélection des souches probiotiques, la combinaison de souches utilisée et le choix des doses administrées sont des paramètres cruciaux pour des effets durables en termes de rémission ou d'amélioration dans la phase aiguë de la maladie [7].

En plus de l'approche thérapeutique basée sur l'administration de bactéries probiotiques, la stimulation des populations microbiennes intestinales pourrait constituer une piste prometteuse pour de nouvelles thérapies. Des études ont montré des effets bénéfiques de prébiotiques (composants alimentaires naturels indigestes permettant d'améliorer la santé en influençant favorablement la flore intestinale suite à une stimulation sélective de certaines bactéries probiotiques [92]) dans des modèles animaux de MICI. L'administration d'oligo-fructose ou de lactulose chez le rat traité à l'acide trinitrobenzène sulfonique (TNBS) ainsi que l'administration d'inuline chez les rats présentant une colite induite au Dextran Sulfate Sodium (DSS) permettent d'améliorer les symptômes inflammatoires [15]. Cependant comme pour les probiotiques, une sélection judicieuse du prébiotique selon la maladie à traiter est essentielle puisque contrairement à l'inuline, l'oligo-fructose n'a pas d'effet bénéfique chez les rats traités au DSS [16]. Les prébiotiques ont également été testés chez l'homme. Pour examiner l'efficacité des prébiotiques, 10 patients atteints de MC ont reçu de l'oligo-fructose (15 g/j) et l'évolution de la maladie a été suivie 3 semaines. Ce traitement a réduit significativement l'indice de Harvey-Bradshaw de 9,8 à 6,9 (cet indice reflète l'activité de la MC) [14].

L'utilisation de symbiotiques (mélange d'une bactérie probiotique et d'un composé prébiotique) s'est aussi avérée efficace pour le traitement d'une CU chez 18 patients. En effet, un mois de traitement avec le symbiotique constitué de *Bifidobacterium longum* ( $4 \times 10^{11}$  bactéries/jour) et d'inuline-oligo-fructose (un prébiotique métabolisé par ce probiotique) réduit de façon significative les marqueurs d'inflammation de la muqueuse et améliore les scores histologiques et de sigmoïdoscopies (examen du rectum et du colon) [17].

## ***I.D Interaction des bactéries commensales intestinales avec l'hôte et effets sur la santé***

Le microbiote intestinal joue un rôle dans la modulation de la réponse immunitaire en état d'homéostasie ou de dysbiose intestinale. Cependant, même s'il existe plusieurs bactéries commensales étudiées pour leurs effets immuno-modulateurs, les mécanismes moléculaires impliqués restent quasiment inconnus. Nous présenterons ci-dessous quelques-unes des bactéries commensales décrites comme ayant des effets anti-inflammatoires in vitro dans des modèles cellulaires ou in vivo dans des modèles murins ou chez l'Homme.

a) Plusieurs bactéries du genre *Bacteroides* ont des effets anti-inflammatoires. *Bacteroides fragilis* protège la souris contre la colite induite par le TNBS en stimulant la production des cellules T régulatrices de l'inflammation (Treg) dans le colon. Cet effet protecteur est médié par le polysaccharide A (PSA) extracellulaire de *B. fragilis* [93]. L'internalisation du PSA dans les cellules dendritiques suite à une reconnaissance via la voie de signalisation TLR2/Myd88, active la production des cellules Treg productrices de la cytokine anti-inflammatoire IL-10 ce qui permet d'atténuer l'inflammation due à la colite induite par le TNBS [93]. *Bacteroides thetaiotaomicron* est capable d'atténuer in vitro et in vivo la réponse inflammatoire induite par la bactérie pathogène *Salmonella enterica*. *B. thetaiotaomicron* induit l'export du complexe PPAR- $\gamma$ -RelA (Rel A étant une sous-unité de NF- $\kappa$ B) du noyau des cellules épithéliales intestinales vers le cytoplasme ce qui permet de prévenir la transcription de gènes qui codent pour des composés pro-inflammatoires [94]. PPAR- $\gamma$  (Peroxisome proliferator-activated receptor gamma) un membre de la famille des facteurs transcriptionnels activés par les acides gras présents dans les tissus adipeux et dans l'intestin, joue un rôle dans le métabolisme des lipides et la réponse immunitaire [95]. Les membres de la famille PPAR modulent l'expression de gènes pro-

inflammatoires en inhibant de la translocation nucléaire de NF- $\kappa$ B ce qui empêche l'activation transcriptionnelle du régulon [95]. En plus de son rôle dans l'effet anti-inflammatoire de *B. thetaiotaomicron*, PPAR- $\gamma$  est aussi impliqué dans l'effet immuno-modulateur d'*Enterococcus faecalis* EC16 (isolée du microbiote de nouveau-nés). Cette bactérie exerce un effet anti-inflammatoire in vitro à travers la régulation de l'activation de PPAR- $\gamma$ . En induisant la phosphorylation de PPAR- $\gamma$ , *E. faecalis* déclenche son activation, inhibe la translocation nucléaire de NF- $\kappa$ B (et l'expression de gènes pro-inflammatoires) et permet la transcription de gènes dont les produits sont anti-inflammatoires comme l'IL-10 [96]. Dans un modèle de souris déficientes en IL-2, *Bacteroides vulgatus* protège contre la colite induite par une souche d'*E.coli* par un mécanisme d'action non décrit [97].

b) Les clostridia appartenant au phylum des Firmicutes et au genre Clostridium, sont généralement localisées au niveau du gros intestin où ils semblent exercer des effets immuno-modulateurs. Un mélange de 46 espèces de clostridium protège contre la colite induite par DSS chez la souris en stimulant la production de cellules Treg [98]. Les effecteurs bactériens à l'origine de cet effet sont inconnus cependant, l'implication de récepteurs PRR (Pattern recognition receptors) ainsi que celle de leurs adaptateurs (Myd88, Rip2 et Card9) dans l'effet anti-inflammatoire ont été exclues [6].

c) Comme décrit précédemment, le microbiote des individus atteints de MC présente une diminution du nombre de *Faecalibacterium prausnitzii* [50]. Les effets anti-inflammatoires de cette bactérie ont été analysés dans des modèles cellulaires in vitro ainsi qu'in vivo dans un modèle murin. Testé in vitro sur les cellules caco-2 transfectées avec un gène rapporteur de l'activité du NF- $\kappa$ B, le surnageant de culture de la bactérie induit une réponse anti-inflammatoire en supprimant l'activité NF- $\kappa$ B induite par la cytokine IL-1 $\beta$ , ce qui implique que l'effet anti-inflammatoire pourrait être dû à des métabolites sécrétés par la bactérie. La stimulation des

cellules PBMC (Peripheral Blood Mononuclear Cell) par cette bactérie commensale induit une réponse anti-inflammatoire en augmentant la production de la cytokine IL-10 et en diminuant le niveau d'expression d'IL-12 et d'IFN-gamma qui sont inflammatoires [50]. *F. prausnitzii* induit une réponse anti-inflammatoire in vivo chez les souris avec une colite induite par le TNBS [50]. Toutefois, l'effecteur bactérien ainsi que le mécanisme d'induction de la réponse immuno-modulatrice restent inconnus à ce jour.

## **II. Bactéries probiotiques et santé humaine**

### ***II.A Généralités***

Au cours de ces dernières années, la vague d'études décrivant le rôle des bactéries du microbiote dans la santé humaine a permis de montrer que la colonisation par une flore microbienne diversifiée est importante pour le développement et la régulation du système immunitaire. La rupture de cet équilibre entraîne une perte d'homéostasie métabolique et immunitaire [99]. Certaines bactéries commensales auraient la capacité de contrecarrer les effets d'une dysbiose et de restaurer la tolérance au microbiote [100].

L'effet bénéfique de certaines bactéries sur la santé a commencé à être étudié il y a plus d'une centaine d'années. A la fin des années 1800, Metchnikoff [101] émet l'hypothèse que la consommation de produits fermentés serait bénéfique pour la santé du fait des micro-organismes vivants qu'ils contiennent. Dans les années 1900, Tissier identifie les bifidobactéries dans les selles d'enfants nourris au sein et associe la présence des bifides à une plus faible fréquence de gastro-entérites que chez les enfants nourris au biberon [102]. Il en conclut que ces bactéries sont associées à un tractus intestinal sain. En 2001, un groupe d'experts de l' « Organisation des Nations Unies pour l'alimentation et l'agriculture » (FAO) et de l' « Organisation mondiale de la santé » (OMS) définissent les probiotiques

comme étant « **des micro-organismes vivants qui, administrés en quantité suffisante, confèrent des effets bénéfiques sur la santé de leur hôte au-delà de leurs valeurs nutritionnelles** » [35]. Par conséquent, une bactérie du microbiote peut être qualifiée de probiotique après isolement, purification et mise en évidence d'effets bénéfiques lorsqu'elle est administrée à l'hôte.

Les probiotiques peuvent avoir divers effets bénéfiques sur la santé humaine par exemple, ils diminuent l'intolérance au lactose et les maladies diarrhéiques, ils stimulent le système immunitaire et jouent un rôle dans le traitement des maladies inflammatoires ou encore, ils pourraient prévenir certaines maladies telles que le cancer, les allergies, ainsi que les infections systémiques [10, 103].

Dans la suite de ce chapitre, nous présenterons l'effet probiotique de souches de lactobacilles et bifidobactéries, les deux genres précédemment étudiés par Metchnikoff et Tissier possédant des effets anti-inflammatoires reconnus. Plusieurs études rapportent l'effet positif de l'administration orale de ces bactéries dans la rémission de patients atteints de certaines MICI, voire la diminution des symptômes de crises inflammatoires (partie qui sera développée par la suite).

## ***II.B Les bactéries probiotiques des genres *Lactobacillus* et *Bifidobacterium****

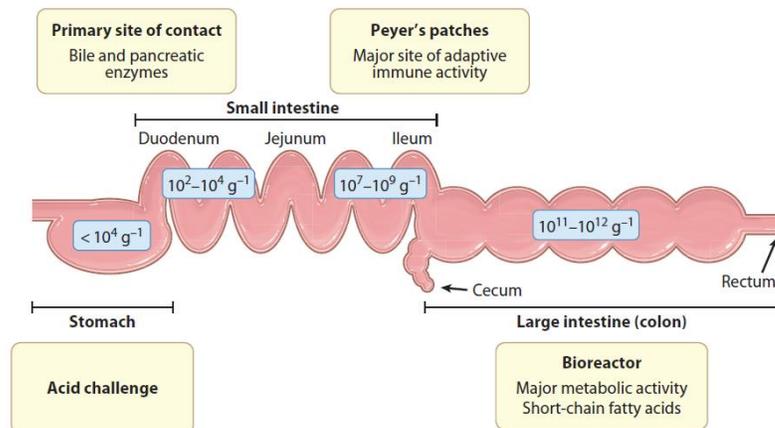
Les lactobacilles sont des bactéries à Gram positif appartenant au genre des Firmicutes. Elles sont immobiles et ont une forme de bacilles ou de coccobacilles isolés ou en chaînettes. Elles sont utilisées comme bactéries fermentaires dans plusieurs produits laitiers comme le fromage et le yaourt. Il existe 145 espèces connues du genre *Lactobacillus* [104, 105]. Ces bactéries peuvent être isolées à partir de différentes niches écologiques principalement à partir de la bouche, de l'intestin et du vagin d'êtres humains et d'animaux mais aussi à partir d'aliments tels que des végétaux, les viandes, les poissons et les produits laitiers. Notons ici qu'au niveau de l'écosystème intestinal, les lactobacilles font partie de la flore sous-dominante [106].

Les bifidobactéries sont des bactéries à Gram positif constituant un genre d'actinobactéries. Elles sont immobiles et présentent une forme de bacilles d'aspects irréguliers. On connaît 30 espèces différentes de bifidobactéries dont certaines sont identifiées comme probiotiques [107]. Ces bactéries sont naturellement présentes dans l'intestin qu'elles colonisent durant la première semaine après la naissance puis leur abondance diminue avec l'âge : elles représentent 95% du microbiote chez l'enfant pour seulement 3% chez l'adulte [107].

Des souches de ces deux genres bactériens sont capables d'agir en tant que probiotiques dans l'écosystème intestinal.

## II.C Niche écologique, lactobacilles et bifidobactéries intestinaux

L'intestin humain présente sur toute sa longueur plusieurs écosystèmes (Figure 7). Les bactéries du genre *Bifidobacterium* ainsi que plusieurs *Lactobacilli* font partie de ces écosystèmes.



**Figure 7. Représentation schématique de l'écosystème intestinal humain (d'après [114]).** Les caractéristiques des différentes régions de l'intestin sont encadrées et colorées en jaune. Le nombre de bactéries par région est encadré et coloré en bleu. Au cours de leur passage à travers l'intestin, les bactéries alimentaires font face au stress acide dans l'estomac, suivi du stress enzymatique de la bile et du pancréas dans l'intestin grêle. L'abondance des bactéries probiotiques diminue en avançant dans l'intestin jusqu'à devenir une population minoritaire dans le gros intestin, où réside un grand nombre de bactéries endogènes. L'iléon correspond à la région de l'intestin où le système immunitaire est le plus actif, ce qui explique que cette région est généralement décrite comme une importante zone d'interaction avec le système immunitaire pour les probiotiques. Au niveau du gros intestin, les bifidobactéries (espèces endogènes et apportées par l'alimentation) contribuent à la dégradation de l'aliment et du glycane. Cela permet de générer une variété d'acides gras à courtes chaînes comme l'acétate, le lactate, le propionate et le butyrate.

Généralement, les bactéries sélectionnées comme probiotiques sont choisies pour leur résistance au passage à travers la partie supérieure du tractus digestif. En effet pour délivrer leurs effets bénéfiques à la santé, les probiotiques ingérés oralement au sein d'un aliment vont devoir résister à de nombreuses agressions (Figure 7). Les bactéries probiotiques rentreront également en compétition avec d'autres espèces du microbiote pour les substrats et l'adhésion éventuelle à l'épithélium intestinal.

## ***II.D Propriétés intrinsèques de résistance dans le tractus digestif : Lactobacillus et Bifidobacterium***

Certaines bifidobactéries sont capables d'adhérer aux cellules épithéliales intestinales comme décrit *in vitro* sur les cellules Caco-2 [108] grâce à plusieurs propriétés intrinsèques :

- Leurs propriétés hydrophobes augmentent l'adhésion aux surfaces [109,110].
- Leur capacité à dégrader des sucres complexes confère un avantage dans la compétition pour les nutriments et se révèle un facteur déterminant pour l'implantation [111].
- Leur capacité à moduler la réponse immunitaire et à induire la tolérance de l'hôte à leur rencontre [107].

En plus de leur capacité à coloniser l'intestin, les bifidobactéries et les lactobacilles sont résistants aux différentes conditions de stress rencontrées au cours de leur passage dans l'intestin (Figure 7) ; plusieurs composés ont été impliqués dans cette résistance comme décrits ci-dessous. Le rôle potentiel de certains transporteurs dans la résistance aux stress des bifidobactéries a été étudié. En effet, l'expression du gène codant pour le transporteur BL0920 chez *B. Longum* NCC2705 augmente de 20 fois en présence de sels biliaires ce qui suggère qu'il contribue à la résistance au stress biliaire [112]. Une corrélation entre la production d'exopolysaccharides et la résistance aux sels biliaires et au pH acide a été démontrée pour des *Bifidobacterium* isolés des fèces d'enfants et du lait maternel [113]. D'autres composés ont été étudiés chez les lactobacilles : certains gènes impliqués dans la formation

de l'enveloppe cellulaire (biosynthèse des composants de la paroi ainsi que d'acides gras) chez *L. plantarum* sont activés en réponse à un stress biliaire [114].

La capacité des Bifidobactéries à adhérer à l'épithélium intestinal ainsi qu'à résister aux conditions de stress leur permettent de persister dans le tractus digestif. Ainsi la persistance de la souche probiotique *B. breve* UCC2003 s'éleve à une dizaine de jours dans le tractus digestif de souris [115]. Cette capacité à persister dans le tractus digestif est un avantage puisqu'il laisse plus de temps à la bactérie pour agir sur son hôte.

## ***II.E Effets anti-inflammatoires des bactéries probiotiques, lactobacilles et bifidobactéries***

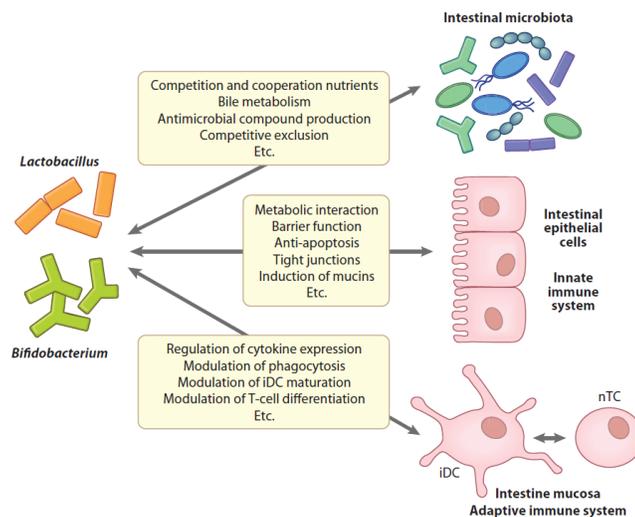
### **II.E.1. Etudes cliniques et mise en évidence de l'effet anti-inflammatoire de certains lactobacilles et bifidobactéries**

Plusieurs études cliniques dont une sélection est présentée ci-dessous, ont confirmé l'intérêt d'étudier l'effet anti-inflammatoire des bifidobactéries et des lactobacilles. Plusieurs études mettent en évidence la diminution de l'abondance de ces genres bactériens chez des patients souffrant de MICI. Ainsi Favier et al rapportent en 1997 que les patients atteints de MC présentent une diminution de concentration de ces deux genres bactériens dans leurs flores fécales par rapport aux sujets sains [116]. De façon similaire, des patients atteints d'IBS présentent une abondance réduite de bifidobactéries au niveau duodéal par rapport aux individus sains [109]. L'administration orale du mélange probiotique VSL#3 (contenant entre-autres 3 souches de bifidobactéries et 4 souches de lactobacilles) permet de prévenir la rechute de patients atteints de CU [117]. D'autres études sont venues par la suite pour conforter l'hypothèse d'un rôle de certains bifidobactéries et lactobacilles dans le maintien de la

rémission de patients atteints de CU. A titre d'exemple, une souche de *L. rhamnosus* GG a été décrite comme ayant un effet préventif (supérieur à celui du placebo) contre la pouchite [118].

## II.E.2. Modes d'interaction de bactéries probiotiques avec l'hôte

Les bactéries probiotiques des genres bifidobactéries et lactobacilles possèdent des caractéristiques leur permettant de résister aux conditions environnementales stressantes du tractus digestif et d'exercer des effets bénéfiques pour la santé de leur hôte notamment des effets anti-inflammatoires au niveau intestinal. Trois modalités d'interaction avec l'hôte sont actuellement proposées comme rappélé ci-dessous (Figure 8).



**Figure 8. Représentation schématique des différents modes d'interaction des probiotiques avec leur hôte (d'après [119]).**

- Interactions entre bactéries. Il peut s'agir de compétition ou de coopération entre bactéries par exemple pour la digestion des nutriments ou la production de composés antimicrobiens. La compétition peut conduire à l'exclusion de certaines bactéries notamment de pathogènes intestinaux ;
- Interaction avec les cellules épithéliales et le système immunitaire inné. Parmi ces interactions citons, les interactions métaboliques via la production d'acides organiques, produits de la

fermentation bactérienne qui constituent des nutriments pour les cellules épithéliales mais aussi des interactions avec les acteurs de la barrière intestinale qui modulent (i) la production de mucines, (ii) la perméabilité mucosale en agissant sur les jonctions serrées, (iii) l'activité du système immunitaire inné ou encore (iv) la mort cellulaire programmée (apoptose) des cellules épithéliales.

- Interaction avec le système immunitaire adaptatif : modulation de la différenciation des cellules T et régulation de l'expression des cytokines, etc.

#### Interactions bactéries-hôte et effets anti-inflammatoires

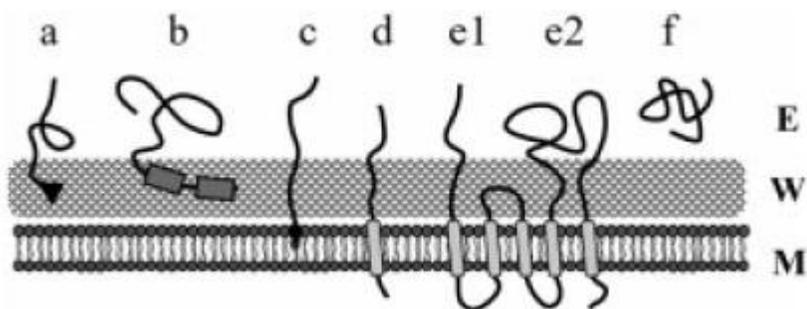
Dans cette partie, nous allons parler des interactions entre bactéries comme moyen pour certaines bactéries probiotiques de protéger l'hôte contre d'éventuelles réponses inflammatoires. Plusieurs travaux montrent que certains lactobacilles et bifidobactéries sont capables de protéger les entérocytes (cellules de l'épithélium intestinal) contre la réponse inflammatoire induite par des pathogènes. A titre d'exemple, *L. acidophilus* Bar13, *L. plantarum* Bar10, *B. longum* et *B. lactis* réduisent l'adhésion des entéropathogènes *E. coli* H10407 et *Salmonella typhimurium*, sur les cellules intestinales Caco-2 [120]. *B. animalis* et *L. casei* produisent des antagonistes à certaines bactéries pathogènes comme *E. coli* [110]. En addition certaines bactéries probiotiques interagissent avec des bactéries commensales du microbiote afin de réguler la production de cytokines pro-inflammatoires et de diminuer l'inflammation. Une étude réalisée par Lan et al a montré que *L. rhamnosus* GG est capable de réguler l'induction de cytokines pro-inflammatoires par *E. coli* et *Bacteroidetes ovatus* [121]. Certains probiotiques sont capables de renforcer la barrière épithéliale et par conséquent de prévenir la translocation bactérienne, comme cela a été montré pour *B. lactis* [122].

L'interaction entre bactéries n'est pas le seul moyen pour exercer des effets anti-inflammatoires sur l'hôte. Plusieurs études mettent en avant la capacité des probiotiques à moduler de façon directe la

réponse immunitaire de l'hôte restaurant ainsi l'équilibre entre les cytokines pro- et anti-inflammatoires. Citons à titre d'exemples, (i) l'augmentation de la sécrétion de cytokines anti-inflammatoires telle que l'IL-10 et l'inhibition de la cytokine pro-inflammatoire IL-8 par certains lactobacilles et bifidobactéries [123–125] et (ii) la diminution par *Lactobacillus casei* Shirota de la sécrétion des cytokines pro-inflammatoires IL-6 et IFN- $\gamma$  (produites par les cellules mononucléaires de la lamina propria) [126].

### II.E.3. Effecteurs bactériens à l'origine des effets anti-inflammatoires de certains probiotiques

Les bactéries probiotiques ayant des effets anti-inflammatoires peuvent induire ces effets par leurs composés exposés à la surface ou sécrétés [119, 127]. Comme rappelé dans la Figure 9, la paroi des



**Figure 9. Représentation schématique de la paroi des bactéries à Gram positif et des différents types protéines susceptibles d'être exposées à la surface.** (a) protéine liée de façon covalente au peptidoglycane, (b) Protéine liée de façon non covalente au peptidoglycane. (c) lipoprotéine liée à la membrane bactérienne. (d) Protéine ancrée à la membrane par une hélice transmembranaire (TMH) N- ou C-terminale. (e) Protéine ancrée à la membrane par plusieurs TMH, (e1) protéine avec une extrémité N- ou C-terminale exposée à la surface, (e2) protéine avec une boucle exposée à la surface (f) protéine sécrétée dans le milieu extracellulaire M, membrane bactérienne ; W, peptidoglycane ; E, protéines exposées à la surface ou sécrétés (D'après [128]).

bactéries à Gram positif est une structure complexe qui contient plusieurs composés caractéristiques.

Cette structure multicouche sert à protéger la bactérie de l'environnement. La bactérie est tout d'abord entourée d'une couche épaisse de peptidoglycane, dans laquelle sont distribuées des protéines (exposées ou pas à la surface). Une membrane sépare le peptidoglycane du milieu intracellulaire microbien. Les protéines qu'on retrouve au niveau du peptidoglycane peuvent être ancrées à la

membrane comme les lipoprotéines. Les lipoprotéines ainsi que les protéines membranaires peuvent aussi être exposées à la surface cellulaire. En plus des protéines de surface, certaines protéines sont sécrétées par la bactérie dans le milieu extracellulaire [119].

Les bactéries probiotiques comme toutes les autres bactéries à Gram+, sont considérées comme des organismes étrangers par le système immunitaire. Les motifs moléculaires associés aux microbes (microbe-associated molecular patterns ou MAMPs) sont des motifs permettant au système immunitaire de reconnaître ces « corps étrangers ». Certains polysaccharides de paroi, le peptidoglycane, des lipoprotéines et l'acide lipotéichoïque sont des MAMPs. Les motifs moléculaires associés aux microbes sont capables d'interagir avec les récepteurs TLR (Toll-like receptor), CLR (C-type lectin receptors), NLR (nucleotide oligomerization domain (NOD) like receptors) qui sont exprimés par différents types cellulaires notamment les cellules épithéliales et immunitaires via lesquelles la réponse immunitaire innée est induite [127]. Des différences de composition et de structure des parois cellulaires peuvent suffire à conférer des propriétés spécifiques de souches ou d'espèces dans l'interaction avec l'hôte par exemple des effets immuno-modulateurs [119].

L'étude des effecteurs bactériens à l'origine d'effets anti-inflammatoires de bactéries probiotiques a permis d'en identifier quelques-uns. Il s'agit de composés exposés à la surface de la bactérie, de molécules sécrétées par la bactérie, ou d'ADN bactérien. Nous rappelons brièvement ci-dessous, les résultats de ces études.

**Serpine de *B. longum* NCC2705.** Un homologue de serpines eucaryotes qui inhibent des protéases à sérine, a été identifié chez *B. longum* NCC2705. Cette serpine inhibe l'élastase des cellules neutrophiles humaines qui est une sérine-protéase libérée en condition d'inflammation et de ce fait, réduit les effets délétères dus à l'activité de ces protéases dans les MICI [129].

**Facteur soluble chez *B. breve* C50.** Un facteur soluble de nature inconnue, capable de réduire la phosphorylation d'I $\kappa$ B (inhibiteur de NF- $\kappa$ B) est produit par *B. breve* C50 ; il permet de réduire l'inflammation en inhibant la voie d'activation de NF- $\kappa$ B [130].

**ADN des bactéries probiotiques du mélange VSL#3.** L'ADN des bactéries probiotiques du mélange VSL#3 (*L. casei*, *L. plantarum*, *L. acidophilus*, *L. bulgaricus*, *B. longum*, *B. breve*, *B. infantis* et *Streptococcus thermophilus*) atténuent la réponse inflammatoire induite par différents stimuli à la fois *in vitro* avec des cellules épithéliales HT-29 et *in vivo* en modèles murins [131].

**Le pilus de *L. rhamnosus* GG.** Le pilus de *L. rhamnosus* GG codé par l'opéron *spa* inhibe la réponse inflammatoire dans des cellules Caco-2 en diminuant le niveau d'interleukine pro-inflammatoire IL-8. Ce motif moléculaire associé aux microbes est produit par plusieurs lactobacilles cependant sa structure (composition, longueur, etc.) varie suivant les souches ; ces variations pourraient expliquer que l'effet probiotique ne soit observé que chez certaines des bactéries possédant le pilus [127].

**Protéines p75 et p40 de *L. rhamnosus* GG.** Un autre effecteur de *L. rhamnosus* GG identifié dans le surnageant de culture est capable de prévenir la mort cellulaire programmée induite par des cytokines pro-inflammatoires et de renforcer les jonctions serrées entre cellules épithéliales. Les deux protéines sécrétées p75 et p40 sont impliquées dans cet effet anti-inflammatoire [132]. La protéine p40 joue un rôle protecteur pour les cellules épithéliales au cours d'une colite induite chez la souris [133]. Les gènes spécifiant p40 et p75 codent respectivement pour un antigène de surface et une glycoside hydrolase associée à la paroi [134].

**LTA et SlpA de *L. acidophilus* NCFM.** L'acide lipotéichoïque (LTA) ainsi que la protéine SlpA pour « S-layer protein A » de *L. acidophilus* NCFM diminuent l'inflammation en induisant la production de la cytokine anti-inflammatoire IL-10 par les cellules dendritiques et en

stimulant le développement des cellules T qui sécrètent de grandes quantités de cytokine IL-4. Le mécanisme d'action moléculaire reste toutefois inconnu à ce jour [135].

**Peptidoglycane.** Les propriétés anti-inflammatoires de certaines bactéries probiotiques peuvent être dues à une spécificité au niveau du peptidoglycane (PEG). Le PEG purifié de *L. salivarius* Ls33 induit un effet anti-inflammatoire *in vitro* et *in vivo* dans un modèle de colite induite par le TNBS. Cet effet protecteur n'est pas retrouvé avec le PEG purifié d'une autre souche de *L. salivarius* non anti-inflammatoire. L'analyse de la différence de structure du PEG entre ces deux souches a mis en évidence un muropéptide (peptide associé à la paroi) additionnel chez *L. salivarius* Ls33 qui serait le composé actif dans l'effet probiotique de cette souche [136].

**Facteur peptidique de faible poids moléculaire chez *Streptococcus salivarius*.** La bactérie commensale *S. salivarius* limite l'activation de NF- $\kappa$ B *in vitro*. Cet effet implique l'action d'un composé peptidique de faible poids moléculaire dont la nature exacte n'est pas précisée à ce jour [137].

**Facteur inconnu chez *L. casei*.** *L. casei* atténue l'inflammation induite au niveau des cellules épithéliales par l'infection à *Shigella flexneri* en bloquant la voie d'activation de NF- $\kappa$ B via la stabilisation d'I $\kappa$ B (inhibition de son ubiquitination) [138, 139].

### III. Alimentation, bactéries alimentaires et santé humaine

L'alimentation joue un rôle important dans notre santé qui va bien au-delà de l'aspect nutritif. En plus de l'impact sur la composition du microbiote intestinal [140], nos changements d'habitudes alimentaires ont été mis en cause dans la recrudescence de maladies telles que l'obésité [21], le diabète de type 2 [141], les maladies cardiovasculaires [142] ainsi que les maladies inflammatoires chroniques de l'intestin [9, 72, 143]. Hormi les effets de l'aliment proprement dit, les micro-

organismes associés aux produits fermentés pourraient avoir un effet sur la santé. Ce chapitre détaille le rôle de l'alimentation et des bactéries alimentaires dans la santé humaine. Nous mettrons sur les produits fermentés laitiers comme le yaourt et leurs bactéries et leur rôle dans la santé en particulier, dans les MICI.

### ***III.A Alimentation et santé humaine***

#### **III.A.1. Impact de l'alimentation sur le microbiote**

La colonisation bactérienne de l'intestin commence à la naissance et elle devient plus stable au bout de la première année de la vie [144]. Plusieurs études se sont intéressées à l'impact de l'alimentation sur le microbiote en début de vie chez des nouveaux nés, en comparant la flore intestinale formée après allaitement à celle après une alimentation à base de préparation lactée. La plus importante observation a été la forte proportion de *Bifidobacterium* chez les enfants nourris au sein, alors que les nouveaux nés ayant reçu la préparation lactée possèdent un microbiote caractérisé par une haute proportion de *Bacteroides* et de bactéries du groupe *Clostridium coccoïdes* [145–148]. Cela pourrait en partie être expliqué par la composition du lait maternel. Il contient de grandes quantités de galacto-oligosaccharides (oligosaccharides contenant du galactose) [149, 150]. qui protègent le nourrisson de la diarrhée et ont un effet bifidogène (c'est à dire qu'ils stimulent le développement des bifidobactéries) [151].

D'autres expériences ont été entreprises afin d'évaluer l'association entre le régime alimentaire et la composition du microbiote intestinal humain adulte. La détermination de la composition du microbiote à partir d'échantillons fécaux de 98 individus a permis de les diviser en 3 groupes distincts appelés entérotypes. Ils diffèrent principalement par la prédominance des genres *Ruminococcus*, *Bacteriodes* ou *Prevotella*. Ces différences dans le microbiote des individus sont en

partie expliquées par leur régime alimentaire [140]. Un régime alimentaire riche en protéines animales, en graisses et pauvre en carbohydrates, similaire à celui des pays industrialisés, est associé au microbiote dominé par des bactéries du genre *Bacteroides* et comportant peu de *Prevotella* (il s'agit de l'« entérotipe *Bacteroides* »). Un régime alimentaire riche en carbohydrates simples et pauvre en protéines animales et en graisses est associé à un microbiote caractérisé par une prédominance de bactéries du genre *Prevotella* et une faible abondance de bactéries du genre *Bacteroides* (correspondant à l'« entérotipe *Prevotella* »).

Bien qu'un régime alimentaire contrôlé (riche en graisses et pauvre en fibres, ou inversement), sur une période de 10 jours, induise des changements détectables dans la composition du microbiote initial, il ne conduit pas à un changement d'entérotipe. Ce résultat indique que l'influence de l'alimentation sur l'entérotipe ne pourrait être observée qu'avec un régime sur du long terme [140, 152].

### **III.A.2. Aliments prébiotiques: impact sur les MICI**

Les prébiotiques correspondent à des composés alimentaires naturels indigestes permettant d'améliorer la santé en influençant favorablement la flore intestinale suite à une stimulation sélective de certaines bactéries probiotiques [92]. Dans l'industrie alimentaire, les prébiotiques sont ajoutés à certains aliments comme des produits lactés ou des barres de céréales, ce qui permet de leur attribuer l'appellation « aliment fonctionnel » (functional food). Depuis une dizaine d'années, certains aliments pour nourrissons sont partiellement enrichis en prébiotiques. Les prébiotiques sont des carbohydrates à courtes chaînes qui ne sont pas digérés dans l'intestin grêle. Ils traversent donc le tube digestif pour atteindre le colon sous une forme inchangée. Au niveau du colon, ils servent de substrats à la flore bactérienne. Parmi les prébiotiques, ceux qui sont présents naturellement dans les aliments et qui sont les plus abondants sont les fructo-oligosaccharides (FOS) et l'inuline (mélange

de polysaccharides). Ces prébiotiques constituent un important réservoir de carbohydrates dans certains légumes, fruits, céréales, ainsi que dans le lait maternel.

L'apport d'inuline et de FOS chez l'adulte (10 g par jour) stimule la prolifération des bactéries lactiques, notamment des bifidobactéries dans l'intestin. Ces bactéries fermentent les fibres prébiotiques et les réduisent en acides gras à chaînes courtes comme l'acétate, le propionate, le butyrate et le lactate [153]. L'efficacité des prébiotiques contre les MICI a été étudiée (i) *in vitro* [154], (ii) dans des modèles animaux (colites induites par le DSS ou le TNBS) [155–158] et (iii) chez des patients atteints de MC ou de CU [159–161]. Ces travaux montrent tous une modulation de la composition du microbiote intestinal par les prébiotiques, associée à une amélioration des symptômes de MC ou de CU chez les patients.

### ***III.B Rôle des produits fermentés et des bactéries lactiques dans la santé humaine***

La fermentation est l'un des plus anciens et des plus importants procédés biologiques utilisés en alimentation. Les premiers produits fermentés ont très probablement été obtenus par acidification spontanée de jus végétaux comme le vin ou la bière ou par contamination spontanée du lait pour la production de fromages ou de yaourt. Les premières preuves de l'existence de produits laitiers fermentés remontent à 8000 ans avant Jésus Christ [162].

La fermentation peut être réalisée à partir de différents aliments, par exemple des végétaux comme la betterave, le concombre, les dattes, et le soja ou des produits d'origine animale comme la viande, le lait ou le poisson. Le procédé de fermentation confère non seulement une saveur particulière au produit mais permet aussi de conserver les aliments.

En 1857, Louis Pasteur a établi un lien entre la fermentation lactique et les bactéries [163]. En 1873, la première culture bactérienne pure fut celle de *Lactococcus lactis* (« *Bacterium lactis* ») obtenue par Joseph Lister [164]. En 1904, Metchnikoff isole le bacille bulgare, *Lactobacillus delbrueckii*

*ssp. bulgaricus* présent dans le yaourt. Il étudie les propriétés acidifiantes de cette bactérie et émet l'hypothèse avant-gardiste que les bactéries contenues dans les laits fermentés ont un effet bénéfique sur la santé [101]. Cette hypothèse l'amène à proposer l'introduction des produits laitiers fermentés dans le régime alimentaire. En 1905, les industries laitières produisent pour la première fois du yaourt à partir de souches isolées à l'Institut Pasteur [165]. En 2012, la consommation moyenne de yaourts et de desserts lactés était estimée à 29,3 kg par français et celle de fromage à 25,7 kg, ce qui a permis d'élire les produits laitiers « aliment préféré des Français » par l'institut CSA d'études et sondages (source <http://www.maison-du-lait.com>).

L'appellation bactéries lactiques, inspirée de la capacité de ces bactéries à fermenter le lactose du lait, désignait initialement à la fois les bactéries coliformes (entérobactéries fermentaires à Gram-) et lactiques. En 1901, Beijerinck fait la découverte que les lactobacilles sont des bactéries à Gram+, ce qui aboutit à la séparation de ces bactéries des coliformes [104]. Les bactéries lactiques deviennent les bactéries à Gram+ convertissant le pyruvate en acide lactique pour produire le NAD+ utilisé dans la glycolyse. Pour se développer, toutes les bactéries lactiques ont besoin d'une source de carbone (carbohydrates fermentescibles) mais certaines d'entre elles ont d'autres exigences nutritionnelles spécifiques, telles que des acides aminés [166, 167], des vitamines et des acides gras [166].

L'industrie agroalimentaire utilise les bactéries lactiques pour fermenter divers aliments et produire par exemple du lait fermenté, du yaourt, des fromages, des olives ou encore de la choucroute.

Les bactéries lactiques agissent sur les caractéristiques organoleptiques du produit fermenté en lui donnant une saveur et une texture caractéristiques. Ces bactéries produisent des molécules comme l'acétaldéhyde, l'éthanol ou l'acide lactique qui sont importantes dans le développement de la saveur du produit final. Elles augmentent aussi la durée de vie du produit fermenté en inhibant le

développement de la flore bactérienne indésirable (pathogène ou d'altération) par l'acidification du milieu ou la production de bactériocines.

Parmi les produits laitiers fermentés, le yaourt, défini par le Codex Alimentarius de 1992 comme un produit du lait fermenté résultant de la fermentation du lactose en acide lactique par *Lactobacillus delbrueckii bulgaricus* et *Streptococcus thermophilus* [168], est l'un des mieux connus. Ces deux bactéries fermentaires sont des bactéries lactiques dont la quantité dans le yaourt doit être au nombre supérieur ou égal à  $10^8$  bactéries par gramme en fin de fabrication [10]. *Lactobacillus delbrueckii bulgaricus* et *Streptococcus thermophilus* peuvent être isolées à partir de selles humaines ce qui reflète leur survie au moins partielle pendant le transit dans le tractus gastro-intestinal [169].

Les effets bénéfiques des produits fermentés peuvent relever de deux types d'actions : (i) la transformation du produit initial par la fermentation modifie sa composition et pourrait modifier son impact sur la santé et (ii) les bactéries fermentaires pourraient interagir avec l'hôte et avoir des effets directs sur la santé. Cette partie du manuscrit décrira les connaissances sur ces deux concepts.

### **III.B.1. Produits fermentés du lait et apport en nutriments, l'exemple du yaourt**

La composition en nutriments du yaourt est définie par la composition en nutriments du lait dont il dérive. Cette dernière est affectée par de nombreux facteurs parmi lesquels nous pouvons citer : les facteurs génétiques, l'alimentation, le stade de lactation et l'âge des vaches laitières. D'autres variables influencent la valeur nutritionnelle du lait, par exemple la température, la durée de chauffage, l'exposition à la lumière ainsi que les conditions de stockage. Le changement de composition du lait intervenant pendant la fermentation lactique influence la valeur nutritionnelle et physiologique du produit final. La composition nutritionnelle finale du produit fermenté est aussi affectée par les espèces et les souches bactériennes utilisées pour la fermentation, ce qui pousse les

industriels à sélectionner les souches de bactéries lactiques les plus intéressantes sur le plan biotechnologique, pour obtenir un produit final avec une qualité organoleptique (texture, saveur, arômes, etc.) attendue.

#### Apport en protéines

Le contenu en protéines des yaourts commercialisés est généralement plus important que celui du lait du fait de l'ajout de poudre de lait écrémé qui augmente le contenu protéique du produit final. Il a été prouvé que les protéines contenues dans le yaourt sont plus facilement digérées par l'homme que les protéines du lait [170] du fait des deux phénomènes suivants :

- Pendant la fermentation, le traitement à la chaleur ainsi que la production d'acide entraînent la coagulation des caséines ce qui peut contribuer à une meilleure digestibilité des protéines du yaourt par rapport à celles du lait [10].
- La prédigestion des protéines du lait dans le yaourt par les bactéries lactiques, *L. delbrueckii* ssp. *bulgaricus* et *S. thermophilus* pourrait faciliter leur digestion par l'homme [170,171]. De plus dans le yaourt, le taux d'acides aminés libres serait plus élevé notamment pour la proline et la glycine que dans le lait.

#### Apport en lactose

Le lait ainsi que ses produits transformés comme le yaourt, sont une source exclusive de lactose dans l'alimentation humaine. Notons ici que le taux de lactose dans le lait est d'environ 6% [10]. Dans l'intestin, le lactose est hydrolysé par la beta-galactosidase humaine en glucose et galactose. Les monosaccharides ainsi produits sont absorbés et utilisés comme source d'énergie.

Pendant la fermentation, *L. delbrueckii* ssp. *bulgaricus* et *Streptococcus thermophilus* hydrolysent 20% à 30% du lactose en glucose et galactose [168] qui est converti en acide lactique. La

diminution de la concentration de lactose au cours de la fermentation du lait explique en partie pourquoi le yaourt est mieux toléré que le lait par les personnes intolérantes au lactose [172–174].

#### Apport en lipides

La matière grasse du lait subit des changements biochimiques pendant le processus de fermentation. De petites quantités d'acides gras libres sont libérées sous l'action des lipases bactériennes. La plupart des yaourts vendus contenant peu de matières grasses, l'hydrolyse des lipides modifie généralement peu la qualité de la plupart des yaourts. Les principales sources d'acide linoléique conjugué (ALC, appartenant à une famille de l'acide linoléique qui est un acide gras essentiel) dans notre alimentation sont la viande et les produits laitiers des mammifères. Une augmentation de la consommation de matières grasses laitières est associée à l'augmentation de la concentration de l'ALC dans le tissu adipeux humain [175] et le lait maternel [176]. L'ALC est décrit comme ayant des propriétés immuno-stimulatrices et anti-carcinogènes [177]. Dans une récente étude avec des cellules cancéreuses de sein et de colon, Kemp *et al.* [178] ont montré que les propriétés anti-carcinogènes de l'ALC résulteraient de la capacité des isomères de l'ALC à inhiber l'expression des cyclines, une famille de protéines impliquées dans la régulation du cycle cellulaire. Cette inhibition ralentirait la progression du cycle cellulaire de la phase G1 à la phase S. De plus, l'ALC induit l'expression du répresseur tumoral p53.

#### Apport en minéraux

Le yaourt est une excellente source de minéraux et de microéléments comme le calcium, le potassium, le phosphore, le magnésium et le zinc. En effet, le yaourt comme tous les autres produits laitiers, fournit la majorité du calcium indisponible dans un régime alimentaire occidental typique. Une des fonctions la plus importante du calcium est son rôle dans la formation et la minéralisation osseuses [10].

### Apport en vitamines

Les produits laitiers sont généralement considérés comme excellents sources de vitamines B2, B3, B6, B9 et B12 [179]. Du fait de la sensibilité des vitamines à la chaleur et à l'oxydation, le procédé de fabrication de yaourt peut diminuer la quantité ou la qualité des vitamines s'il n'est pas bien maîtrisé. Les ferments utilisés peuvent aussi modifier le contenu en vitamines du produit fini car certaines souches de bactéries lactiques nécessitent de la vitamine B pour croître, alors que d'autres en produisent [179].

Le folate aussi appelé vitamine B9 est synthétisé par certaines bactéries lactiques [180, 181]. La capacité à synthétiser ou utiliser la vitamine B9 d'espèces utilisées pour la fermentation du lait et la production de yaourt a été récemment examinée [181]. *Streptococcus thermophilus* et *Bifidobacterium* sont producteurs du folate alors que les lactobacilles épuiseraient le folate présent sous la forme de 5-méthyl-tetrahydrofolate [182]. Cette dernière observation semble toutefois contredite par la description d'une voie complète de biosynthèse du folate dans la souche *L. delbrueckii* ssp. *bulgaricus* ATCC 11842 [167] ce qui suggère que ce phénotype pourrait être souche-dépendant chez les lactobacilles.

### **III.B.2. Produits fermentés du lait et activités enzymatiques, l'exemple du yaourt**

Depuis sa découverte en 1960, la mauvaise digestion ou intolérance au lactose a été l'objet d'importantes recherches. Plus de la moitié de la population adulte mondiale souffre d'intolérance au lactose [10] en raison d'une baisse physiologique de la lactase ( $\beta$ -galactosidase) intestinale [183]. L'intolérance au lactose peut se développer suite à une inflammation ou une perte fonctionnelle de la muqueuse intestinale accompagnant la maladie de Crohn, la maladie cœliaque (intolérance au gluten), le syndrome de l'intestin court (trouble de malabsorption dû à une ablation intestinale) ou bien à une infection bactérienne ou parasitaire [10]. Il est maintenant bien établi que les individus

déficients en lactase tolèrent mieux le lactose du yaourt [184], comparé à d'autres produits laitiers non fermentés [185]. Les explications avancées sont indiquées dans la suite de ce paragraphe [10].

- Pendant la fermentation du lait, le lactose est partiellement hydrolysé, ce qui réduit très significativement la quantité de lactose dans le yaourt par rapport au lait.
- La  $\beta$ -galactosidase des bactéries lactiques reste fonctionnelle dans le produit fermenté et dans l'intestin dont la température et le pH sont compatibles avec l'activité enzymatique. Cette hypothèse implique que les bactéries lactiques consommées avec le yaourt resteraient enzymatiquement actives et amélioreraient la dégradation du lactose après ingestion du produit lacté [174].
- Une étude sur un modèle souris montre que certaines bactéries lactiques (entre autre celles du yaourt) stimulent l'activité lactase des cellules épithéliales intestinales : le suivi des effets de l'apport (i) de lait fermenté par *Lactobacillus casei* DN-114 001 ou (ii) de yaourt sur l'épithélium de l'intestin grêle de souris, a montré une augmentation de l'activité lactase au niveau de la bordure en brosse des cellules [186].

Certaines cultures bactériennes semblent avoir plus d'activité protéolytique que d'autres. Dans le yaourt, *L. delbrueckii ssp. bulgaricus* possède une activité protéolytique plus importante pendant la fermentation du lait et le stockage que celle observée pour *S. thermophilus*. Cela a été étudié par le suivi de l'augmentation de la concentration en peptides et acides aminés libres après la fermentation du lait par chacune des deux espèces séparément [187].

### ***III.C Adaptation et persistance des bactéries lactiques des produits fermentés dans le tractus digestif***

Les lactobacilles constituent une partie de la flore microbienne du gros intestin et de l'intestin grêle.

La survie de ces bactéries lactiques dans l'environnement gastrique est entre autre influencée par le

pH acide, l'exposition aux enzymes digestives et aux sels biliaires [188]. La capacité de survie à ces conditions dépend des souches [10]. Dans un groupe de patients souffrant de gastrite atrophique et d'hypochlorhydrie, *L. gasseri* a survécu au passage à travers le tractus gastro-intestinal alors que *S. thermophilus* et *L. bulgaricus* sont inactivés [189]. L'espèce *Bifidobacterium* a été détectée dans les selles jusqu'à 8 jours après son ingestion [190]; elle survivrait donc au passage gastrique.

### ***III.D Propriétés immuno-modulatoires des bactéries lactiques alimentaires***

L'intérêt porté aux bienfaits des bactéries fermentaires des produits laitiers sur la santé humaine a commencé il y a plus d'une centaine d'années [101]. Au cours des dernières décennies, un intérêt particulier s'est développé pour l'isolement de souches qui exercent des effets immuno-modulateurs *in vitro* ou *in vivo* en plus de leurs capacités fermentaires et technologiques. Cet intérêt s'est traduit par l'identification de nouveaux probiotiques. La plupart des études ont porté sur les effets immuno-modulateurs des bactéries lactiques dans l'intestin [10]. Quelques exemples emblématiques de ces études sont présentés ci-dessous.

#### *Effets immuno-modulatoires de Lactobacillus plantarum*

*L. plantarum* est une bactérie utilisée dans l'industrie dans de nombreuses applications : fermentation de produits alimentaires en saumure (choucroute, olives, cornichons), fermentation du lait pour la production de certains fromages (fromage à pâte pressée cuite, Feta, etc.) ainsi que pour la fabrication de certains produits de charcuterie comme les saucissons secs. En plus de ses propriétés fermentaires, plusieurs études établissent les effets immuno-modulateurs de différentes souches de *L. plantarum*. La souche *L. plantarum* ACA-DC146, inhibe tout à la fois l'adhésion d'*E. coli* et de *Salmonella typhimurium* aux cellules Caco-2 et induit une importante sécrétion de cytokines pro-inflammatoires (IL-12, TNF- $\alpha$ , IFN- $\gamma$ ) par les cellules PBMC comparée à d'autres souches d'espèces *L. casei*, *L. paracasei* [191]. L'administration orale de *L. plantarum* K68 aux

souris Balb/c avec une colite induite par le DSS, réduit la production de cytokines pro-inflammatoires (TNF- $\alpha$ , IL- $\beta$  et IL-6) [192]. La souche *L. plantarum* 299v atténue l'inflammation colique dans un modèle de souris déficientes en IL-10, par la diminution de la production d'IL-12, d'IFN- $\gamma$  et de l'immunoglobuline G2a pour la muqueuse [193]. Chez des fumeurs, l'ingestion de cette même souche agit comme un agent cardio-protecteur en diminuant la concentration de l'insuline, de leptine et d'IL-6 dans le plasma [194]. *L. plantarum* Lp91 module l'expression de gènes codant pour des cytokines pro-inflammatoires (baisse de l'expression de TNF- $\alpha$  et COX2) et anti-inflammatoires (augmentation de l'expression de IL-10, IL-4, IL-6 et COX1) dans un modèle de souris de colite induite par le TNBS [195]. Enfin, la souche probiotique, *L. plantarum* LP-onlly atténue l'inflammation dans un modèle de colite chez la souris comme l'atteste la baisse du score d'inflammation et des dommages histologiques [196].

#### Effets immuno-modulateurs de *Lactobacillus helveticus*

*Lactobacillus helveticus* est un ferment industriel le plus souvent utilisé dans la fermentation du lait pour la production fromagère. Plusieurs souches présentent des effets immuno-modulateurs établis [197]. Citons quelques une des études. *L. helveticus* MIMLh5 testée pour ses effets immuno-modulateurs *in vivo* et *ex-vivo*, stimule la production du TNF- $\alpha$  et de IL-2 par les cellules dendritiques de la moelle osseuse de souris et réduit l'activité du facteur transcriptionnel NF- $\kappa$ B (réduction de l'expression des cytokines pro-inflammatoire TNF- $\alpha$ , IL-8 et IL-6) dans la lignée cellulaire hypopharyngienne humaines FaDU. La souche *L. helveticus* HY7801 réduit l'activité du NF- $\kappa$ B ainsi que l'expression des cytokines pro-inflammatoires chez les souris infectées par *Candida albicans*. Le lait fermenté par *L. helveticus* R398 induit une augmentation de l'IL-10 et une diminution de l'IL-6 dans le sérum de souris et il stimule la production de cellules sécrétrices d'IgA dans l'intestin grêle de souris. Quant à la fraction peptidique du lait fermenté avec *L. helveticus* R389, elle induit une réponse Th2 (induisant d'importantes quantités d'IL-4 dans le sérum) *in vivo*

dans un modèle murin. Enfin, l'ingestion du lait fermenté par la souche *L. helveticus* M92 augmente la concentration d'IgA sécrétoire dans les fèces et dans le sang des souris.

#### Effets immuno-modulateurs de *Streptococcus thermophilus*

*S. thermophilus* est surtout connue comme ferment pour la production de yaourt (en association avec *L. bulgaricus*). Des effets immuno-modulateurs sont décrits pour plusieurs souches de cette espèce.

*S. thermophilus* THS utilisée dans la fabrication du yaourt induit une importante production d'IL-12 pro-inflammatoire par des macrophages, plus importante que celles de différentes souches de lactobacilles, bifidobactéries, Lactocoques, Leuconostokes et Propionibactéries. Le rapport IL-10 sur IL-12 (qui renseigne sur le pouvoir anti-inflammatoire) obtenu après stimulation des macrophages avec *S. thermophilus* THS est relativement faible par rapport à ceux obtenus pour les autres souches étudiées. De plus, cette souche stimule l'expression de SOCS3 (suppressor of cytokine signalling 3) qui contrôle l'expression de gène codant pour des cytokines pro-inflammatoires [198]. La consommation de lait fermenté par *S. thermophilus* CRL1190 prévient l'apparition de gastrite (maladie inflammatoire de la paroi de l'estomac) chez la souris Balb/c par modulation de la réponse immunitaire et le maintien de la couche de mucus intestinale [199].

#### Effets immuno-modulateurs des Propionibactéries

Les propionibactéries sont des bactéries fermentaires retrouvées dans le lait et les produits laitiers, principalement connues pour leurs utilisations dans l'industrie fromagère et plus particulièrement dans la production de fromage de type suisse comme l'emmental. Plusieurs études concluent aux propriétés immuno-modulatrices des Propionibactéries. L'évaluation des effets anti-inflammatoires de 10 souches fermentaires de Propionibactéries (*P. freudenreichii* CIRM-BIA1, CIRM-BIA118, CIRM-BIA456, ITGP18, ITGP20, SI48, SI41, LSIP11, et LSIP23 et *P. jensenii* CIRM-BIA455) sur les cellules PBMC a montré une induction de l'IL-10 anti-inflammatoire variable suivant les

souches. Parmi les dix souches sélectionnées les deux souches *P. freudenreichii* BIA118 et SI48 montrent des effets protecteurs dans deux modèles de colites chez la souris [200].

*P. freudenreichii* ET-3, produit le DHNA (1,4-Dihydroxy-2-naphthoic acid), précurseur de la ménaquinone (Vitamine K<sub>2</sub>). Ce composé stimule la croissance des bifidobactéries dans l'intestin humain [201] et atténue l'inflammation en supprimant l'expression de cytokines pro-inflammatoires [202] dans un modèle de souris déficientes en IL-1 et *in vitro* sur une lignée cellulaire de macrophages. L'ingestion orale de DHNA produit par *P. freudenreichii* par des souris ayant une colite induite au DSS améliore le score histologique, le taux de survie des souris et diminue l'expression de cytokines pro-inflammatoires comme l'IL-6, l'IL-1 $\beta$  et le TNF- $\alpha$  [203].

#### Effets immuno-modulateurs de *Lactobacillus casei*

*L. casei* est une bactérie lactique isolée entre-autres du lait et du fromage. Cette espèce est utilisée comme levain dans les fermentations alimentaires (fromages, olives). Les effets immuno-modulateurs observés avec certaines souches sont présentés ci-dessous.

*L. casei* DN-114 001 survit au transit intestinal [204] et stimule le système immunitaire durant le transit en interagissant avec les plaques de Peyer, en stimulant la production d'IgA, de cellules CD4+ ainsi que d'anticorps spécifiques de la souche stimulatrice [205]. *L. casei* DN-114 001 interagit aussi avec la muqueuse intestinale humaine et peut sensiblement diminuer la libération des cytokines (IL-8, TNF- $\alpha$ ) pro-inflammatoires chez les malades de Crohn [206,207]. *L. casei* Shiota ACA-DC 6002 inhibe l'adhésion d'*E. coli* et de *S. typhimurium* aux cellules Caco-2 et induit une importante sécrétion de la cytokine anti-inflammatoire IL-10 par les cellules PBMC par rapport à d'autres souches des espèces *L. plantarum* et *L. paracasei* [191].

#### Effets immunomodulateurs de bactéries lactiques en mélange

- Les effets de la consommation de lait fermenté par *L. casei*, *L. acidophilus* ou deux bactéries, sur les systèmes de défenses spécifiques et non spécifiques de la souris ont été étudiés. La

consommation du lait fermenté par l'une des bactéries ou les deux pendant 8 jours augmente l'activité phagocytaire des macrophages péritonéaux ainsi que l'activité lymphocytaire [208].

- *L. acidophilus*, *L. rhamnosus* et *L. delbrueckii*, limitent la production RANTES (un membre de la superfamille de cytokines IL-8) et d'IL-8, par la lignée de cellules épithéliales intestinale HT-29 [209]. Rappelons qu'une forte production de ces chimiokines exacerbe la réponse inflammatoire [10].
- L'association de *L. heveticus* MIMLh5 et de *S. salivarius* ST3 stimule la réponse immunitaire innée des macrophages U937 humains *in vitro* [197].
- L'association de *L. heveticus* R0052, *B. longum* et *S. thermophilus*, utilisées pour la fermentation de boissons à base de soja ou de lait de vache, induit une baisse de l'expression de IL-8 par les lignées cellulaires épithéliales HT-29 et T84 préalablement stimulées par la cytokine pro-inflammatoire TNF- $\alpha$  [197].
- La consommation de yaourt contenant *L. bulgaricus* et *S. thermophilus* vivants pendant 4 mois augmente la production d'IFN- $\gamma$ , une cytokine ayant pour rôle de défendre l'organisme vis à vis de pathogènes [10].

### ***III.E Effets anti-inflammatoires de L. delbrueckii***

*L. delbrueckii* est utilisé dans l'industrie laitière pour ses propriétés fermentaires. La sous-espèce *bulgaricus* est utilisée pour la production de yaourt alors que la sous-espèce *lactis* est traditionnellement utilisée pour la fabrication de fromages du type parmesan.

L'utilisation de *L. delbrueckii* ssp. *lactis* comme ferment pour la fabrication des fromages à pâtes dures comme l'Emmental, le Comté et le Beaufort (328 tonnes de fromage à pâte cuite ont été fabriquées en France en 2012 par 232 industries laitières) [210], devient de plus en plus répandue du fait de son importante capacité à transformer le galactose en acide lactique pendant l'étape de

pressage du fromage. La consommation de ce type de fromage apporte ainsi entre  $10^5$  et  $10^9$  *L. delbrueckii* par jour et par personne.

En plus de cette importante consommation de fromage en France, 1,5 millions de tonnes de yaourt et autres laits fermentés sont produites annuellement. La consommation est estimée à 60 grammes consommés par personne et par jour. Cela correspond à une consommation de  $10^9$  à  $10^{10}$  bactéries de *L. delbrueckii* par jour et par personne [en considérant qu'un yaourt contient  $10^7$  bactéries de ce type par gramme (concentration minimal légale suivant le Décret n°88-1203 du 30 décembre 1988 relatif aux laits fermentés et au yaourt, version consolidée au 03 avril 1997)].

Par conséquent, entre la consommation de fromages et de yaourts, les français ingèrent quotidiennement de grandes quantités de *L. delbrueckii* d'où l'intérêt d'étudier les effets immuno-modulateurs éventuels des ferments initialement sélectionnés pour leurs propriétés technologiques. Plusieurs études apportent des arguments en faveur d'effets probiotiques de *L. delbrueckii*; nous les résumons ci-dessous.

En 1908, Metchnikoff a émis l'hypothèse que les bactéries du yaourt, *L. delbrueckii* et *S. thermophilus*, exerceraient des effets bénéfiques sur la santé du consommateur [101]. Le produit probiotique VSL#3 contenant un mélange de six bactéries lactiques dont *L. delbrueckii* améliore les symptômes de colite ulcéreuse dans des essais cliniques chez l'homme [211, 212]. *L. delbrueckii* appartient au groupe acidophilus de *Lactobacilli*. Ce groupe de bactéries présente une grande similarité génomique et contient de nombreuses espèces décrites et utilisées comme probiotiques, par exemple, *L. acidophilus*, *L. johnsonii* et *L. gasseri*. La souche *L. delbrueckii ssp. bulgaricus* CRL423 affecte la production de cytokines dans l'intestin grêle de souris [213] et la souche *L. delbrueckii ssp. bulgaricus* OLL1181 inhibe le développement de colite induite par le DSS chez la souris par l'induction de la voie AhR (aryl hydrogen receptor), une voie de régulation de l'inflammation dans le gros intestin [214].

L'importance du choix de la souche de *L. delbrueckii* étudiée pour l'effet immuno-modulateur est devenue une évidence suite à des études de génomes de souches de *L. bulgaricus* [167,215–218] qui révèlent un génome en évolution qui tend à une spécialisation et à une adaptation de la bactérie à sa niche écologique. Toutes ces observations ont été le moteur pour une étude réalisée sur les propriétés immuno-modulatrices de différentes souches *L. delbrueckii* [219].

### III.E.1. Effets anti-inflammatoires de *L. delbrueckii*

Une collection de 57 souches de *L. delbrueckii* a été testée *in vitro* en mesurant l'activité du facteur transcriptionnel NF- $\kappa$ B (nuclear factor kappa B), un régulateur majeur de la réponse inflammatoire, dans la lignée de cellules épithéliales intestinales humaines HT-29 [219]. En condition où le facteur transcriptionnel NF- $\kappa$ B est induit par la cytokine pro-inflammatoire TNF- $\alpha$ , l'ajout simultané de *L. delbrueckii* réduit l'activité de NF- $\kappa$ B par rapport au contrôle positif sans ajout de bactéries (Figure 10). L'effet anti-inflammatoire dépend des souches : les souches de *L. delbrueckii*, CNRZ327 et CNRZ333 présentent un effet anti-inflammatoire important (~70%) alors que d'autres comme ATCC11842, exercent un effet anti-inflammatoire plus modéré (~50%) à faible (~20%).



**Figure 10. Effets des souches de *L. delbrueckii* sur l'activité du facteur transcriptionnel NF- $\kappa$ B, induite par le TNF- $\alpha$  d'après [219].** Les barres indiquent l'activité du gène rapporteur luciférase sous contrôle d'un promoteur inducible par NF- $\kappa$ B dans la lignée de cellules épithéliales intestinales HT-29 -, correspond au contrôle négatif (sans TNF- $\alpha$  et sans bactérie) ; T, correspond au contrôle positif (avec TNF- $\alpha$  et sans bactérie). Les autres barres présentent

l'activité NF- $\kappa$ B en présence de souches de *L. delbrueckii* (indiquée en abscisse) et du TNF- $\alpha$ . Les barres présentant un écart type correspondent à la moyenne entre trois à cinq mesures indépendantes de l'activité NF- $\kappa$ B. Les autres barres représentent une mesure ou moyenne de deux mesures. Les barres noires correspondent aux mesures d'activité obtenues avec *L. delbrueckii* ssp. *lactis*, les barres grises correspondent aux mesures d'activité obtenues avec *L. delbrueckii* de la ssp. *bulgaricus*. Les barres hachurées correspondent à des *lactobacilles* commensaux utilisés comme contrôles positifs.

Parmi les 37 souches de *L. delbrueckii* ssp. *bulgaricus* et les 20 souches de *L. delbrueckii* ssp. *lactis* testées, celles qui présentent l'effet anti-inflammatoire le plus élevé (CNRZ327 et CNRZ333) appartiennent à la sous-espèce *lactis*, alors que les souches avec les plus faibles effets anti-inflammatoires appartiennent aux deux sous-espèces. La majorité des souches de *L. delbrueckii* testées répriment l'activité NF- $\kappa$ B de façon plus importante que les souches probiotiques *L. salivarius* Ls33, *L. acidophilus* NCFM et *L. gasseri* ATCC 33323. L'effet des CNRZ327 et CNRZ333 n'est pas dû à leur activité protéolytique qui aurait pu dégrader le TNF- $\alpha$  [219], puisqu'il est observé avec différents activateurs de NF- $\kappa$ B. *L. delbrueckii* bloque l'activation de NF- $\kappa$ B en réduisant la phosphorylation d'I $\kappa$ B, empêchant ainsi la dégradation de cet inhibiteur de NF- $\kappa$ B [219]. Les résultats suivants suggèrent que des protéines de surface des souches seraient responsables de l'effet anti-inflammatoire.

- l'effet anti-inflammatoire a été observé en utilisant des bactéries en phase exponentielle de croissance qui ont été lavées avant la mise en contact avec les cellules HT-29; ces conditions devraient éliminer les composés éventuellement libérés par une lyse bactérienne;
- Le surnageant de culture n'a pas d'effet anti-inflammatoire dans le même test d'activité NF- $\kappa$ B;
- L'effet immuno-modulateur est perdu après un traitement des bactéries à la trypsine éliminant les protéines exposées à la surface suivi par des lavages.

Des études *in vivo* ont montré que la souche CNRZ327 réduit significativement les dommages induits par le DSS chez la souris, un modèle de colite induite montrant des similarités avec la CU chez l'homme [219, 220].

#### **IV. Approches « omiques » pour l'identification d'effecteurs probiotiques**

Comme discuté dans les chapitres précédents, les bactéries probiotiques peuvent avoir différents effets bénéfiques sur la santé. Bien que certains de ces effets soient bien établis, les effecteurs bactériens qui les produisent restent pour la plupart inconnus. L'approche classique de recherche d'effecteurs bactériens à l'origine de l'effet probiotique de certaines bactéries consiste à procéder par élimination d'hypothèses préalablement établies. Pour identifier la nature l'effecteur bactérien (protéine, sucre, lipide, etc), sa localisation cellulaire (cytoplasmique, membranaire, sécrété, exposé à la surface), ses interactions potentielles avec d'autres molécules, les voies de signalisation impliquées dans son effet, etc, un plan expérimental est établi afin de caractériser la molécule recherchée en répondant progressivement à chacun des aspects.

Avec les approches « omiques », l'échelle des données analysables ainsi que la méthodologie de recherche ont considérablement changées. Il s'agit de techniques à haut débit qui génèrent d'importantes quantités de données aux différents niveaux biologiques de la séquence des gènes à la traduction en protéines, permettant d'étudier différents niveaux et leurs régulations qui conditionnent le fonctionnement des systèmes organiques [221].

Ainsi, pour la recherche d'effecteurs bactériens par des approches omiques, les techniques de séquençage et d'étude du transcriptome et du protéome, permettent de rassembler toutes les données génomiques et protéomiques des organismes étudiés, de traiter ces données (à travers des études comparatives par exemple) dans le but d'identifier les premières indications permettant de définir un ensemble de gènes candidats avant de confirmer expérimentalement leurs effets.

## ***IV.A Génomique comparative, corrélation entre phénotypes et présence ou absence de gènes***

La génomique comparative est une approche de comparaison de génomes ou de leur contenu en gènes qui peut être appliquée à des micro-organismes phylogénétiquement proches ou plus éloignés (par exemple comparaison entre souches d'une même espèce ou de différentes espèces). L'objectif de l'étude comparative peut être une meilleure compréhension de l'évolution des espèces ou la mise en évidence de gènes d'intérêt dont la présence est corrélée à un phénotype donné.

La première séquence complète d'un génome bactérien a été publiée en 1995 [222]. 15 ans après, plus de 1000 génomes bactériens étaient complètement séquencés [223]. Actuellement, les séquences génomiques complètes de plus de 2300 bactéries sont accessibles dans des bases de données publiques (source <http://mbgd.genome.ad.jp/>).

### **IV.A.1. Génomique comparative et identification d'effecteurs bactériens potentiellement à l'origine des effets probiotiques**

Dans l'étude des bactéries probiotiques, la génomique comparative permet d'identifier des gènes potentiellement impliqués dans l'adaptation de la bactérie aux conditions environnementales du tractus intestinal, dans le maintien du probiotique dans l'intestin ainsi que dans le contact avec les cellules de l'hôte pour induire les effets bénéfiques définis.

L'étude des séquences génomiques de lactobacilles probiotiques a permis d'identifier certaines caractéristiques de ce groupe bactérien généralement corrélés à leur niche écologique et à l'adaptation à cet environnement. Le génome des lactobacilles a ainsi révélé un important degré d'auxotrophie (incapacité d'un organisme vivant à synthétiser un composé organique nécessaire à son développement) pour les acides aminés par exemple ainsi que pour d'autres composés nécessaires à la cellule bactérienne. Cette auxotrophie est compensée par la présence d'un large

nombre de transporteurs permettant d'acquérir ces molécules à partir du milieu extracellulaire [119].

La comparaison entre génomes de lactobacilles isolés du tractus digestif et ceux isolés des plantes ou du lait permet d'identifier des gènes dont la distribution diffère entre ces deux types isolats.

Ainsi, les génomes de *L. bulgaricus* et *L. helveticus* [167,218,224], qui sont toutes deux adaptées au milieu lait, contiennent plusieurs fragments de gènes (pseudogènes) dont les produits contribuent à l'utilisation de divers carbohydrates. Cela s'explique par leur adaptation au milieu lait riche en lactose. Inversement, les lactobacilles isolés du tractus digestif comme *L. plantarum WCFS1* [225], possèdent généralement un important nombre de gènes dont les produits permettent l'internalisation et l'utilisation des sucres du milieu extracellulaire [218]. Le profil de présence ou absence d'enzymes nécessaires à la survie dans l'intestin comme l'hydrolase de sels biliaires dépend de la niche écologique du lactobacille considéré [226]. L'abondance des gènes codant pour des protéines de surface bactérienne qui permettent la fixation au mucus (mucus-binding proteins) semblent également corrélés à la persistance dans le tractus digestif de certains lactobacilles probiotiques [227–229].

La caractérisation des propriétés de surface cellulaires chez les probiotiques a été au centre des études de génomiques comparatives vu qu'elles jouent probablement un rôle clef dans l'interaction avec l'hôte. La variabilité de composition et de structure de la surface bactérienne conduit à des propriétés différentes suivant les espèces voire les souches qui pourraient expliquer les effets immuno-modulateurs spécifiques de certains probiotiques. J'indique ci-dessous deux exemples d'effecteurs bactériens de surface initialement identifiés par des études de génomique comparative.

L'étude de génomique comparative entre les deux lactobacilles phylogénétiquement proches, *L. rhamnosus GG* (la souche probiotique la plus étudiée de cette espèce) et *L. casei* (souche alimentaire non adapté au tractus digestif) visait à identifier les déterminants génétiques

pouvant expliquer l'adaptation de *L. rhamnosus* à la niche intestinale. Malgré la grande similarité génomique des deux espèces, le génome de *L. rhamnosus* GG contient plusieurs centaines de gènes qui n'ont pas d'homologue chez *L. casei*. Parmi ces gènes, un opéron codant pour des structures similaires aux pili (pili-like operon) a été identifié [119]. Le rôle joué par ces pili dans l'adhésion aux cellules épithéliales et dans l'immuno-modulation a ensuite été confirmé expérimentalement [222, 223].

L'étude de génomique comparative entre *L. johnsonii* NCC533 (souche probiotique) et *L. gasseri* avait pour but d'identifier les déterminants génétiques de *L. johnsonii* NCC533 pouvant expliquer les interactions avec les cellules épithéliales [232]. Malgré la grande similarité de séquence entre les deux génomes bactériens (>95%), le génome de *L. johnsonii* NCC533 code pour une structure fimbriale (long filaments) qui pourrait avoir un rôle dans l'adhésion aux cellules épithéliales [232]; Cette hypothèse n'a pas encore été confirmée expérimentalement.

En conclusion, on peut dire que la combinaison d'approches de génomique comparative pour identifier des gènes dont la présence diffère entre organismes de phénotypes différents et d'analyse de séquence de gènes candidats pouvant indiquer une fonction potentielle moléculaire est une bonne stratégie pour identifier les mécanismes moléculaires à l'origine du phénotype probiotique de certaines bactéries d'intérêt.

#### ***IV.B Transcriptomique comparative, corrélation entre phénotype probiotique et expression de gènes***

Le transcriptome est défini comme l'ensemble des ARNm. A partir des données génomiques, l'approche de transcriptomique comparative consiste à comparer les profils d'expression de gènes entre espèces, sous-espèces ou souches et de corrélérer les niveaux d'expression avec le phénotype

étudié. Soulignons ici que cette approche ne permettra d'étudier que les gènes présents dans les bactéries considérées.

La puce à ADN correspond à l'une des techniques permettant d'étudier le transcriptome. Cette technique permet de quantifier le niveau d'expression relatif d'un grand nombre de gènes simultanément dans la cellule.

#### **IV.B.1. Transcriptomique comparative et identification d'effecteurs bactériens potentiellement à l'origine des effets probiotiques**

L'approche transcriptomique pour l'étude du mode d'action des probiotiques a permis de confirmer l'adaptation des bactéries à leur niche écologique et d'identifier des gènes dont l'expression est corrélée à l'interaction avec l'hôte. Ci-dessous sont présentés d'importants exemples d'études transcriptomiques à ce sujet.

La persistance dans le tractus digestif de souris des souches de *L. johnsonii* NCC533 et ATCC33200 diffère significativement (12 versus 5 jours). L'hypothèse est que les gènes de *L. johnsonii* par NCC533 spécifiquement exprimés durant son passage dans le tractus digestif seraient impliqués dans la meilleure persistance de cette souche par rapport à ATCC33200. L'étude du transcriptome de chacune des souches pendant le passage dans le tractus digestif de souris monocolonisées a révélée 233 gènes par NCC533 exclusivement exprimés. Une deuxième étude comparant le transcriptome de NCC533 en croissance *in vitro* ou *in vivo*, pendant le passage dans le tractus digestif de souris monocolonisées, a révélé 174 gènes fortement exprimés dans le jéjunum des souris. La combinaison des deux études révèle 3 loci spécifiquement exprimés par NCC533 et fortement exprimés *in vivo* [233,234]. Les fonctions codées par ces loci, pourraient être à l'origine de la persistance de *L. johnsonii* NCC533 dans le tractus digestif de souris.

La comparaison des transcriptomes de *L. plantarum* WCFS1 dans le cæcum de souris monocolonisée et en croissance *in vitro* a révélé la surexpression de gènes impliqués dans le transport et le métabolisme des carbohydrates ainsi que la répression de certains gènes codant pour des fonctions de surface bactérienne comme les gènes de D-alanylation de l'acide lipoteichoïque qui médie l'effet sur le système immunitaire de l'hôte [235]. La première étude du profil d'expression de gènes d'une bactérie probiotique dans l'intestin humain a été celle de la souche *L. plantarum* 299v. Le transcriptome a été réalisé à partir de bactéries isolées de biopsies de la muqueuse du gros intestin de patients atteints de cancer du colon ayant consommé volontairement cette bactérie probiotique. Comparé au transcriptome de *L. plantarum* 299v dans des conditions de croissance *in vitro*, le transcriptome de la souche réalisé à partir des biopsies révèle un profil d'expression de gènes spécifique à l'environnement intestinal humain. Dans l'intestin humain, *L. plantarum* 299v adapte tout particulièrement ses capacités métaboliques pour l'acquisition de carbohydrates, l'expression des exopolysaccharides et des composés protéiques de la surface cellulaire [236].

La comparaison entre le transcriptome de *L. plantarum* WCFS1 dans le modèle de souris monocolonisées [235] avec celui de *L. plantarum* 299v dans les biopsies du gros intestin humain [236] a permis d'établir un core transcriptome de *L. plantarum* dans les intestins de mammifères [236]. Cette étude transcriptomique comparative a identifié les gènes potentiellement impliqués dans l'adaptation de *L. plantarum* à l'écosystème intestinal et fourni de nouvelles cibles pour l'analyse moléculaire des interactions bactérie-hôte qui affectent la santé humaine [236].

#### ***IV.C Protéomique comparative et corrélation entre phénotype probiotique et protéome***

Les approches de génomique et de transcriptomique se basent respectivement sur la composition en gènes et leur niveau d'expression. L'approche de protéomique se base sur le contenu en protéines et le protéome est défini comme l'ensemble des protéines d'un organisme donné. Plusieurs études protéomiques ont été réalisées dans le but d'identifier les propriétés permettant aux probiotiques de persister dans le tractus intestinal et d'interagir avec l'hôte pour exercer leurs effets bénéfiques [237, 238].

La majorité de ces études s'est intéressée aux protéines extracellulaires considérées comme des facteurs clé dans le dialogue bactérie-hôte du fait de leurs possibles interactions directes avec les cellules de la muqueuse intestinale.

##### **IV.C.1. Protéomique comparative et identification d'effecteurs bactériens potentiellement à l'origine des effets probiotiques**

Comme c'est le cas pour les études génomiques et transcriptomiques, la protéomique permet de mieux comprendre les mécanismes moléculaires et d'identifier des voies métaboliques activées ou désactivées au cours de l'adaptation de la bactérie à la niche intestinale. Des études protéomiques ont mis en évidence des modes d'action partagés par différentes espèces bactériennes pour s'adapter aux conditions de stress rencontrées dans la niche intestinale : (1) la surexpression des protéines chaperons (famille Clp, complexe GroES/GroEL, DnaK, et DnaJ) permettant de contrecarrer le repliement anormal de protéines, et leur dénaturation; (2) l'augmentation du nombre de protéines impliquées dans le métabolisme des carbohydrates afin d'augmenter la production d'énergie; (3) l'augmentation de l'activité de l'enzyme  $F_0-F_1$  ATPase assurant l'homéostasie du pH cytoplasmique chez les micro-organismes anaérobies en facilitant l'extrusion des proton; (4) la modification du

niveau d'expression de plusieurs protéines impliquées dans la transcription (RNA polymérase), la traduction (facteurs d'élongation et protéines ribosomales), dans la biosynthèse de nucléotides et d'acides aminés et dans la réparation de l'ADN [239, 240].

L'approche de protéomique a également révélé des stratégies spécifiques de certaines espèces pour faire face aux conditions de stress [237]. Parmi ces études, l'analyse du protéome de *L. rhamnosus* GG dans les conditions de stress acide *in vitro*, met en évidence une inhibition de la glutamine synthétase ainsi qu'une importante phosphorylation de différentes protéines impliquées dans le métabolisme des carbohydrates [241]. Chez les bactéries du genre lactobacillus, les analyses protéomiques suggèrent que le contrôle intracellulaire du pH peut se faire par l'activation de la voie de l'arginine déiminase [237].

Revenons à *L. rhamnosus* comme exemple d'étude d'effets probiotiques par protéomique comparée [242] entre la souche probiotique *L. rhamnosus* GG isolée du TD et la souche Lc705 isolée de produits laitiers. La comparaison des données protéomiques a révélé plus de 90 protéines de *L. rhamnosus* GG et 150 protéines dans Lc705 qui ne présentent pas d'homologue dans l'autre souche. Ces protéines différenciellement exprimées sont impliquées dans la formation de biofilm, le fonctionnement de phages, le remodelage de la paroi cellulaire bactérienne et l'immuno-modulation [242].

En conclusion, l'approche protéomique donne des résultats prometteurs pour la compréhension des mécanismes impliqués dans les effets probiotiques. Les protéines exposées à la surface bactérienne sont particulièrement importantes à étudier, vu qu'elles représentent la première ligne d'interaction avec les cellules de la muqueuse. Mis à part quelques exceptions [243], le rôle des protéines sécrétées dans le dialogue probiotique-hôte n'a pas été assez étudié alors qu'elles pourraient être des facteurs centraux dans ces interactions.

L'approche protéomique peut être combinée à une ou plusieurs autres approches « omiques » pour étudier les effets probiotiques. La combinaison entre différentes approches « omiques » permet d'observer le comportement du probiotique face à son hôte dans sa globalité : (1) la génomique permet d'identifier le répertoire génétique de la bactérie probiotique et de corrélérer le profil de présence/absence de gènes au caractère bénéfique du micro-organisme; (2) la transcriptomique mesure le niveau d'expression des gènes du probiotique en contact avec les cellules de l'hôte, ce qui permet de corrélérer la surexpression ou l'inhibition de certains gènes avec cette interaction; (3) l'approche protéomique notamment l'étude des protéines exposées à la surface et sécrétées permet d'identifier des protéines spécifiques aux espèces probiotiques et susceptibles d'interagir directement avec l'hôte pour induire des effets bénéfiques.

# **V. Résultats**

## **V.A Genome sequence of *Lactobacillus delbrueckii* ssp. *lactis* CNRZ327, a dairy bacterium with anti-inflammatory properties.**

### Introduction

*L. delbrueckii* est l'une des espèces les plus utilisées dans l'industrie laitière. *L. delbrueckii* ssp. *lactis*, est utilisée comme ferment dans la production de certains fromages à pâte dure, d'un grand intérêt économique dans l'industrie fromagère. Des études récentes ont montré que certaines souches de cette ssp. possèdent des effets anti-inflammatoires *in vitro* et *in vivo*. C'est notamment le cas de la souche CNRZ327 qui est apparue parmi les plus immunodulatrices des 57 souches de *L. delbrueckii* testées par Santos-Rocha et al [219]. Un/des composés de nature protéique et exposé/s à la surface de ce lactobacille est/sont impliqués dans l'effet immuno-modulateur. C'est pourquoi le génome de cette souche de *L. delbrueckii* ssp. *lactis* a été sélectionné pour séquençage avec une étape de finition afin d'étudier ensuite aussi précisément que possible le surfaceome de CNRZ327.

L'objectif de cette première partie de travail a été d'obtenir après finition, le génome complet à l'exception des séquences répétées de cette souche anti-inflammatoire afin d'avoir à notre disposition toute l'information génétique pour la recherche du/des effecteur/s bactérien/s immuno-modulateur/s. Rappelons qu'à ce jour, il n'existe aucun génome complet de la ssp. *lactis* dans les bases de données ce qui rajoute un intérêt au génome de cette souche.

**Genome sequence of *Lactobacillus delbrueckii* ssp. *lactis* CNRZ327, a dairy bacterium with anti-inflammatory properties.**

Running title: *Lactobacillus delbrueckii* ssp. *lactis* CNRZ327 genome.

Hela El Kafsi<sup>1,2,\*</sup>, Johan Binesse<sup>1,2,\*</sup>, Valentin Loux<sup>3</sup>, Julien Buratti<sup>3</sup>, Samira Boudebbouze<sup>1,2</sup>, Rozenn Dervyn<sup>1,2</sup>, Amal Hammani<sup>3</sup>, Emmanuelle Maguin<sup>1,2</sup>, and Maarten van de Guchte<sup>1,2,#</sup>

<sup>1</sup>INRA, UMR1319 Micalis, Jouy en Josas, France

<sup>2</sup>AgroParisTech, UMR Micalis, Jouy en Josas, France

<sup>3</sup>INRA, UR1077 Mathématique Informatique et Génome, Jouy en Josas, France

\*These authors contributed equally.

#Corresponding author

e-mail: [maarten.vandeguchte@jouy.inra.fr](mailto:maarten.vandeguchte@jouy.inra.fr)

## Abstract

*Lactobacillus delbrueckii* ssp. *lactis* CNRZ327 is a dairy bacterium with anti-inflammatory properties, *in vitro* and *in vivo*. Here we report the genome sequence of this bacterium, which appears to contain not less than 215 IS elements, an exceptionally high number regarding the strains' small genome size.

## Announcement

*Lactobacillus delbrueckii* is a member of the acidophilus complex, a group of lactobacilli related to *Lactobacillus acidophilus*. This group comprises a relatively high number of bacteria with alleged probiotic properties, particularly in immune modulation, mostly isolated from the human gut environment (*L. acidophilus*, *Lactobacillus johnsonii*, *Lactobacillus gasseri*). *L. delbrueckii*, rather known from its dairy applications in yogurt and cheesemaking, is an atypical member of this group (1, 2) characterized by a genomic GC content of about 50% as opposed to about 35% for the aforementioned bacteria, and its immune modulation potential has often been ignored. We recently identified *L. delbrueckii* ssp. *lactis* CNRZ327 as a strain with strong immune modulation potential, capable of down-regulating nuclear factor kappa B (NF- $\kappa$ B) activation in gut epithelial HT29 cells *in vitro*, and improving the *in vivo* symptoms of dextran sodium sulfate (DSS) induced colitis in mice, a model of human ulcerative colitis (3, 4).

Here, we report the genome sequence of *L. delbrueckii* ssp. *lactis* CNRZ327, which was determined to near completion: A draft sequence was generated by 454 paired-end sequencing (Roche Life Sciences) followed by sequence assembly using Newbler 2.6 (Roche). The resulting 33 scaffolds were ordered using Mauve aligner (5), with the earlier published genome sequence of *L. delbrueckii* ssp. *bulgaricus* ATCC 11842 (2) as the reference. The resulting scaffold order could in most cases be confirmed using polymerase chain reaction (PCR) amplification of scaffold-linking sequences, while the position and orientation of remaining scaffolds was determined using

multiplex PCR (6). 33 scaffolds and 33 PCR products together formed a circular chromosome. PCR product size and end-sequences typically converged to indicated the presence of a ribosomal RNA operon or an IS element (identified by a transposase-coding sequence) between the scaffolds, in which case no further sequencing of the PCR products was undertaken. Only if the product size indicated the presence of additional, possibly non-repeated, sequences, the PCR product was sequenced in order to obtain a complete overview of chromosomally encoded functions. The same approach was used to analyze the 128 gaps between contigs, within the scaffolds, of which 127 appeared to contain IS elements. A total of 1938538 bp of chromosomal sequence was assembled and the total size of the remaining gaps (containing repeated sequences) was estimated at 167 kbp, yielding a total genome size of approximately 2.1 Mbp. Sequence annotation was performed using the AGMIAL annotation platform (7), and protein localization was predicted using SurfG+ (8).

The *L. delbrueckii* ssp. *lactis* CNRZ327 genome appears to contain not less than 215 IS elements, an exceptionally high number regarding the strains' small genome size (9, 10) and compared to the 56 IS elements for the earlier reported genome of the *L. delbrueckii* ssp. *bulgaricus* strain ATCC11842 (2). In spite of this abundance of IS elements which could have resulted in massive gene inactivation and genomic rearrangements, strain CNRZ327 appears to have more extensive metabolic capacities than strain ATCC11842 and the genomes of the two strains appear to be essentially co-linear.

**Nucleotide sequence accession number.** The genome sequence of *L. delbrueckii* ssp. *lactis* CNRZ327 has been deposited at DDBJ/EMBL/GenBank under accession number xxxx.

**Acknowledgments.** This project was supported by INRA AIP Bio-Ressources and by a grant from the Région Ile-de-France. J. Binesse is supported by the project "Surfing" (ANR-2010-ALIA-016) financed by the French National Research Agency (ANR).

## References

1. **Nicolas P, Bessieres P, Ehrlich SD, Maguin E, van de Guchte M.** 2007. Extensive horizontal transfer of core genome genes between two *Lactobacillus* species found in the gastrointestinal tract. *BMC evolutionary biology* **7**:141.
2. **van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, Robert C, Oztas S, Mangenot S, Couloux A, Loux V, Dervyn R, Bossy R, Bolotin A, Batto JM, Walunas T, Gibrat JF, Bessieres P, Weissenbach J, Ehrlich SD, Maguin E.** 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proceedings of the National Academy of Sciences of the United States of America* **103**:9274-9279.
3. **Santos Rocha C, Gomes-Santos AC, Garcias Moreira T, de Azevedo M, Diniz Luerce T, Mariadassou M, Longaray Delamare AP, Langella P, Maguin E, Azevedo V, Caetano de Faria AM, Miyoshi A, van de Guchte M.** 2014. Local and systemic immune mechanisms underlying the anti-colitis effects of the dairy bacterium *Lactobacillus delbrueckii*. *PLoS One* **In press**.
4. **Santos Rocha C, Lakhdari O, Blottiere HM, Blugeon S, Sokol H, Bermu'dez-Humara'n LG, Azevedo V, Miyoshi A, Dore J, Langella P, Maguin E, van de Guchte M.** 2012. Anti-inflammatory properties of dairy lactobacilli. *Inflammatory bowel diseases* **18**:657-666.
5. **Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT.** 2009. Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics (Oxford, England)* **25**:2071-2073.
6. **Sorokin A, Lapidus A, Capuano V, Galleron N, Pujic P, Ehrlich SD.** 1996. A new approach using multiplex long accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and sequencing. *Genome research* **6**:448-453.
7. **Bryson K, Loux V, Bossy R, Nicolas P, Chaillou S, van de Guchte M, Penaud S, Maguin E, Hoebeke M, Bessieres P, Gibrat JF.** 2006. AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system. *Nucleic acids research* **34**:3533-3545.
8. **Barinov A, Loux V, Hammani A, Nicolas P, Langella P, Ehrlich D, Maguin E, van de Guchte M.** 2009. Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. *Proteomics* **9**:61-73.
9. **Robinson DG, Lee MC, Marx CJ.** 2012. OASIS: an automated program for global investigation of bacterial and archaeal insertion sequences. *Nucleic acids research* **40**:e174.
10. **Touchon M, Rocha EP.** 2007. Causes of insertion sequences abundance in prokaryotic genomes. *Molecular biology and evolution* **24**:969-981.

## Conclusion

L'étape de finition du génome de *L. delbrueckii* ssp. *lactis* CNRZ327 a révélé un génome d'une taille plus grande que celle déjà décrite pour la ssp. *bulgaricus*. La partie manquante du génome qui a été complétée grâce à l'étape de finition, est majoritairement composée de séquences répétées de type séquences d'insertions (IS).

Les résultats marquants de ce travail sont d'une part que le nombre total d'IS présentes chez CNRZ327 (215 IS) est important par rapport à la taille du génome et d'autre part, que l'abondance des IS dans ce génome de *L. delbrueckii* ssp. *lactis* est beaucoup plus élevé que tout ce qui a pu être observé dans les génomes complets de *L. delbrueckii* actuellement disponibles.

## **V.B *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus*: a chronicle of evolution in action.**

### Introduction

*Lactobacillus delbrueckii* ssp. *lactis* et *L. delbrueckii* ssp. *bulgaricus* sont deux bactéries lactiques largement utilisées en industrie laitière pour leurs propriétés fermentaires, respectivement dans la fabrication du fromage et du yaourt. L'étude antérieure du premier génome complet de la ssp. *bulgaricus*, a révélé un génome en phase active et rapide d'évolution, perdant des gènes superflus en adaptation à l'environnement lait de la bactérie.

Dans ce chapitre, nous étudions pour la première fois si la conclusion de l'étude antérieure s'applique à la ssp. *lactis* et discutons de la diversité inter- et intra- sous-espèces de *L. delbrueckii* dans le contexte de l'adaptation à l'environnement actuel de la bactérie.

***Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus*: a chronicle of evolution in action.**

Hela El Kafsi<sup>1,2</sup>, Johan Binesse<sup>1,2</sup>, Valentin Loux<sup>3</sup>, Julien Buratti<sup>3</sup>, Samira Boudebbouze<sup>1,2</sup>, Rozenn Dervyn<sup>1,2</sup>, Sean Kennedy<sup>1,2\*</sup>, Nathalie Galleron<sup>1,2\*</sup>, Benoît Quinquis<sup>1,2\*</sup>, Jean-Michel Batto<sup>1,2</sup>, Bouziane Moumen<sup>1,2</sup>, Emmanuelle Maguin<sup>1,2</sup>, and Maarten van de Guchte<sup>1,2,#</sup>

<sup>1</sup>INRA, UMR1319 Micalis, Jouy en Josas, France

<sup>2</sup>AgroParisTech, UMR Micalis, Jouy en Josas, France

<sup>3</sup>INRA, UR1077 Mathématique Informatique et Génome, Jouy en Josas, France

#Corresponding author

e-mail: [maarten.vandeguchte@jouy.inra.fr](mailto:maarten.vandeguchte@jouy.inra.fr)

**Background.** *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus* are lactic acid producing bacteria that are largely used in dairy industries, notably in cheese-making and yogurt production. An earlier in-depth study of the first completely sequenced ssp. *bulgaricus* genome revealed the characteristics of a genome in an active phase of rapid evolution, in what appears to be an adaptation to the milk environment. Here we examine for the first time if the same conclusions apply to the ssp. *lactis*, and discuss intra- and inter-subspecies genomic diversity in the context of evolutionary adaptation.

**Results.** Both *L. delbrueckii* ssp. show the signs of reductive evolution through the elimination of superfluous genes, thereby limiting their carbohydrate metabolic capacities and amino acid biosynthesis potential. In the ssp. *lactis* this reductive evolution has gone less far than in the ssp. *bulgaricus*. Consequently, the ssp. *lactis* retained more extensive carbohydrate metabolizing capabilities than the ssp. *bulgaricus* but, due to high intra-subspecies diversity, very few carbohydrate substrates, if any, allow a reliable distinction of the two ssp.. We further show that one of the most important traits, lactose fermentation, of one of the economically most important dairy bacteria, *L. delbrueckii* ssp. *bulgaricus*, relies on horizontally acquired rather than deep ancestral genes. In this sense this bacterium may thus be regarded as a natural GMO *avant la lettre*.

**Conclusions.** The dairy lactic acid producing bacteria *L. delbrueckii* ssp. *lactis* and ssp. *bulgaricus* appear to represent different points on the same evolutionary track of adaptation to the milk environment through the loss of superfluous functions, where the ssp. *bulgaricus* has progressed further away from the common ancestor.

### **Keywords**

*Lactobacillus delbrueckii, bulgaricus, lactis, genome, comparative genomics, adaptation, evolution.*

## Background

The thermophilic lactic acid producing bacterium *Lactobacillus delbrueckii* has a long history of application in dairy fermentations where the subspecies (spp.) *bulgaricus* is mainly known for its use in yogurt making while the ssp. *lactis* is traditionally used in the production of Parmesan and Emmental-type cheeses. The ssp. *bulgaricus* and *lactis* are historically distinguished on the basis of their capacity to metabolize different carbohydrates [1]. At the molecular level, a difference in the 16s rRNA sequence has been documented [2], and strains have been classified in either of the ssp. using Multi Locus Sequence Typing (MLST) [3]. The molecular basis for differing carbohydrate metabolism phenotypes has only in some cases been elucidated [4, 5].

The first complete genome sequence of *L. delbrueckii* ssp. *bulgaricus* (*L. bulgaricus*) revealed an ongoing adaptation to the protein-rich milk environment through the loss of superfluous amino acid biosynthesis functions, many of which are still recognizable in the form of pseudogenes while others have completely disappeared [6]. The nature of a number of other pseudogenes, remnants of genes involved in the transport and metabolism of various carbohydrates, pointed to an ancestor that most probably evolved in an environment where plant-carbohydrates were readily available. Evolutionary adaptation to the milk environment appears coherent with traditional yogurt-making practice, which involved the sequential transfer of samples of yogurt cultures to fresh milk. With the first records of yogurt (kisim) dating to 3200 before Christ [7], one would logically predict that *L. bulgaricus* adapted to this environment over time [6].

The ssp. *lactis* which phenotypically can be distinguished from the ssp. *bulgaricus* by its more extensive carbohydrate metabolizing capabilities, notably including the fermentation of various sugars of vegetal origin like maltose, mannose, saccharose, and trehalose [1], is less studied at the genomic level. We therefore sequenced the genomes of 6 *L. delbrueckii* strains and here present an analysis of these sequences and 4 earlier established genome sequences, together representing 5 ssp.

*bulgaricus* strains and 5 *ssp. lactis* strains, to further explore the differences between the two subspecies of this economically important bacterium, notably regarding their metabolic capacities and adaptation to the milk environment.

## Methods

**Bacterial strains.** The bacterial strains of which the genome sequences were determined in this work were obtained from the INRA collection and belong to the subspecies *L. delbrueckii ssp. bulgaricus* (strains Vib27 and Vib44) or *L. delbrueckii ssp. lactis* (strains CNRZ327, CNRZ333, CNRZ226, CNRZ700). For as far documented, the *ssp. bulgaricus* strains were originally isolated from yogurt while the *ssp. lactis* strains were derived from Emmental cheese or the starter cultures used to produce this type of cheese.

**Genome sequencing.** Genome sequences for 6 *L. delbrueckii* strains, 2 of the *ssp. bulgaricus* and 4 of the *ssp. lactis*, were generated by 454 paired-end sequencing (Roche Life Sciences) followed by sequence assembly using Newbler 2.6 (Roche). The resulting scaffolds were ordered using Mauve aligner [23] with the earlier published genome sequence of *L. bulgaricus* ATCC 11842 [6] as the reference. For one strain, *L. delbrueckii ssp. lactis* CNRZ327, a finishing protocol was applied in order to evaluate the level of completeness of the sequence [11].

**Genome annotation.** Genome annotation was performed using AGMIAL [24], preferably transferring annotations from the *L. bulgaricus* ATCC11842 annotation [6] where appropriate. Insertion sequences were attributed to IS families using ISfinder (<https://www-is.biotoul.fr>) [12]. Pseudogenes were detected by aligning predicted CDS against Uniprot release 2013\_06 using blastX. In case of CDS aligning with longer or smaller proteins in the database ( $\pm 10\%$  threshold), the database protein is back-aligned to the genome sequence with tblastn. If the resulting alignment is longer than the original CDS, with the presence of a frameshift or a stop codon in the genome sequence, the original predicted CDS is marked as pseudogene-fragment.

**KEGG pathway annotation.** Proteins from strains ATCC11842 and CNRZ327 were assigned to KEGG (Kyoto Encyclopedia of Genes and Genomes) ortholog groups and mapped to KEGG pathways using the KEGG Automatic Annotation Server (KAAS, <http://www.genome.jp/kegg/kaas/>) [14] with parameters “BBH method” and genes data set “eco, bsu, sau, lmo, lla, spy, spn, ste, lpl, lpj, ljo, ljf, lac, lsa, lsl, ldb, lbu, lbr, lca, lcb, lga, lre, lrf, lhe, lfe, lrh, lrl”. The results were visualized using KEGG or ipath (<http://pathways.embl.de/>) [15].

**Protein clustering.** In order to determine the core proteomes of *L. delbrueckii* and the two ssp. studied, protein sequences encoded by non-pseudogenes in the different genomes were compared using blastP to define groups of orthologous proteins by single linkage clustering (e-value < 10<sup>-3</sup>; > 78 % identity over > 76 % of the longest sequence length).

**Multi Locus Sequence Typing.** MLST analysis was performed with MEGA5 [25] as described in [3], using the conserved parts of 7 housekeeping genes (*fusA*, *gyrB*, *hsp60*, *ileS*, *pyrG*, *recA*, *recG*) and including the strains with known *L. delbrueckii* subspecies attribution studied in [3].

**Phylogenetic reconstruction.** MEGA5 [25] was used to construct strain phylogenetic trees (1000 bootstrap replications). Trees were drawn using jplot [27].

**Carbohydrate fermentation profiling.** Carbohydrate fermentation profiles of the strains in this study were established using API 50 CH (Biomérieux) according to the instructions of the supplier, with readout after 48 hr of incubation at 42 °C.

## Results

***L. delbrueckii* ssp. *bulgaricus* genomes are smaller than ssp. *lactis* genomes.** We sequenced the genomes of 6 *L. delbrueckii* strains to near completion: 4 strains belonging to the ssp. *lactis*, and two strains from the ssp. *bulgaricus* (Table 1). The complete genome sequences of 4 other strains classified as *L. delbrueckii* ssp. *bulgaricus* (ATCC11842 [6], ATCC BAA365 [8], 2038 [9], NDO2

[10]) were retrieved from Genbank. An analysis of the 16S rRNA gene sequences suggested that one of these strains, NDO2, was misclassified and in fact does not belong to the *ssp. bulgaricus* as its 16S sequence does not contain the characteristic two *EcoRI* sites [2]. This conclusion was corroborated by the results of an analysis of inter-strain relationships using MLST (Fig. S1) where strain NDO2 does not make part of the well-defined *ssp. bulgaricus* cluster but instead appears to belong to the *ssp. lactis* or the *ssp. delbrueckii*. An alignment of the 16S rRNA sequence of strain NDO2 with the sequences of the *ssp. lactis* and *ssp. bulgaricus* strains used in this study and the sequence of the *L. delbrueckii ssp. delbrueckii* type strain ATCC9649 [5] revealed 6 positions in which strain NDO2 shared nucleotides conserved in the *ssp. lactis* and *bulgaricus* strains that differ from the *ssp. delbrueckii*, and 3 positions in which strain NDO2 shared nucleotides conserved in the *ssp. lactis* and *delbrueckii* strains that differ from the *ssp. bulgaricus* (Table S1). Taking the results of the different analyses together, we consider strain NDO2 as a representative of the *ssp. lactis*.

Among the genomes that had earlier been sequenced to completion, the three *ssp. bulgaricus* genomes (strains ATCC 11842, ATCC BAA-365 and 2038) are smaller (1,857 to 1,873 kbp) than the *ssp. lactis* NDO2 genome (2,126 kbp) (Table 1). This observation appears to be corroborated by the estimated genome sizes of the newly sequenced strains (estimations taking assembled sequence and gap size estimations from paired end sequencing into account), which tend to be smaller for the *ssp. bulgaricus* than for the *ssp. lactis* (Table 1). An in-depth analysis of one of the newly sequenced genomes, of *L. delbrueckii ssp. lactis* CNRZ327, learned that virtually all remaining gaps in the sequence assembly were due to the presence of one or more repeated elements (IS elements, ribosomal RNA operons) at these sites [11].

**Table 1. Characteristics of *L. delbrueckii* ssp. *bulgaricus* and *L. delbrueckii* ssp. *lactis* genomes**

	<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> strains					<i>L. delbrueckii</i> ssp. <i>lactis</i> strains				
	ATCC 11842	ATCC BAA365	2038	VIB27	VIB44	NDO2	CNRZ226	CNRZ327 (e)	CNRZ333	CNRZ700
Assembled genome size (a)	1864998	1856951	1872918	1838091	1810332	2125753	1904440	1844879 (1938538)	1996651	1989632
Estimated genome size (b)	N/A	N/A	N/A	1853236	1818010	N/A	1910730	1969493 (2105434)	2051578	2085621
Number of contigs	1	1	1	32	27	1	21	161 (571)*	87	333
Number of scaffolds	1	1	1	14	14	1	10	33 (1)	23	75
Average sequencing depth	-	-	-	86	94	-	71	78	77	56
Number of CDS (c)	1466	1380	1333	1783	1711	1666	1665	1525	1721	1593
Number of pseudogene- fragments (d)	630	341	459	388	423	346	390	545	381	408
Number of CDS with unknown function	642	294	343	442	434	317	361	315	369	345
Overall GC content (%)	49.7	49.7	49.68	49.41	49.72	49.56	49.83	49.81	48.18	49.48
GC content of CDS	50.8	51.2	51.9	51.7	51.8	51.5	52.02	52.2	51.6	51.8
GC content of CDS at codon position 3 (%)	65.0	64.8	64.9	66.0	66.7	64.0	67.0	65.1	63.4	67.4
CDS as % of genome sequence	73.4	68.3	69.2	77.1	76.5	75	77.4	62.9	75.3	68.4
Number of rrn operons	9	9	9	-	-	9	-	9	-	-
	<b>Protein localization prediction</b>									
Cytoplasmic	1089	996	958	1346	1277	1245	1237	1140	1272	1182
Membrane	225	227	208	247	253	242	248	223	259	237
Surface exposed	86	101	115	118	115	119	119	101	123	113
Secreted	69	56	52	72	66	60	61	61	67	61

a, without paired end sequencing results.

b, assembled sequence plus estimated size of sequence gaps (estimations on the basis of paired end sequencing results).

c, not counting pseudogenes.

d, corresponding to CDS annotated as “fragment”.

e, numbers in parentheses represent data after genome finishing.

\*The increase in the number of contigs after genome finishing is due to the addition of sequence fragments in the original sequence gaps.

Assuming that the same explanation holds for the other genomes which were sequenced to comparable depths (Table 1), we conclude that except for repeated elements the genome sequences are practically complete and allow a comparison of the predicted proteomes (see below).

***L. delbrueckii* ssp. *lactis* genomes contain more IS elements.** The finishing and detailed analysis of the *L. delbrueckii* ssp. *lactis* CNRZ327 genome revealed the presence of an extremely high number of IS elements (Table 2): 215 IS elements were detected of which 178 could be attributed to known families using IS Finder (<http://www-is.biotoul.fr>) [12].

**Table 2. IS families in *L. delbrueckii* ssp. *bulgaricus* and *L. delbrueckii* ssp. *lactis***

IS family	<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> ATCC11842	<i>L. delbrueckii</i> ssp. <i>lactis</i> CNRZ327
IS30	16	57
IS 256	0	45
ISL3	2	33
IS110	8	17
IS3	1	16
IS4	12	9
ISL30	0	1
other	17	37
<b>Total</b>	<b>56</b>	<b>215</b>

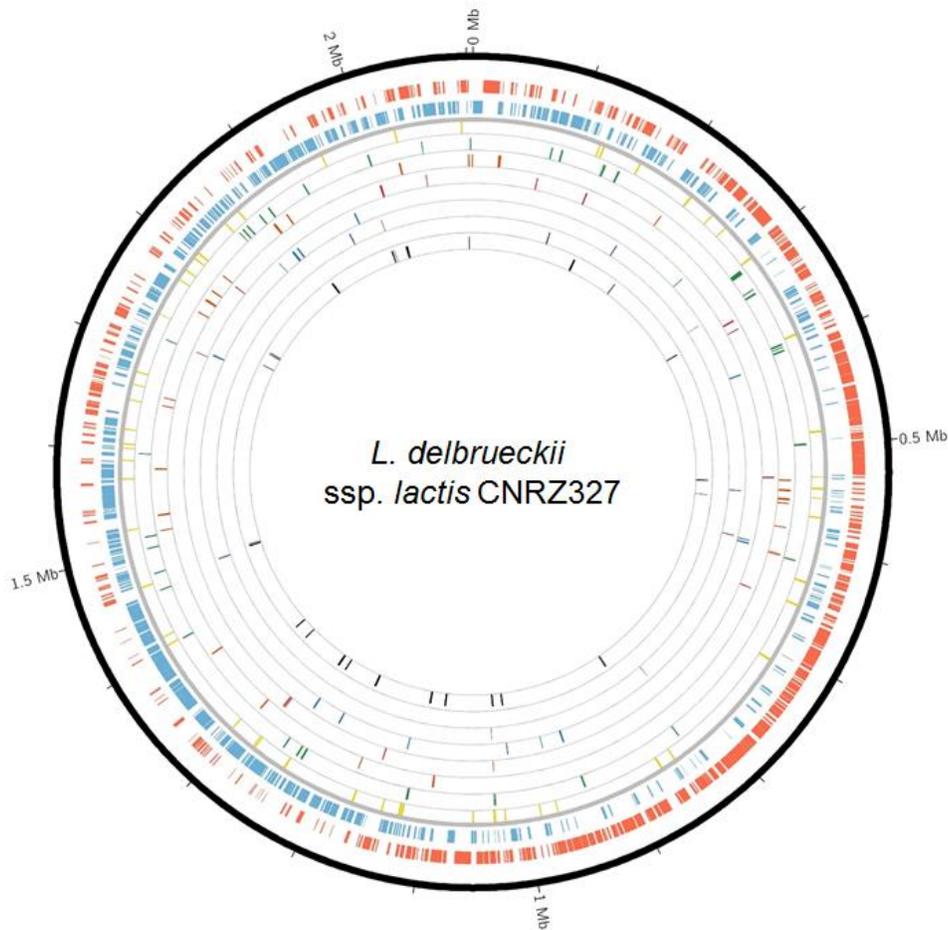
Numbers indicate the frequency of occurrence of different IS elements in *L. delbrueckii* ssp. *bulgaricus* ATCC11842 [6] and *L. delbrueckii* ssp. *lactis* CNRZ327 (this study) genomes.

The earlier published complete *L. delbrueckii* ssp. *lactis* NDO2 genome [10] also appears to contain far more (133) IS elements than the complete genomes of *L. delbrueckii* ssp. *bulgaricus* strains

ATCC 11842, ATCC BAA365, and 2038 (56, 29, and 54, respectively [6,8,9] (Table S2)). If the number of remaining gaps in the newly sequenced genomes of the present study is an indication for the number of IS elements, as it proved to be in the case of strain CNRZ327, the same tendency is found in these genomes. Except for *L. delbrueckii* ssp. *lactis* CNRZ226, the ssp. *lactis* genomes contain significantly more IS elements than the ssp. *bulgaricus* genomes, even if the number of IS elements can also vary considerably between strains of the same subspecies. Differences in numbers of IS elements thus appear to contribute significantly to the differences in genome sizes between ssp. *bulgaricus* and ssp. *lactis* strains.

When comparing *L. delbrueckii* ssp. *lactis* CNRZ327 and *L. delbrueckii* ssp. *bulgaricus* ATCC11842, striking differences are observed in the numbers of IS elements of the IS256, IS30 and ISL3 families, which are all largely overrepresented in strain CNRZ327 (Table 2). The ubiquitous IS256 family (45 copies in strain CNRZ327) is even completely absent from *L. delbrueckii* ssp. *bulgaricus* ATCC11842 and 2038 (not shown), but it can be found in the *L. delbrueckii* ssp. *bulgaricus* ATCC BAA365 genome (2 copies; not shown). In *L. delbrueckii* ssp. *lactis* NDO2, 7 copies are present (not shown). These observations illustrate that not only the global numbers of IS but also the numbers of IS elements of a given family largely vary between strains, including within the same ssp.. No ssp. *lactis* or ssp. *bulgaricus* specific IS elements were found.

In contrast with *L. bulgaricus* ATCC11842 where a region of 415 kbp (22% of the genome) is exempt of IS elements [6], in *L. lactis* CNRZ327 the IS elements seem more or less randomly distributed over the genome although a region with relatively few IS elements can be distinguished between positions 620 kb and 920 kb (Fig 1).



**Fig. 1. Genome atlas of *L. delbrueckii* ssp. *lactis* CNRZ327.** The nine circles (outer to inner) show: Circle 1, CDS (excluding pseudogenes and transposases) on positive (red) or negative (blue) strand; Circle 2, transposases of the IS30 family (yellow); Circle 3, transposases of the IS256 family (green); Circle 4, transposases of the ISL3 family (orange); Circle 5, transposases of the IS110 family (red); Circle 6, transposases of the IS3 family (blue); Circle 7, transposases of the IS4 family (purple); Circle 8, transposases of ISL30 family (black) Circle 9, transposases of unknown family (grey).

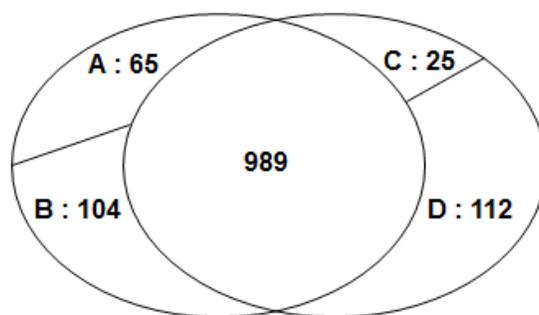
Several locations on the genome, situated in between scaffolds of the original sequence assembly, appear to be insertion hotspots where varying combinations of two or three different IS elements are found in close proximity or in a nested configuration (not shown). 4 cases of nested IS were observed: IS256 or IS110 inserted in ISs of unknown family, IS 256 in ISL3, and IS110 in IS30.

IS elements in CNRZ327 are nearly exclusively found in intergenic regions (not shown), and thus do not appear to have played an important role in the shaping of the functional gene repertoire of the ssp..

**Gene repertoires and metabolic capacities.** Both the spp. *bulgaricus* and the spp. *lactis* show distinct signs of genomes in an active state of evolution. An aberrant GC content at the third codon positions in coding sequences, a sign of rapid ongoing evolution first observed in the *L. delbrueckii* spp. *bulgaricus* ATCC 11842 genome [6], is also observed in the other spp. *bulgaricus* strains, as well as in the spp. *lactis* strains (Table 1). High numbers of pseudogenes, another sign of rapid evolution, are also observed in both spp. (Table 1). In *L. delbrueckii* spp. *bulgaricus* strain ATCC11842 the occurrence of pseudogenes appeared to reflect an adaptation to the protein rich milk environment through the elimination of superfluous genes, notably involved in amino acid biosynthesis and carbohydrate metabolism [6]. In *L. delbrueckii* spp. *lactis* CNRZ327 as well as in the other *L. delbrueckii* strains studied here, we observed a same tendency of elimination of genes involved in amino acid biosynthesis and carbohydrate metabolism (see below). In both spp., the spontaneous deterioration of genes thus appears to have played an important role in the shaping of the functional gene repertoire of the species, through the inactivation of superfluous genes in an apparent adaptation to the environment.

In order to compare the remaining gene repertoires and metabolic capacities of the spp. *bulgaricus* and the spp. *lactis* we used two *in silico* methods and experimentally evaluated the potential of using different mono- and di-saccharides. First, we performed a clustering of orthologous proteins in order to determine the extent of the spp. *bulgaricus* core proteome, the spp. *lactis* core proteome, and the common core proteome of the two subspecies. While one might have expected that the sizes of the respective core proteomes would reflect the earlier mentioned difference in genome size between the two spp., this appears not to be the case: the results presented in Fig. 2 show that the two spp. have similarly sized core proteomes, which largely overlap.

Core proteome *ssp. lactis*    Core proteome *ssp. bulgaricus*

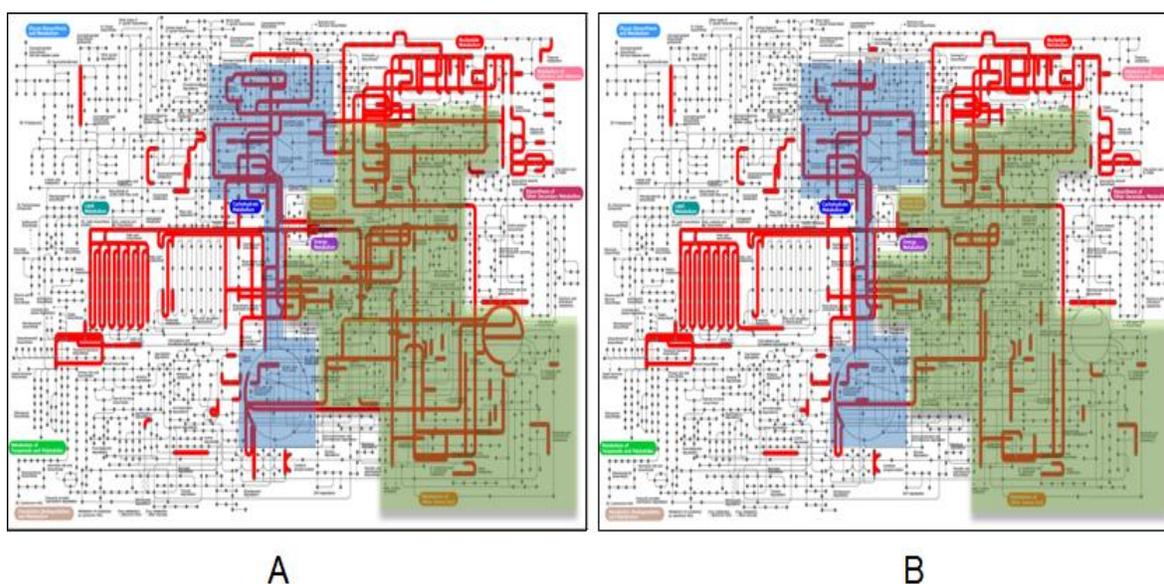


**Fig. 2. Core proteomes of *L. delbrueckii* ssp. *lactis* and ssp. *bulgaricus*.** Ovals represent the core proteomes of *L. delbrueckii* ssp. *lactis* and *L. delbrueckii* ssp. *bulgaricus*, respectively. The overall core of the 5 ssp. *lactis* and the 5 ssp. *bulgaricus* strains in this study consists of 989 proteins. A, 65 proteins are present in all ssp. *lactis* strains and absent from all 5 ssp. *bulgaricus* strains; B, 104 proteins are present in all ssp. *lactis* strains and absent from 1 to 4 ssp. *bulgaricus* strains; C, 25 proteins are present in all ssp. *bulgaricus* strains and absent from all 5 ssp. *lactis* strains; D, 112 proteins are present in all ssp. *bulgaricus* strains and absent from 1 to 4 ssp. *lactis* strains.

When comparing the proteins from the core proteome of the ssp. *bulgaricus* that are lacking from all 5 ssp. *lactis* strains at the one hand, and the proteins from the core proteome of the ssp. *lactis* that are lacking from all 5 ssp. *bulgaricus* strains at the other (Fig 2; supplementary tables S3 and S4), it is striking that the majority (17 out of 25) of the ssp. *bulgaricus*-specific proteins have no known function while the ssp. *lactis*-specific proteins, which are more than twice as many, nearly all (49 out of 65) have a known (general or detailed) function. The latter proteins are mainly involved in carbohydrate and amino-acid metabolism (see below). The 8 ssp. *bulgaricus* specific proteins to which a function could be attributed are mainly involved in membrane transport. For 24 among the 65 ssp. *lactis* specific proteins, fragments of coding sequences (pseudogenes) can be found in one or more of the ssp. *bulgaricus* strains. Likewise, for 5 of the 25 ssp. *bulgaricus* specific proteins gene fragments can be found in one or more of the ssp. *lactis* strains. These observations suggest that at least part of the subspecies specificity results from the differential loss of ancestral genes.

A second *in silico* comparison was made using KEGG [13] functional annotations as a starting point. For this purpose, proteins from the predicted proteomes of strains ssp. *bulgaricus*

ATCC11842 and *ssp. lactis* CNRZ327 were assigned to KEGG ortholog groups and mapped to KEGG pathways using the KEGG Automatic Annotation Server (KAAS) [14]. Matching maps were compared, and for a selection of differences between the two strains the comparison was extended to the 10 strains used in this study using homology search based analyses. Not unexpectedly, the results of this analysis to a great extent corroborate the results of the core proteomes analysis for as far as proteins with known functions are concerned, and in addition put a certain number of differences in a metabolic pathway context. A first impression of the global metabolic capacities of the two strains obtained using Ipath [15] suggests that these capacities are more extensive for the *ssp. lactis* than for the *ssp. bulgaricus* (Fig. 3), notably regarding carbohydrate and amino acid metabolism which will be discussed in more detail in the following sections together with some other features of relevance for dairy industrial bacteria.



**Fig. 3. Metabolic capacities of *L. delbrueckii ssp. lactis* CNRZ327 and *L. delbrueckii ssp. bulgaricus* ATCC 11842.** Metabolic pathway analysis was performed using KEGG [13, 14]. Graphs were generated using ipath [15]. A, *L. delbrueckii ssp. lactis* CNRZ327; B, *L. delbrueckii ssp. bulgaricus* ATCC 11842; red, enzyme functions identified in the respective genomes; highlighted in blue, carbohydrate metabolism pathways; highlighted in green, amino acid biosynthesis pathways.

**Carbohydrate metabolism potential is much more reduced in the *ssp. bulgaricus* than in the *ssp. lactis*.** An overview of the experimentally determined carbohydrate metabolism potential of

four *ssp. lactis* and four *ssp. bulgaricus* strains (Table 3) confirms that the former are capable of metabolizing a more extensive range of mono- and di-saccharides, mostly of plant origin, as has been reported earlier. However, a high level of variability was observed within the *ssp. lactis*, both with regard to the number (from 6 to 12) and the nature of the carbohydrates that could be fermented by a given strain. The *ssp. bulgaricus* strains fermented from 2 to 4 different carbohydrates. Because of this intra *ssp.* variability, and in contrast with the earlier description of the *ssp.* [1], only one of the 14 substrates for which fermentation was observed allowed to distinguish the two *ssp.* in our set of eight strains: N-acetylglucosamine was fermented by the *ssp. lactis* but not by the *ssp. bulgaricus*.

**Table 3. Carbohydrate fermentation profiles of *L. delbrueckii* strains.**

	Galactose	D-glucose	D-fructose	D-mannose	N-acetylglucosamin	Amygdalin	Arbutin	Esculin	Salicin	Cellobiose	Maltose	Lactose	Sucrose	Trehalose	Total
<i>Lactobacillus delbrueckii</i> <i>ssp. lactis</i> CNRZ226	-	+	+	+	+	+	+	+	+	+	+	-	+	+	12
<i>Lactobacillus delbrueckii</i> <i>ssp. lactis</i> CNRZ327	-	+	+	+	+	-	-	+	-	-	-	+	-	-	6
<i>Lactobacillus delbrueckii</i> <i>ssp. lactis</i> CNRZ333	+	+	+	+	+	-	+	+	+	-	+	+	+	+	12
<i>Lactobacillus delbrueckii</i> <i>ssp. lactis</i> CNRZ700	+	+	-	+	+	-	-	+	-	-	-	+	-	+	7
<i>Lactobacillus delbrueckii</i> <i>ssp. bulgaricus</i> ATCC11842	-	+	-	+	-	-	-	-	-	-	-	+	-	-	3
<i>Lactobacillus delbrueckii</i> <i>ssp. bulgaricus</i> ATCC BAA-365	-	-	-	+	-	-	-	-	-	-	-	+	-	-	2
<i>Lactobacillus delbrueckii</i> <i>ssp. bulgaricus</i> VIB27	-	+	-	+	-	-	-	+	-	-	-	+	-	-	4
<i>Lactobacillus delbrueckii</i> <i>ssp. bulgaricus</i> VIB44	-	-	-	+	-	-	-	-	-	-	-	+	-	-	2

Carbohydrate fermentation profiles were experimentally determined using API 50 CH (Biomérieux). +, fermentation observed; -, no fermentation observed. Results were reproduced in two independent experiments.

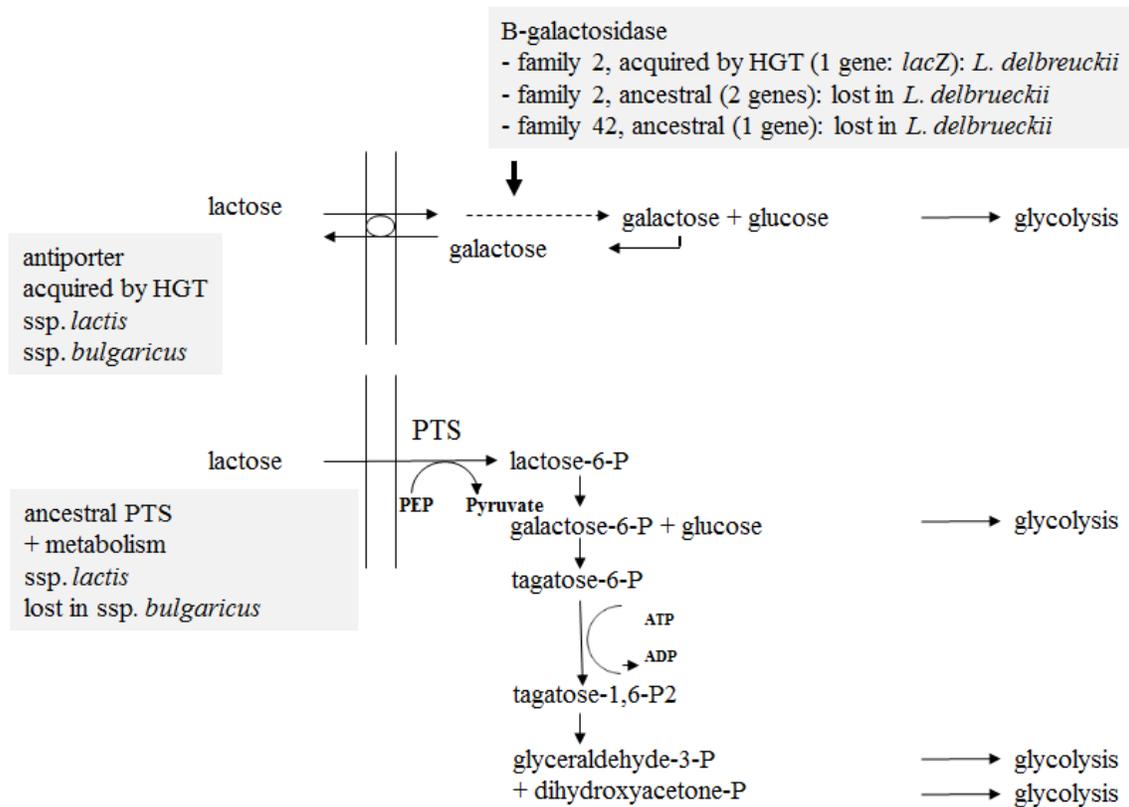
With only few exceptions, the observed fermentation profiles could be predicted on the basis of *in silico* metabolic pathway analyses using existing and newly established genome sequences (Table S5). The results of these *in silico* analyses also suggested that the *ssp. lactis* would be able to degrade starch, in contrast to the *ssp. bulgaricus* (Table S5). This prediction was experimentally

confirmed by growth on M17 medium containing starch as the only carbon source and a starch degradation assay (results not shown).

Apart from the difference in carbohydrate metabolism potential between the two ssp., the *in silico* analyses also revealed that when a carbohydrate metabolic pathway is inactive in the ssp. *lactis* this is mostly due to the lack or fragmentation of one or two pathway specific genes while in several cases in the ssp. *bulgaricus* the same pathway is completely lacking or only represented by pseudogenes. This observation suggests that the ssp. *bulgaricus* represents a more advanced state of elimination of ancestral carbohydrate metabolic pathways than the ssp. *lactis*.

The results of the *in silico* analysis also indicate that for the uptake of carbohydrates *L. delbrueckii* mostly relies on active transport using PTS systems or ABC transporters (Table S5). A noticeable exception is the uptake of the milk sugar lactose in the ssp. *bulgaricus*, which will be described below.

**Lactose metabolism in the ssp. *bulgaricus* relies on horizontally acquired genes.** The comparative genomics approach using KEGG pathways led to some particularly interesting observations with regard to the uptake and metabolism of the milk sugar lactose, a key feature of dairy lactic acid bacteria. While for the ssp. *bulgaricus* it is well established that lactose enters the cell via an antiporter which, after cleavage of lactose by a  $\beta$ -galactosidase, extrudes the non-metabolizable galactose moiety of the molecule [4, 16, 17], it appears that nearly all the ssp. *lactis* strains in addition possess a lactose PTS system to import and phosphorylate lactose, a phospho- $\beta$ -galactosidase to cleave the lactose-6-P formed, and the enzymes needed to metabolize the galactose-6-phosphate liberated in the latter reaction (Table S5). Using this system, the ssp. *lactis* would thus be able to metabolize not only the glucose moiety of lactose, like the ssp. *bulgaricus*, but also the galactose moiety (Fig. 4).



**Fig. 4. Lactose transport and metabolism pathways in *L. delbrueckii* and its ancestors.**

In *L. delbrueckii* *ssp. bulgaricus*, lactose uptake relies on a lactose-galactose antiporter which has been acquired by horizontal gene transfer while the ancestral lactose PTS system has been lost. *L. delbrueckii* *ssp. lactis* contains both transport systems. In *L. delbrueckii* *ssp. bulgaricus*, lactose metabolism relies on a  $\beta$ -galactosidase which has been acquired by horizontal gene transfer while the ancestral  $\beta$ -galactosidases have been lost. *L. delbrueckii* *ssp. lactis* contains the same horizontally acquired  $\beta$ -galactosidase and, in addition, the pathways to completely metabolize lactose-6-P generated by the lactose PTS-system.

This PTS system, the phospho- $\beta$ -galactosidase and the galactose-6-P metabolizing enzymes appear to be rare in the lactobacilli of the acidophilus group and in lactobacilli in general. The phylogeny of the lactose specific PTS system component (LacE) and the phospho- $\beta$ -galactosidase (LacG) appears to be coherent with the 16S rRNA phylogeny (Fig. S2), however, suggesting that they are encoded by ancestral genes which have been lost in most lactobacilli, including *L. delbrueckii* *ssp. bulgaricus*. The first signs of a loss of this system are also found in the *ssp. lactis*, where in strain CNRZ226 the whole system is missing, like in the *ssp. bulgaricus* (Table S5).

In contrast, several lines of evidence suggest that the lactose-galactose antiporter permease and the functionally associated  $\beta$ -galactosidase have most probably been acquired by horizontal gene transfer in the ancestral *L. delbrueckii* lineage: 1) while some of the closely related bacteria in the *Lactobacillus acidophilus* group contain a  $\beta$ -galactosidase (belonging to the glycoside hydrolase family 2) encoded in two genes (*LacL* and *LacM*, for large and small subunit, respectively), the homologous family 2  $\beta$ -galactosidase of *L. delbrueckii* is of the type that resembles that found in *Streptococcus*, encoded in one gene (*lacZ*) (Fig. S3) and 2) the lactose-galactose antiporter (lactose permease, encoded by *lacS*) of *L. delbrueckii* shows a higher resemblance to its *Streptococcus* homologues than to its homologues in closely related lactobacilli, causing a disagreement between the *lacS*-based phylogeny and 16S rRNA-based phylogeny indicative of horizontal transfer between the *L. delbrueckii* lineage and the *Streptococcus* lineage (Fig. S3). It is interesting to note that several lactobacilli of the acidophilus group (*L. johnsonii*, *L. acidophilus*, *L. crispatus*, *L. helveticus*) and some other lactobacilli in addition to the two-gene family 2  $\beta$ -galactosidase contain LacA, a  $\beta$ -galactosidase of the glycoside hydrolase family 42, which also appears to be ancestral (Fig. S4), showing the same phylogeny as the 16S rRNA genes. This family 42 protein is not present in *L. delbrueckii*.

It thus appears that, in spite of the presence of ancestral genes encoding a lactose permease,  $\beta$ -galactosidases belonging to two different glycoside hydrolase families, a lactose PTS, and a phospho- $\beta$ -galactosidase in the *L. acidophilus* group, the lactose fermenting capacity of one of the most important dairy bacteria in this group, *L. delbrueckii* ssp. *bulgaricus*, relies on horizontally acquired rather than deep ancestral genes. As earlier reported [5], the *lac* repressor is inactivated in the ssp. *bulgaricus* strains. Inactivation of this repressor and the loss of the galactose metabolic pathway in the spp. *bulgaricus* have been presented as possibly having contributed to the selection of this ssp. for fast fermentation of milk where lactose is abundant [5, 18]. Furthermore, the

ancestral PTS-lactose system still visible in the *ssp. lactis* suggests that the ancestor has its origin in a low lactose environment, where the PTS system would ensure a level of efficiency of lactose transport and energy harvesting that the antiporter system cannot provide.

**Amino acid biosynthesis capacities are more severely reduced in the *ssp. bulgaricus* than in the *ssp. lactis*.** A second key to the comprehension of *L. delbrueckii*'s metabolic capacities is the protein richness of the milk environment. Earlier studies of the *L. bulgaricus* ATCC11842 genome revealed a highly reduced capacity of amino acid biosynthesis which may be explained as an adaptation to this environment where the bacterium's proteolytic system can liberate amino acids from the abundant milk proteins, thus rendering amino acid biosynthesis pathways superfluous [6]. This view is confirmed in the other *L. bulgaricus* strains in this study (Table S6). In the *ssp. lactis*, the same tendency can be observed albeit in a less advanced state: fewer amino acid biosynthesis pathways are inactivated (Table S6), in inactivated pathways less genes are inactivated (Table S6), and inactivated genes are less fragmented than in the *ssp. bulgaricus* (not shown). All strains of the *ssp. lactis* are predicted to synthesize lysine, aspartate and arginine, while most of the *ssp. bulgaricus* strains cannot synthesize these amino-acid (Table S6). Neither of the two *ssp.* would be able to produce phenylalanine, tyrosine, tryptophan, glycine, serine, alanine, glutamate, valine, leucine, isoleucine or histidine.

Both *ssp.* contain an important proteolytic system with about 30 proteins annotated as proteinases or peptidases (not shown). This number is not different, however, from the number found in *Lactobacillus gasseri*, a closely related *Lactobacillus acidophilus* group member from the gastrointestinal tract (not shown). The only significant adaptation of the proteolytic system of both *L. delbrueckii ssp. bulgaricus* and *ssp. lactis* appears to be the presence of the major cell wall bound protease PrtB, which is responsible for milk casein degradation [19]. This protease is not found in

closely related lactobacilli. The most similar protein sequence is found in *Lactobacillus ruminis* (50% amino acid identity; results not shown).

**Less genes involved in acid tolerance in the ssp. *bulgaricus* than in the ssp. *lactis*.** *L. delbrueckii* acidifies its growth medium through the production of lactate, and acid tolerance therefore is an important aspect of its physiology. In spite of this, the analysis of the ssp. *bulgaricus* ATCC 11842 genome revealed the presence of only few of the diverse systems that have been implicated in acid tolerance in lactic acid bacteria [20], and suggested that an H<sup>+</sup> transporting ATPase constitutes its primary defense, with a possible contribution of cation:proton antiporters and two ornithine decarboxylases [6]. For the latter enzymes, *L. bulgaricus* appears to rely on ornithine produced by *Streptococcus thermophilus*, its companion bacterium in yogurt fermentation [6]. This vision is corroborated by the analysis of the other ssp. *bulgaricus* strains in the present study (Table S7). The ssp. *lactis* genomes in contrast reveal a more diverse set of genes involved in acid stress resistance where, in addition to the functions present in the ssp. *bulgaricus*, ornithine as well as pH lowering ammonia and CO<sub>2</sub> can be produced through the arginine deiminase pathway (encoded by the *arcA*, *arcB* and *arcC* genes). Ornithine can then be used to import more arginine through an arginine-ornithine antiporter (encoded by *arcD*), or be decarboxylated by ornithine decarboxylase for which the ssp. *lactis* strains possess one gene copy. In one of the ssp. *lactis* strains, CNRZ700, *arcA* appears to be a pseudogene, while the same is true for the *arcD* gene in strains CNRZ327 and NDO2 (Table S7). In the ssp. *bulgaricus* strains, *arcD* is present as a pseudogene, while no traces are found of the *arcA*, B, or C genes.

The ssp. *lactis* genomes also contain remnants of another gene that has been implicated in acid tolerance, encoding a glutamate decarboxylase, of which no trace can be found in the ssp. *bulgaricus* strains. Both the arginine deiminase pathway and glutamate decarboxylase can be found

in other lactobacilli of the acidophilus group, and the above mentioned observations suggest that *L. delbrueckii* ssp. *bulgaricus* has lost these functions while the ssp. *lactis* is following the same way.

## **Discussion**

*Lactobacillus delbrueckii* is one of the economically most important bacteria in dairy industries, where the ssp. *bulgaricus* is mainly known from its application in yogurt fermentation and the ssp. *lactis* from its use in the production of Parmesan and Emmenthal type cheeses. The classical procedure of distinguishing the two ssp. on the basis of their phenotypical features, in particular their carbohydrate fermentation profiles, has clearly shown its limits, as again confirmed in this study. While the ssp. *lactis* strains in this study were on an average capable of fermenting a larger pallet of (up to 12) mono- and disaccharides than the ssp. *bulgaricus* strains, a high level of variability was observed between strains of the same ssp. and only one monosaccharide substrate was identified that allowed to reliably discriminate the two ssp., even among the small number of strains studied.

Our *in silico* analyses revealed that the ssp. *lactis* would, in contrast to the ssp. *bulgaricus*, be able to degrade the complex plant carbohydrate starch. This feature, which has not been reported before, was experimentally confirmed. In our test panel of 8 strains, starch thus constituted the second carbohydrate that allowed to discriminate the two *L. delbrueckii* ssp., and it will be of interest to see if this easily testable phenotypical criterion holds when tested on a larger collection of strains.

The *in silico* analysis of the first complete genome sequence of an *L. bulgaricus* strain [6] indicated that this strain derived from an ancestor that would have been capable of fermenting a number of carbohydrates of plant origin, and suggested that *L. bulgaricus* adapted to the milk environment by reductive evolution in which superfluous metabolic pathways went through a process of gene inactivation and elimination. Here, we extend this observation to four other *L.*

*bulgaricus* genomes and demonstrate that the differential loss of genes involved in carbohydrate metabolism in different strains is at the basis of the observed variability in fermentation capacities. For the ssp. *bulgaricus* strain ATCC BAA365, adaptation to the milk environment has gone so far that it only metabolizes one sugar, mannose, other than the milk sugar lactose. In this study we extend these genome-scale observations for the first time to the ssp. *lactis*, where in four newly established and one existing genome sequence a similar, but less advanced, adaptation to the milk medium can be recognized. Intriguingly, while for any of 14 mono- and di-saccharides studied at least one and on an average 4 out of 8 strains lost the capacity to ferment that carbohydrate, all *L. bulgaricus* and *L. lactis* strains studied preserved their capacity to ferment the plant sugar mannose.

A second major adaptation to the milk environment concerns the loss of amino acid biosynthesis pathways, earlier observed in *L. bulgaricus* [6] and here also in the ssp. *lactis*. Like in the case of carbohydrate metabolism, this adaptation appears to be less advanced in the ssp. *lactis* than in the ssp. *bulgaricus*. Explained by the ample availability of amino acids in the form of milk proteins, this adaptation is coherent with the presence of the major cell wall bound protease PrtB responsible for the first step in the degradation of milk proteins in both ssp.. The gene encoding this proteinase is not found in closely related lactobacilli.

From these examples, the general picture emerges that both ssp. are in the course of adaptation to the protein and lactose rich milk environment, where the ssp. *bulgaricus* has advanced further in this direction. This conclusion is in line with an earlier study by Germond et al. [5] who postulated on the basis of 16S rRNA analysis that the ssp. *bulgaricus* had evolved further away from the common ancestor than the ssp. *lactis*. The common ancestor would have evolved in an environment where various carbohydrates from plant origin constituted the main carbon sources, as can be deduced from the presence of (remnants of) genes involved in their fermentation present in the various strains analyzed.

Interestingly, our analyses reveal that while the ssp. *bulgaricus* and *lactis* possess an acquired lactose-galactose antiporter to import the milk sugar lactose, the common ancestor possessed a dedicated PTS system, which is still present in the ssp. *lactis*, for this purpose. Where an antiporter is the system of choice in the lactose-rich milk environment, a PTS system which excels in conditions where the substrate concentration is low points to an ancestral environment where lactose was present, but in low concentrations. Together with the indication that the ancestral environment contained carbohydrates from plant origin, this may suggest that the ancestor evolved in the mammalian digestive tract, an environment where both conditions can be met. This hypothesis is coherent with the fact that most of the known closely related lactobacilli of the acidophilus group are gut isolates, and further backed by the presence of genes coding for putative mucus binding proteins in 4 of the 5 ssp. *lactis* genomes analyzed (not shown).

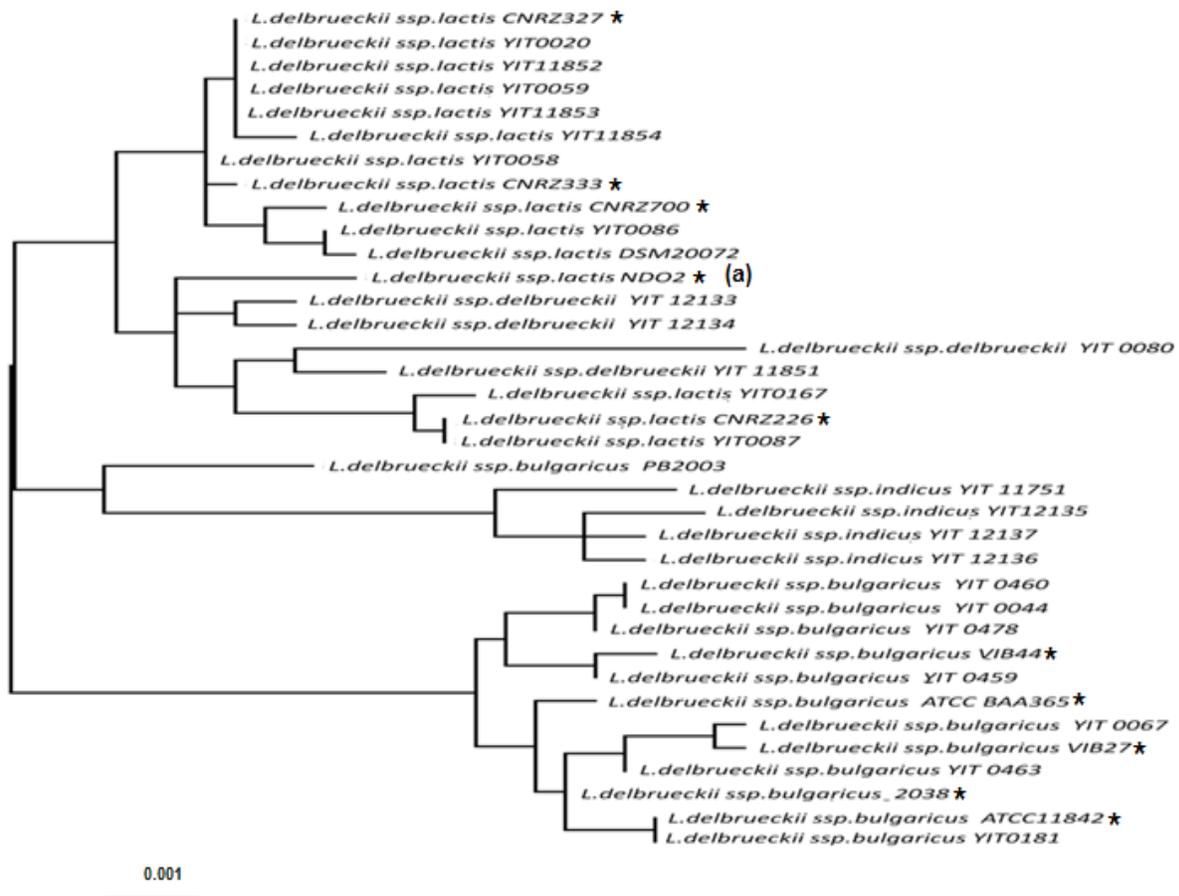
With genomes that similar which represent different stages of adaptation to the milk environment, what is the difference between the ssp. *lactis* and *bulgaricus* that is important to dairy industries? Earlier reports mentioned the capacity of faster and more reliable milk fermentation by the ssp. *bulgaricus* than by the ssp. *lactis*, which was explained by the inactivation of the *lac* repressor in the former resulting in the constitutive expression of  $\beta$ -galactosidase, the enzyme that cleaves lactose to yield glucose and galactose [21]. Likewise, the protease prtB has been reported to be constitutively expressed in *L. bulgaricus*, while being tightly regulated in the ssp. *lactis* [22]. These examples indicate that at least part of the industrially relevant difference between the two ssp. may be found in gene regulation rather than in gene content. High intra ssp. variability makes it difficult to pinpoint further industrially relevant features, and in view of the evolutionary picture emerging from the present study chances are high that desired properties can be found among either ssp. From a dairy industrial point of view, the distinction of the two ssp. may therefore look artificial.

## Conclusions

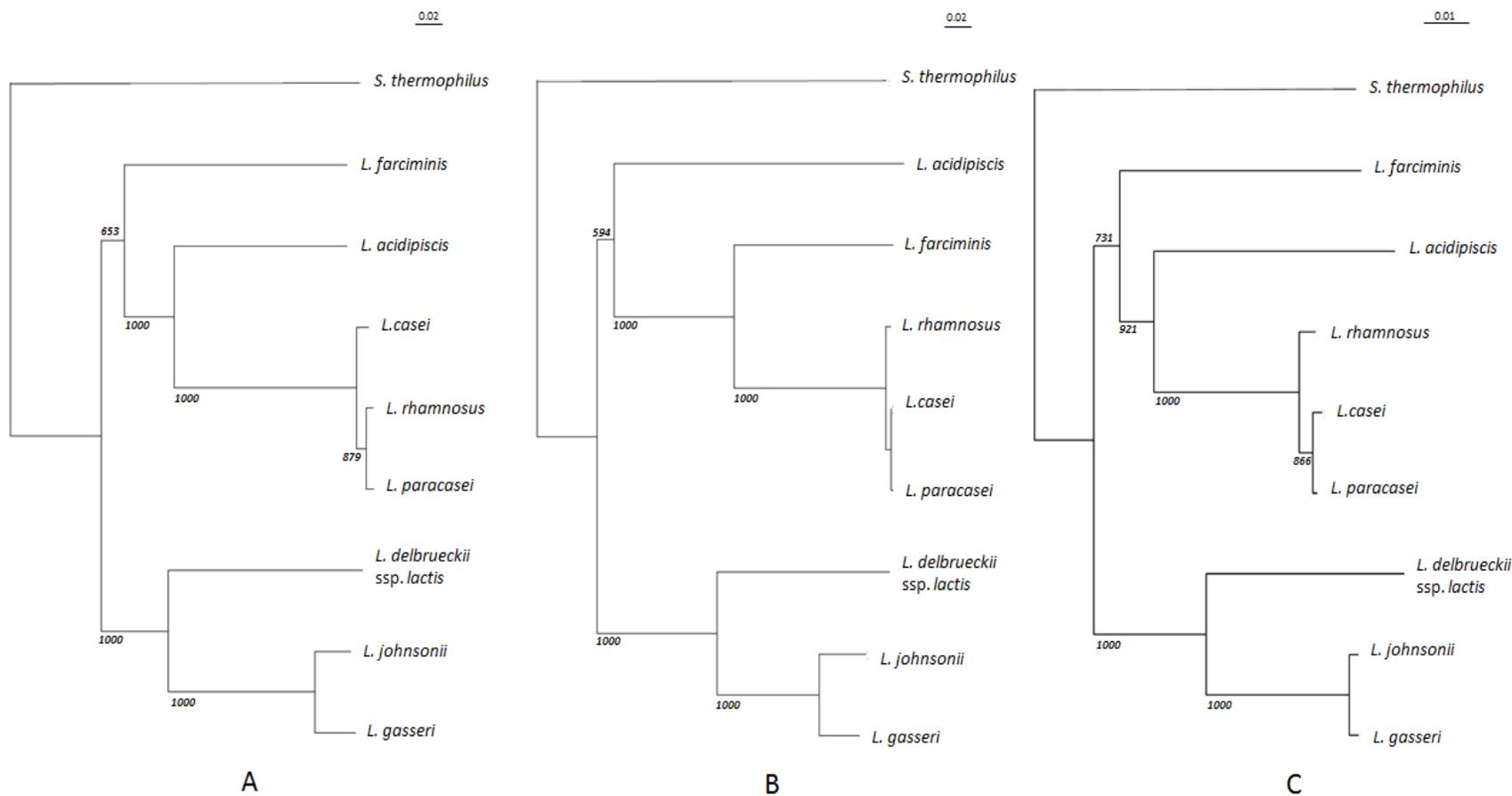
The dairy lactic acid producing bacteria *L. delbrueckii* ssp. *lactis* and ssp. *bulgaricus* appear to represent different points on the same evolutionary track of adaptation to the milk environment through the loss of superfluous functions, where the ssp. *bulgaricus* has progressed further away from the common ancestor. Interestingly, it appears that one of the most important traits, lactose fermentation, of one of the economically most important dairy bacteria, *L. delbrueckii* ssp. *bulgaricus*, relies on horizontally acquired rather than deep ancestral genes. In this sense this bacterium may thus be regarded as a natural GMO *avant la lettre*.

## Supplementary information

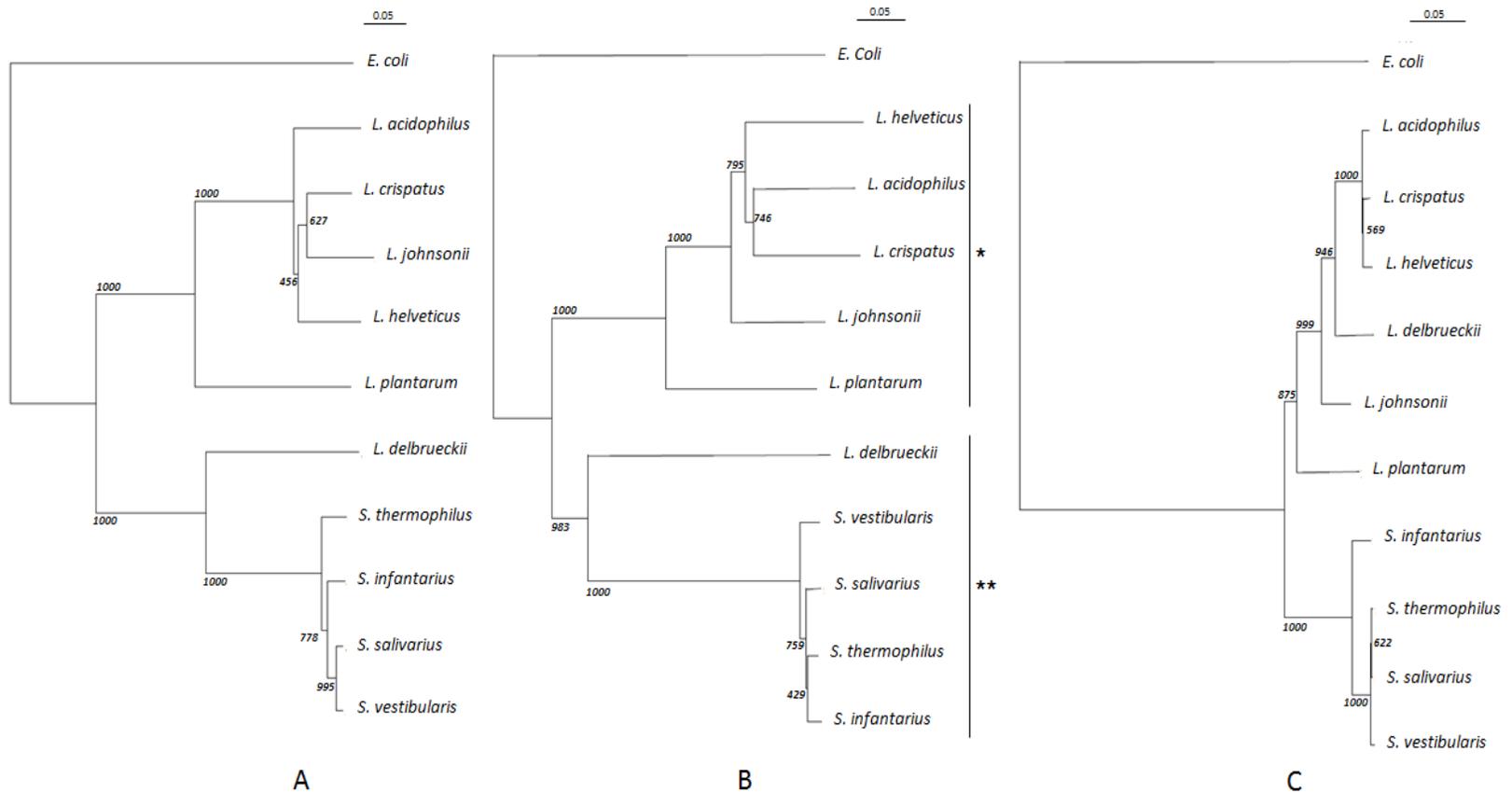
### Supplementary figures



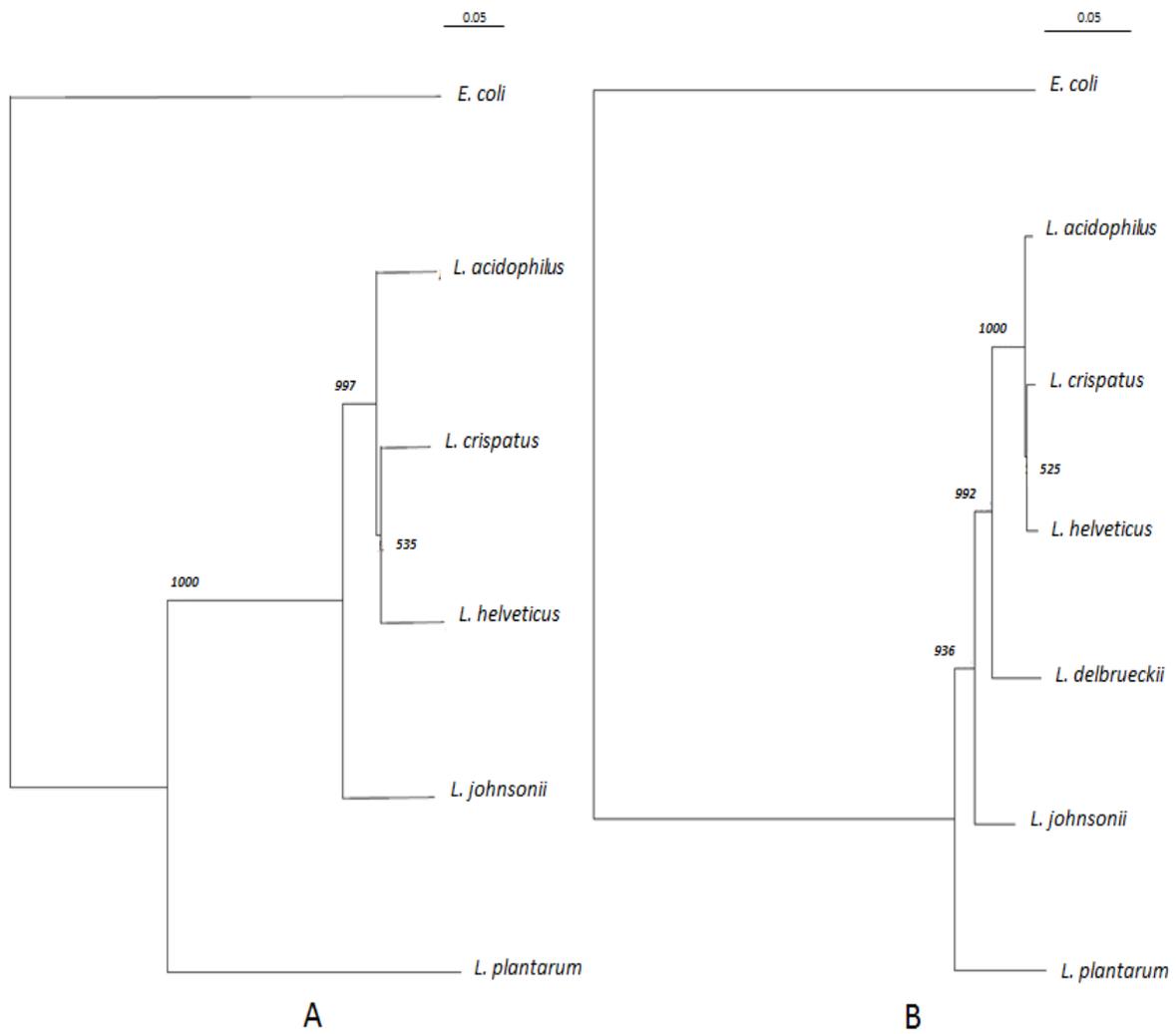
**Fig. S1. Phylogenetic analysis of *L. delbrueckii* strains using MLST.** The phylogenetic tree was constructed using MEGA 5 software [25][244]. \*, strains used in the present study; (a) originally classified as *L. delbrueckii* ssp. *bulgaricus* in [10][245]. The scale bar represents the mean number of nucleotide substitutions per site.



**Fig. S2. Coherence between 16S rRNA-based phylogeny and *lacE* and *lacG* based phylogenies.** Alignment of nucleotide sequences and tree construction were performed using ClustalW [26],[246] and trees were drawn using njplot [27]. **A**, lactose specific PTS system component (*lacE* gene) phylogeny; **B**, phospho- $\beta$ -galactosidase (*lacG* gene) phylogeny; **C**, 16S rRNA phylogeny. Numbers indicate bootstrap values; The scale bar represents the mean number of nucleotide substitutions per site.



**Fig. S3. Inconsistency between 16S rRNA-based phylogeny and *lacS* and  $\beta$ -galactosidase gene based phylogenies.** Alignment of nucleotide sequences and tree construction were performed using ClustalW [26],[246] and trees were drawn using njplot [27]. **A**, lactose permease (*lacS* gene) phylogeny; **B**,  $\beta$ -galactosidase (*lacL* or *lacZ* gene) phylogeny; **C**, 16S rRNA phylogeny. \*,  $\beta$ -galactosidase large subunit (*lacL* gene), belonging to the glycoside hydrolase family 2; \*\*, homologous family 2  $\beta$ -galactosidase encoded in one gene (*lacZ*). Numbers indicate bootstrap values; The scale bar represents the mean number of nucleotide substitutions per site.



**Figure S4. Coherence between 16S rRNA-based phylogeny and *lacA* based phylogeny.** Alignment of nucleotide sequences and tree construction were performed using ClustalW [26],[246] and trees were drawn using njplot [27]. **A**, family 42  $\beta$ -galactosidase (*lacA* gene) phylogeny; **B**, 16S rRNA phylogeny. Numbers indicate bootstrap values; The scale bar represents the mean number of nucleotide substitutions per site.

## Supplementary tables

**Table S1. *L. delbrueckii* subspecies specific nucleotides in 16S rRNA sequences.**

Strain	Nucleotide position								
	7	9	21	224	432	690*	859	1273	1455
<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> ATCC11842	T	A	G	T	T	T	G	G	G
<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> ATCC BAA-365	T	A	G	T	T	T	G	G	G
<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> 2038	T	A	G	T	T	T	G	G	G
<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> VIB27	T	A	G	T	T	T	G	G	G
<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> VIB44	T	A	G	T	T	T	G	G	G
<i>L. delbrueckii</i> ssp. <i>lactis</i> ND02 (a)	T	A	G	C	C	C	G	G	G
<i>L. delbrueckii</i> ssp. <i>lactis</i> CNRZ226	T	A	G	C	C	C	G	G	-
<i>L. delbrueckii</i> ssp. <i>lactis</i> CNRZ327	T	A	G	C	C	C	G	G	G
<i>L. delbrueckii</i> ssp. <i>Lactis</i> CNRZ333	T	A	G	C	C	C	G	G	G
<i>L. delbrueckii</i> ssp. <i>lactis</i> CNRZ700	T	A	G	C	C	C	G	G	G
<i>L. delbrueckii</i> ssp. <i>delbrueckii</i> ATCC9649	A	C	-	C	C	C	A	A	A

16S rRNA sequences from *L. delbrueckii* strains belonging to the ssp *bulgaricus*, *lactis*, and *delbrueckii* were aligned using Clustalw [26]. Only differences that allow to distinguish subspecies are shown. Nucleotide position, position in the ATCC11842 sequence.

\*Mutation generating an *EcoRI* site in the *L. delbrueckii* ssp. *bulgaricus* 16S rDNA. a, originally classified as ssp. *bulgaricus* [10].

**Table S2. Number of IS elements in *L. delbrueckii* ssp. *bulgaricus* and ssp. *lactis*.**

Strain	Number of IS	Genome size (Mbp)
<i>Lactobacillus delbrueckii</i> ssp. <i>bulgaricus</i> ATCC 11842	56	1,86
<i>Lactobacillus delbrueckii</i> ssp. <i>bulgaricus</i> ATCC BAA-365	29	1,85
<i>Lactobacillus delbrueckii</i> ssp. <i>bulgaricus</i> 2038	54	1.87
<i>Lactobacillus delbrueckii</i> ssp. <i>lactis</i> ND02	133	2,12
<i>Lactobacillus delbrueckii</i> ssp. <i>lactis</i> CNRZ327	215	2,11

**Table S3. *L. delbrueckii* ssp. *bulgaricus* specific proteins.**

Locus_tag	Protein	Function
Ldb_1456	Putative permease	Transport/binding protein and lipoproteins
Ldb_0014	ABC transporter, ATP-binding/permease protein	Transport/binding protein and lipoproteins
Ldb_1891	Permease	Transport/binding protein and lipoproteins
Ldb_2094	ABC transporter, ATP-binding/permease protein	Transport/binding protein and lipoproteins
Ldb_0502	Putative permease	Transport/binding protein and lipoproteins
Ldb_1775	Ornithine decarboxylase	Metabolism of amino acids and related molecules
Ldb_0262	Pyridoxine 5 -phosphate oxidase V related favin-nucleotide-binding protein	Pyridoxal phosphate biosynthesis
Ldb_1782	Transposase	Transposon and IS
Ldb_1287	Hypothetical protein	ND
Ldb_1296	Hypothetical membrane protein	ND
Ldb_0158	Hypothetical protein	ND
Ldb_1547	Hypothetical protein	ND
Ldb_1548	Hypothetical protein	ND
Ldb_1547	Hypothetical protein	ND
Ldb_1682	Hypothetical protein	ND
Ldb_1743	Hypothetical protein	ND
Ldb_1766	Hypothetical membrane protein	ND
Ldb_1781	Conserved hypothetical membrane protein	ND
Ldb_1797	Hypothetical protein	ND
Ldb_0061	Hypothetical protein	ND
Ldb_0890	Hypothetical protein	ND
Ldb_0313	Hypothetical protein	ND
Ldb_0074	Hypothetical protein	ND
Ldb_0921	Hypothetical protein	ND
Ldb_0915	Hypothetical protein	ND

Presented proteins make part of the core proteome of the ssp. *bulgaricus* (5 strains), and are lacking from all 5 ssp. *lactis* strains in this study. Locus tag, identifier of the corresponding protein in *L. bulgaricus* ATCC11842 [6]. ND, not determined.

**Table S4. *L. delbrueckii* ssp. *lactis* specific proteins.**

Locus_tag	Protein	Function
LBCNRZ327_03570	ABC superfamily ATP binding cassette transporter, ABC protein	Transporter / binding protein
LBCNRZ327_02370	Oligopeptide binding protein	Transporter / binding protein
LBCNRZ327_00535	ABC transporter substrate binding protein	Transporter / binding protein
LBCNRZ327_01470	Oligopeptide ABC superfamily ATP binding cassette transporter, binding protein	Transporter / binding protein
LBCNRZ327_02360	Periplasmic oligopeptide-binding family protein	Transporter / binding protein
LBCNRZ327_02715	Maltose ABC superfamily ATP binding cassette transporter, permease protein	Transporter / binding protein of carbohydrates
LBCNRZ327_02725	Sugar ABC superfamily ATP binding cassette transporter, ABC protein	Transporter / binding protein of carbohydrates
LBCNRZ327_07155	Mannitol PTS, EIIA	Transporter / binding protein of carbohydrates
LBCNRZ327_02730	$\beta$ -phosphoglucomutase	Metabolism of carbohydrates and related molecules
LBCNRZ327_08295	Phosphoenolpyruvate synthase	Metabolism of carbohydrates and related molecules
LBCNRZ327_03650	Phosphoglycerate mutase	Transport and metabolism of carbohydrates
LBCNRZ327_02735	Maltose phosphorylase	Metabolism of carbohydrates and related molecules
LBCNRZ327_05415	Peptidoglycan-binding protein	Metabolism of carbohydrates and related molecules
LBCNRZ327_08780	Ribose-5-phosphate isomerase B	Metabolism of carbohydrates and related molecules
LBCNRZ327_05435	Phospho- $\beta$ -glycosidase	Metabolism of carbohydrates and related molecules
LBCNRZ327_07165	Mannitol-1-phosphate 5-dehydrogenase	Metabolism of carbohydrates and related molecules
LBCNRZ327_02750	Lacl family transcriptional regulator	Metabolism of carbohydrates and related molecules
LBCNRZ327_03615	Carbamoyl-phosphate synthase large chain 2	Metabolism of amino acids and related molecules
LBCNRZ327_09715	Ornithine carbamoyltransferase	Metabolism of amino acids and related molecules
LBCNRZ327_09720	Carbamate kinase	Metabolism of amino acids and related molecules
LBCNRZ327_09725	Arginine deiminase	Metabolism of amino acids and related molecules
LBCNRZ327_08995	3-dehydroquinate dehydratase	Metabolism of amino acids and related molecules
LBCNRZ327_03960	Argininosuccinate synthase	Metabolism of amino acids and related molecules
LBCNRZ327_04140	Glutamate 5-kinase	Metabolism of amino acids and related molecules
LBCNRZ327_04145	Gamma-glutamyl phosphate reductase	Metabolism of amino acids and related molecules
LBCNRZ327_04855	NAD-dependent aldehyde dehydrogenase	Metabolism of amino acids and related molecules
LBCNRZ327_05000	Glutamine ABC superfamily ATP binding cassette transporter, permease protein	Metabolism of amino acids and related molecules
LBCNRZ327_06445	Dipeptidase A	Metabolism of amino acids and related molecules
LBCNRZ327_08630	Proline-5-carboxylate reductase	Metabolism of amino acids and related molecules
LBCNRZ327_05430	D-serine/D-alanine/glycine:H <sup>+</sup> symporter	Transport/ binding of amino-acids
LBCNRZ327_05805	HNH endonuclease domain protein	Metabolism of nucleotides and nucleic acids
LBCNRZ327_05810	Dihydroorotase	Metabolism of nucleotides and nucleic acids
LBCNRZ327_08930	Aspartate carbamoyltransferase	Metabolism of nucleotides and nucleic acids
LBCNRZ327_03965	Acyl-CoA dehydrogenase	Lipid transport and metabolism
LBCNRZ327_03980	Transcriptional regulator (LysR family)	Transcription regulation
LBCNRZ327_05400	LysR substrate binding domain protein	Transcription regulation
LBCNRZ327_03570	Putative repressor LexA	Transcription regulation
LBCNRZ327_07145	Mannitol operon transcriptional antiterminator	Transcription regulation
LBCNRZ327_09025	LysR substrate binding domain protein	Transcription regulation
LBCNRZ327_06440	Signal recognition particle-docking protein FtsY	Protein secretion
LBCNRZ327_06170	Possible (S)-2-hydroxy-acid oxidase	Glyoxylate and dicarboxylate metabolism
LBCNRZ327_00415	Transposase	Transposon and IS
LBCNRZ327_00575	Acetyltransferase, GNAT family	N-acetyltransferase activity
LBCNRZ327_00695	Signal transduction diguanylate cyclase	phosphorus-oxygen lyase activity
LBCNRZ327_02595	Hemerythrin HHE cation binding domain protein	phosphorelay sensor kinase activity
LBCNRZ327_02600	N-6 adenine-specific DNA methylase YitW	site-specific DNA-methyltransferase (adenine-specific) activity
LBCNRZ327_03405	Glutathione-disulfide reductase	flavin adenine dinucleotide binding
LBCNRZ327_10040	Glutathione reductase	Posttranslational modification, protein turnover, chaperones
LBCNRZ327_10055	Possible ATPase involved in chromosome partitioning	Cell cycle control, cell division, chromosome partitioning
LBCNRZ327_10070	DNA starvation/stationary phase protection protein Dps	Inorganic ion transport and metabolism
LBCNRZ327_08230	DegV family protein	ND
LBCNRZ327_09545	Conserved domain protein	ND
LBCNRZ327_10310	Predicted nucleoside-diphosphate-sugar epimerase	ND
LBCNRZ327_10450	Putative uncharacterized protein	ND
LBCNRZ327_07930	Putative uncharacterized protein	ND
LBCNRZ327_05725	Putative uncharacterized protein	ND
LBCNRZ327_00715	Hypothetical protein	ND
LBCNRZ327_02430	Conserved domain protein	ND
LBCNRZ327_03420	Putative uncharacterized protein	ND
LBCNRZ327_04640	Putative uncharacterized protein	ND
LBCNRZ327_04730	Putative uncharacterized protein	ND
LBCNRZ327_04940	Putative uncharacterized protein	ND
LBCNRZ327_05055	Conserved domain protein	ND
LBCNRZ327_07360	Putative uncharacterized protein	ND
LBCNRZ327_07365	Putative uncharacterized protein	ND

Presented proteins make part of the core proteome of the ssp. *lactis* (5 strains), and are lacking from all 5 ssp. *bulgaricus* strains in this study. Locus tag, identifier of the corresponding protein in *L. lactis* CNRZ327. ND, not determined.

**Table S5. Genes involved in carbohydrate metabolism in *L. delbrueckii* ssp. *lactis* and ssp. *bulgaricus* strains.**

EC number	Enzyme	Gene name	ssp. <i>lactis</i> strains					ssp. <i>bulgaricus</i> strains				
			CNRZ226	CNRZ327	CNRZ333	CNRZ700	NDO2	11842	BAA365	VIB27	VIB44	2038
<b>Galactose metabolism</b>												
-	Transporter not identified	-										
2.7.1.6	Galactokinase	<i>galK</i>				F						
2.7.7.12	Galactose 1-P uridylyltransferase	<i>galT</i>				F						
	API profile		-	-	+	+	ND	-	-	-	-	ND
<b>Sucrose transport and metabolism</b>												
2.7.1.69	PTS family sucrose porter, EIIBC component	<i>ScrA</i>				F		F	F	F	F	F
3.2.1.26	Sucrose 6P hydrolase (sucrase)	<i>ScrB</i>				F		F	F	F	F	F
2.7.1.4	Fructokinase	<i>ScrK</i>				F		F	F	F	F	F
	API profile		+	-	+	-	ND	-	-	-	-	ND
<b>Maltose transport and metabolism</b>												
-	Maltose ABC transporter permease protein (1)	<i>MalE/F/G</i>				F		F	F	F	F	F
2.4.1.8	Maltose phosphorylase (1)	<i>mapA</i>										
5.4.2.6	$\beta$ -phosphoglucomutase (1)	<i>pgmA</i>										
-	Maltose ABC transporter permease protein (2)	<i>malE/F/G</i>				F		F	F	F	F	F
3.2.1.20	Oligo-1,6-glucosidase (2)	<i>mall</i>										

2.7.1.69	PTS –Glucose specific EIIA (3)	-										
3.2.1.122	Maltose-6'-phosphate glucosidase (3)	<i>malH</i>		F						F		
	API profile		+	-	+	-	ND	-	-	-	-	ND
<b>Cellobiose transport and metabolism</b>												
2.7.1.69	PTS system, cellobiose-specific IIABC component	<i>celA/B/C</i>		F	F	F	F	F	F	F	F	F
3.2.1.86	6-Phospho-β-glucosidase	<i>pbg3</i>				F					F	
5.3.1.9	Glucose-6-phosphate isomerase	-										
-	Transcriptional regulator	<i>yeca</i>										
	API profile		+	-	-	-	ND	-	-	-	-	ND
<b>Arbutin transport and metabolism</b>												
2.7.1.69	PTS system, arbutin-specific	-				F						
3.2.1.86	6-phospho-β-glucosidase	<i>pbg3</i>				F					F	
5.3.1.9	Glucose-6-phosphate isomerase	-										
	API profile		+	-	+	-	ND	-	-	-	-	ND
<b>Salicin transport and metabolism</b>												
-	Unidentified transporter	-										
3.2.1.86	6-Phospho-β-glucosidase	<i>pbg3</i>				F					F	
5.3.1.9	Glucose-6-phosphate isomerase	-										
	API profile		+	-	+	-	ND	-	-	-	-	ND
<b>D-glucose transport and metabolism</b>												
2.7.1.69	PTS system, glucose/glucoside-specific enzyme	-										



Fructose transport and metabolism										
2.7.1.69	PTS system, fructose-specific enzyme IIBC component (1)	<i>fruB</i>								
2.7.1.56	1-Phosphofructokinase (1)	<i>fruk</i>								
-	Fructose repressor (1)	<i>fruR</i>					F		F	
-	Unidentified Fructose permease (2)									
2.7.1.4	Fructokinase (2)	<i>ScrK</i>			F		F	F	F	F
2.7.1.11	6-Phosphofructokinase(2)	<i>pfk2</i>								
4.1.2.13	Fructose-bisphosphate aldolase (2)	<i>fba</i>								
	API profile		+	+	+	-	ND	-	-	-
N-acetylglycosamin transport and metabolism										
-	Unidentified transporter	-								
3.5.1.25	N-acetylglucosamine-6-phosphate deacetylase	<i>nagA</i>								F
3.5.99.6	Glucosamine-6-phosphate deaminase	<i>nagB</i>				F	F	F	F	F
	API profile		+	+	+	+	ND	-	-	-
Trehalose transport and metabolism										
2.7.1.69	PTS family trehalose porter, IIBC component	<i>treB</i>								
3.2.1.93	Trehalose-6-phosphate hydrolase	<i>dexS</i>								
	API profile		+	-	+	+	ND	-	-	-
Esculin transport and metabolism										

-	Pathway not described by KEGG	-											
	API profile		+	+	+	+	ND	-	-	+	-	ND	
<b>Amygdalin transport and metabolism</b>													
-	Pathway not described by KEGG	-											
	API profile		+	-	-	-	ND	-	-	-	-	ND	
<b>Starch metabolism</b>													
3.2.1.135	Neopullulanase*	<i>nplT</i>											
3.2.1.70	Glucan 1,6-alpha -glucosidase *	<i>dexB</i>						F	F	F	F	F	
	Degradation profile		+	+	+	+	ND	-	-	-	-	ND	

Grey box, gene is present; white box, gene is absent; F, gene fragment(s) only. API, results of experimental determination of capacity to ferment the carbohydrate studied: +, fermentation observed; -, no fermentation observed; ND, not determined. (1), (2), (3) designate alternative pathways.

\*These genes are absent from table S4 generated by single linkage protein clustering due to limitations imposed by the choice of criteria for automatic protein clustering.

**Table S6. Genes involved in amino acid metabolism in *L. delbrueckii* ssp. *lactis* and ssp. *bulgaricus* strains.**

EC number	Enzyme	Gene name	ssp. <i>lactis</i> strains					ssp. <i>bulgaricus</i> strains				
			CNRZ226	CNRZ327	CNRZ333	CNRZ700	NDO2	11842	BAA365	VIB27	VIB44	2038
<b>Threonine biosynthesis</b>												
2.7.2.4	Aspartate kinase	<i>ask</i>										
1.2.1.11	Aspartate-semialdehyde dehydrogenase	-										
1.1.1.3	Homoserine dehydrogenase	<i>thrA</i>	F									
2.7.1.39	Homoserine kinase	<i>thrB</i>										
4.2.3.1	Homoserine synthase	<i>thrC</i>										
<b>Lysine biosynthesis</b>												
2.7.2.4	Aspartokinase	<i>lysC</i>										
1.2.1.11	Aspartate semialdehyde dehydrogenase	<i>Asd</i>										
4.3.3.7	Dihydrodipicolinate synthase	<i>dapA</i>										
1.3.1.26	Dihydrodipicolinate reductase	<i>dapB</i>										
2.3.1.89	2,3,4,5Tetrahydropyridine-2,6-dicarboxylate N-acetyl-transferase	<i>dapD</i>										
2.6.1.-	Transaminase	<i>aspC1</i>										
3.5.1.47	N-acetyldiaminopimelate deacetylase	<i>hipO</i>										
5.1.1.7	Diamino pimelate epimerase	<i>dapF</i>										

4.1.1.20	Diamino pimelate decarboxylase	<i>lysA</i>												
<b>Arginine biosynthesis</b>														
6.3.1.2	Glutamine synthetase	<i>glnA</i>												
2.7.2.2	Carbamate kinase	<i>arcC</i>												
2.1.3.3	Ornithine carbamoyl transferase	<i>argF</i>												
6.3.4.5	Arginino succinate synthase	<i>argG</i>												
4.3.2.1	Arginino succinate lyase	<i>argH</i>							F	F	F	F	F	
3.5.3.6	Arginine deiminase	<i>arcA</i>												
<b>Aspartic acid biosynthesis</b>														
4.3.1.1	Aspartate ammonia-lyase	<i>aspA</i>												
6.3.5.4	Asparagine synthetase	<i>asnB</i>												
2.6.1.1	Aspartate aminotransferase	<i>aspC1</i>												
<b>Proline biosynthesis</b>														
2.7.2.11	Glutamate 5-kinase	<i>proB</i>		F					F	F	F	F	F	
1.2.1.41	Gamma-glutamyl phosphate reductase	<i>proA</i>												
1.5.1.2	Pyroline-5-carboxylate reductase	<i>proC</i>												
<b>Serine biosynthesis</b>														
1.1.1.95	Phosphoglycerate dehydrogenase	<i>serA</i>												
2.6.1.52	Phosphoserine aminotransferase	<i>serC</i>												
3.1.3.3	Phosphoserine phosphatase	<i>serB</i>												
<b>Tryptophan biosynthesis</b>														

1.1.1.95	Phosphoglycerate dehydrogenase	<i>serA</i>																	
2.6.1.52	Phosphosérine aminotransferase	<i>serC</i>																	
3.1.3.3	Phosphoserine phosphatase	<i>serB</i>																	
4.2.1.20	Tryptophan synthase $\beta$ -chain	<i>trpB</i>																	
<b>Tyrosine biosynthesis</b>																			
4.2.1.10	3-Dehydroquinate dehydratase	<i>aroD</i>																	
1.1.1.25	Shikimate dehydrogenase	<i>aroE</i>	F		F	F	F												
2.7.1.71	Shikimate Kinase	<i>aroK</i>	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
2.5.1.19	3-Phosphoshikimate 1-carboxyvinyl transferase	<i>aroA</i>						F	F	F	F	F	F	F	F	F	F	F	F
4.2.3.5	Chorismate synthase	<i>aroC</i>							F	F	F	F							
5.4.99.5	Chorismate mutase	<i>pheB</i>																	
1.3.1.12	Prephenate dehydrogenase	-																	
2.6.1.57	Aromatic amino acid aminotransferase	-																	
<b>Phenylalanine biosynthesis</b>																			
4.2.1.10	3-Dehydroquinate dehydratase	<i>aroD</i>																	
1.1.1.25	Shikimate dehydrogenase	<i>aroE</i>	F		F	F	F												
2.7.1.71	Shikimate Kinase	<i>aroK</i>	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
2.5.1.19	3-Phosphoshikimate 1-carboxyvinyl transferase	<i>aroA</i>						F	F	F	F	F	F	F	F	F	F	F	F
4.2.3.5	Chorismate synthase	<i>aroC</i>							F	F	F	F							
5.4.99.5	Chorismate mutase	<i>pheB</i>																	
4.2.1.51	Prephenate dehydratase	-																	

2.6.1.57	Aromatic amino acid aminotransferase	-											
<b>Cysteine biosynthesis</b>													
2.3.1.30	Serine O-acetyltransferase	<i>cysE</i>						F	F	F	F	F	F
2.5.1.49	Cysteine synthase	<i>cysK</i>											
<b>Methionine biosynthesis</b>													
2.3.1.30	Serine O-acetyltransferase	<i>cysE</i>						F	F	F	F	F	F
2.5.1.49	Cysteine synthase	<i>cysK</i>											
4.4.1.1	Cystathionine gamma-lyase	-											
4.4.1.8	Cystathionine β-lyase	<i>patC</i>											
2.1.1.10	Homocysteine S-methyltransferase	-											
<b>Glutamine biosynthesis</b>													
6.3.1.2	Glutamine synthetase	<i>glnA</i>											
<b>Alanine biosynthesis</b>													
2.6.1.21	D-amino-acid transaminase	-											
2.6.1.41	D-methionine—pyruvate transaminase	-											
5.1.1.1	Alanine racemase	<i>Alr</i>											
<b>Glutamic acid biosynthesis</b>													
1.4.1.2/3/4	Glutamate dehydrogenase	-											
2.6.1.1	Aspartate aminotransferase	-											
1.4.1.13/14	Glutamate synthase	-											
<b>Glycine biosynthesis</b>													

2.1.2.1	Glycine hydroxymethyltransferase	-												
4.1.2.5	Threonine aldolase	-												
<b>Valine biosynthesis</b>														
4.1.3.18	Acetolactate synthase	-												
5.4.99.3	2-Acetolactate mutase	-												
1.1.1.86	Ketol-acid reductoisomerase	-												
4.2.1.9	Dihydroxy-acid dehydratase	-												
<b>Leucine biosynthesis</b>														
4.1.3.18	Acetolactate synthase	-												
5.4.99.3	2-Acetolactate mutase	-												
1.1.1.86	Ketol-acid reductoisomerase	-												
4.2.1.9	Dihydroxy-acid dehydratase	-												
<b>Isoleucine biosynthesis</b>														
4.3.1.19	Threonine ammonia-lyase	-												
4.1.3.18	Acetolactate synthase	-												
1.1.1.86	Ketol-acid reductoisomerase	-												
4.2.1.9	Dihydroxy-acid dehydratase	-												
2.6.1.42	Branched chain amino acid aminotransferase	-												
<b>Histidine biosynthesis</b>														
2.7.1.15	Ribokinase	<i>rbsK</i>												
2.7.6.1	Ribose-P pyrophosphokinase	<i>Prs</i>												

2.4.2.17	ATP phosphoribosyltransferase	-																	
3.6.1.31	Phosphoribosyl-ATP pyrophosphatase	-																	
3.5.4.19	Phosphoribosyl-AMP cyclohydrolase	-																	
5.3.1.16	1-(5-Phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase	-																	
2.4.2.-	Pentosyl ribosyltransferase	-																	
4.2.1.19	Imidazoleglycerol-phosphate dehydratase	-																	
2.6.1.19	Histidinol phosphate aminotransferase	-																	
3.1.3.15	Histidinol phosphatase	-																	
1.1.1.23	Histidinol dehydrogenase	-																	
<b>Asparagine biosynthesis</b>																			
6.3.1.1	Aspartate-ammonia ligase	<i>asnA</i>																	
6.3.5.4	Asparagine synthetase	<i>asnB</i>																	

Grey box, gene is present; white box, gene is absent; F, gene fragment(s) only.

**Table S7. Genes involved in acid stress resistance.**

EC number	Function	Gene Name	<i>ssp. lactis</i> strains					<i>ssp. bulgaricus</i> strains				
			CNRZ226	CNRZ327	CNRZ333	CNRZ700	NDO2	11842	BAA365	VIB27	VIB44	2038
<b>Proton pumps and related function</b>												
3.6.3.14	H <sup>+</sup> transporting ATPase / ATP synthase, A subunit	<i>atpB</i>										
3.6.3.14	H <sup>+</sup> transporting ATPase / ATP synthase, C subunit	<i>atpE</i>										
3.6.3.14	H <sup>+</sup> transporting ATPase / ATP synthase, B subunit	<i>atpF</i>										
3.6.3.14	H <sup>+</sup> transporting ATPase / ATP synthase, delta subunit	<i>atpH</i>										
3.6.3.14	H <sup>+</sup> transporting ATPase / ATP synthase, alpha subunit	<i>atpA</i>										
3.6.3.14	H <sup>+</sup> transporting ATPase / ATP synthase, gamma subunit	<i>atpG</i>										
3.6.3.14	H <sup>+</sup> transporting ATPase / ATP synthase, beta subunit	<i>atpD</i>										
3.6.3.14	H <sup>+</sup> transporting ATPase / ATP synthase, epsilon subunit	<i>atpC</i>										
<b>Other functions</b>												
4.1.1.15	Glutamate decarboxylase	<i>gadA</i>		F	F	F	F					
4.1.1.17	Ornithine decarboxylase 1	-										
4.1.1.17	Ornithine decarboxylase 2	-										
3.5.3.6	Arginine deiminase	<i>arcA</i>				F						
2.1.3.3	Ornithine carbamoyltransferase	<i>arcB</i>										
2.7.2.2	Carbamate kinase	<i>arcC</i>										
-	Arginine-ornithine antiporter	<i>arcD</i>		F			F	F	F	F	F	F

Grey box, gene is present; white box, gene is absent; F, gene fragment(s) only.

**Acknowledgments.** This project was supported by INRA AIP Bio-Ressources and by a grant from the Région Ile-de-France. J. Binesse is supported by the project "Surfing" (ANR-2010-ALIA-016) financed by the French National Research Agency (ANR).

## References

1. Weiss N, Schillinger U, Kandler O: **Lactobacillus lactis, Lactobacillus leichmannii and Lactobacillus bulgaricus, Subjective Synonyms of Lactobacillus delbrueckii, and Description of Lactobacillus delbrueckii subsp. lactis comb. nov. and Lactobacillus delbrueckii subsp. bulgaricus comb. nov.** *Systematic and applied microbiology* 1983, **4**(4):552-557.
2. Giraffa G, De Vecchi P, Rossetti L: **Note: identification of Lactobacillus delbrueckii subspecies bulgaricus and subspecies lactis dairy isolates by amplified rDNA restriction analysis.** *Journal of applied microbiology* 1998, **85**(5):918-924.
3. Tanigawa K, Watanabe K: **Multilocus sequence typing reveals a novel subspeciation of Lactobacillus delbrueckii.** *Microbiology* 2011, **157**(Pt 3):727-738.
4. de Vos WM, Vaughan EE: **Genetics of lactose utilization in lactic acid bacteria.** *FEMS microbiology reviews* 1994, **15**(2-3):217-237.
5. Germond JE, Lapierre L, Delley M, Mollet B, Felis GE, Dellaglio F: **Evolution of the bacterial species Lactobacillus delbrueckii: a partial genomic study with reflections on prokaryotic species concept.** *Molecular biology and evolution* 2003, **20**(1):93-104.
6. van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, Robert C, Oztas S, Mangenot S, Couloux A *et al*: **The complete genome sequence of Lactobacillus bulgaricus reveals extensive and ongoing reductive evolution.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(24):9274-9279.
7. Teuber M: **Lactic acid bacteria.** In: *Biotechnology*. Edited by Rehm H-J, Reed G, Pühler A, Stadler P, vol. 1, 2nd edn. Weinheim: Verlag Chemie; 1993: 325-366.
8. 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.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(42):15611-15616.
9. Hao P, Zheng H, Yu Y, Ding G, Gu W, Chen S, Yu Z, Ren S, Oda M, Konno T *et al*: **Complete sequencing and pan-genomic analysis of Lactobacillus delbrueckii subsp. bulgaricus reveal its genetic basis for industrial yogurt production.** *PloS one* 2011, **6**(1):e15964.
10. Sun Z, Chen X, Wang J, Zhao W, Shao Y, Guo Z, Zhang X, Zhou Z, Sun T, Wang L *et al*: **Complete genome sequence of Lactobacillus delbrueckii subsp. bulgaricus strain ND02.** *Journal of bacteriology* 2011, **193**(13):3426-3427.
11. El Kafsi H, Binesse J, Loux V, Buratti J, Boudebouze S, Dervyn R, Hammani A, Maguin E, van de Guchte M: **Genome sequence of Lactobacillus delbrueckii ssp. lactis CNRZ327, a dairy bacterium with anti-inflammatory properties.** *Submitted*.
12. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M: **ISfinder: the reference centre for bacterial insertion sequences.** *Nucleic acids research* 2006, **34**(Database issue):D32-36.
13. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic acids research* 2000, **28**(1):27-30.
14. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: **KAAS: an automatic genome annotation and pathway reconstruction server.** *Nucleic acids research* 2007, **35**(Web Server issue):W182-185.
15. Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P: **iPath2.0: interactive pathway explorer.** *Nucleic acids research* 2011, **39**(Web Server issue):W412-415.

16. Hickey MW, Hillier AJ, Jago GR: **Transport and metabolism of lactose, glucose, and galactose in homofermentative lactobacilli.** *Applied and environmental microbiology* 1986, **51**(4):825-831.
17. Poolman B, Royer TJ, Mainzer SE, Schmidt BF: **Lactose transport system of *Streptococcus thermophilus*: a hybrid protein with homology to the melibiose carrier and enzyme III of phosphoenolpyruvate-dependent phosphotransferase systems.** *Journal of bacteriology* 1989, **171**(1):244-253.
18. Poolman B: **Precursor/product antiport in bacteria.** *Molecular microbiology* 1990, **4**(10):1629-1636.
19. Gilbert C, Atlan D, Blanc B, Portailier R, Germond JE, Lapierre L, Mollet B: **A new cell surface proteinase: sequencing and analysis of the prtB gene from *Lactobacillus delbrueckii* subsp. *bulgaricus*.** *Journal of bacteriology* 1996, **178**(11):3059-3065.
20. van de Guchte M, Serror P, Chervaux C, Smokvina T, Ehrlich SD, Maguin E: **Stress responses in lactic acid bacteria.** *Antonie van Leeuwenhoek* 2002, **82**(1-4):187-216.
21. Lapierre L, Mollet B, Germond JE: **Regulation and adaptive evolution of lactose operon expression in *Lactobacillus delbrueckii*.** *Journal of bacteriology* 2002, **184**(4):928-935.
22. Gilbert C, Blanc B, Frot-Coutaz J, Portailier R, Atlan D: **Comparison of cell surface proteinase activities within the *Lactobacillus* genus.** *J Dairy Res* 1997, **64**:561-571.
23. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT: **Reordering contigs of draft genomes using the Mauve aligner.** *Bioinformatics (Oxford, England)* 2009, **25**(16):2071-2073.
24. Bryson K, Loux V, Bossy R, Nicolas P, Chaillou S, van de Guchte M, Penaud S, Maguin E, Hoebeke M, Bessieres P *et al*: **AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system.** *Nucleic acids research* 2006, **34**(12):3533-3545.
25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Molecular biology and evolution* 2011, **28**(10):2731-2739.
26. Thompson JD, Gibson TJ, Higgins DG: **Multiple sequence alignment using ClustalW and ClustalX.** *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]* 2002, **Chapter 2**:Unit 2 3.
27. Perriere G, Gouy M: **WWW-query: an on-line retrieval system for biological sequence banks.** *Biochimie* 1996, **78**(5):364-369.

## Conclusion

Les deux bactéries lactiques *L. delbrueckii* ssp. *lactis* et *L. delbrueckii* ssp. *bulgaricus* sont toutes deux en phase d'évolution conduisant à une adaptation au milieu lait. Toutefois, la ssp. *bulgaricus* semble avoir atteint un stade d'évolution plus avancé que celui de la ssp. *lactis*. L'adaptation des deux ssp. à leur environnement se fait principalement par un phénomène de perte spontanée de gènes ayant des fonctions superflues dans cet environnement.

**V.C Omics approaches to characterize anti-inflammatory properties of  
*Lactobacillus delbrueckii***

## Introduction

*Lactobacillus delbrueckii* est une espèce bactérienne qui compte quatre sous-espèces : *delbrueckii*, *bulgaricus*, *lactis* et *indicus* [247]. Les sous-espèces *bulgaricus* et *lactis* sont les plus utilisées dans l'industrie laitière pour leurs propriétés technologiques de fermentation : la sous-espèce *bulgaricus* est surtout utilisée dans la fermentation de yaourt, alors que la sous-espèce *lactis* est traditionnellement utilisée dans la fabrication de fromage du type parmesan.

En plus du rôle biotechnologique que jouent ces deux sous-espèces bactériennes dans la fabrication des produits laitiers, des preuves de leur potentiel probiotique s'accumulent. Plusieurs études montrent des effets anti-inflammatoires *in vitro* [213, 214, 219] ou un rôle bénéfique de protection contre l'inflammation induite chimiquement dans des modèles de maladies inflammatoires de l'intestin chez la souris [219, 247]. Santos Rocha *et al.* [219] ont établi *in vitro* les effets anti-inflammatoires de 57 souches de *L. delbrueckii* et montrent que l'ampleur de ces effets dépend de la souche considérée. Les mêmes auteurs ont établi que *L. delbrueckii* empêche la phosphorylation de l' $\text{I-}\kappa\text{B}$ , inhibiteur de  $\text{NF-}\kappa\text{B}$ , ce qui empêche la dégradation et conduit à une accumulation de  $\text{NF-}\kappa\text{B}$ , un facteur clé dans l'établissement d'une réponse inflammatoire. Des protéines de la surface bactérienne seraient responsables de l'effet anti-inflammatoire de *L. delbrueckii*.

En se basant sur ces résultats, nous utilisons dans la présente étude des approches comparables de génomique, protéomique et transcriptomique afin d'identifier la ou les protéine(s) responsable(s) des effets anti-inflammatoires de *L. delbrueckii*.

## Matériel et Méthodes

**Souches bactériennes et plasmides utilisées.** Les souches bactériennes ainsi que les plasmides utilisés dans cette étude sont listés dans le tableau 1. Les souches de *L. delbrueckii* ont été obtenues de la collection INRA et appartiennent aux sous-espèces *L. delbrueckii* ssp. *bulgaricus* et *L. delbrueckii* ssp. *lactis*. Lorsque l'origine est connue, les souches de la ssp.

*bulgaricus* proviennent de yaourts, alors que les souches de la ssp. *lactis* ont été isolées d'emmental ou correspondent à des ferments utilisés pour produire ce fromage.

Les souches de *L. delbrueckii* ont été cultivées à 42°C, dans le bouillon MRS (deMan, Rogosa, et Sharpe) ou sur le même milieu solidifié avec 1,5 % d'agar, en condition semi anaérobie. Les souches d'*Escherichia coli* et de *Bacillus subtilis* ont été cultivées à 37°C, dans le bouillon LB (Luria-Betani) sous agitation, ou sur le même milieu solidifié avec 1,5 % d'agar. Des antibiotiques ont été rajoutés aux cultures bactériennes selon les besoins comme suit : Pour *B. subtilis* : 50 µg/ml de spectinomycine, 5 µg/ml de chloramphénicol et 5 µg/ml de kanamycine. Pour *E. coli*, 100 µg/ml de spectinomycine, 100 µg/ml d'ampicilline, 25 µg/ml de kanamycine.

**Tableau 1. Souches bactériennes et plasmides utilisés**

<b>Souches</b>	<b>Génotype</b>	<b>Propriétés</b>	<b>Référence</b>
<i>B. subtilis</i> VI7692	KA8AX derivative; ( <i>yvyD-yvzG-fliT-fliS-fliD-yvyC-hag</i> )::phleo	Déficiente en 9 protéases et en production de flagelline	[248]
<i>E. coli</i> TOP10	F <sup>-</sup> <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139 Δ(ara-leu)</i> 7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG λ-</i>	Bactérie hôte dans le clonage avec une sélection <i>ccdB</i>	Invitrogene
<i>E. coli</i> TG1	F <sup>+</sup> [ <i>traD36 proAB+ lacIq lacZ</i> ΔM15] <i>supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5,</i> (rK <sup>-</sup> mK <sup>-</sup> )	Génération de plasmides multimère utilisés dans la transformation des de <i>B.subtilis</i> compétentes	Lucigene
<i>Lactobacillus delbrueckii</i> ssp. <i>lactis</i> CNRZ327 (CIRM1029)	-	Effet anti-inflammatoire élevé	[249]
<i>Lactobacillus delbrueckii</i> ssp. <i>lactis</i> CNRZ333 (CIRM1512)	-	Effet anti-inflammatoire élevé	-
<i>Lactobacillus delbrueckii</i> ssp. <i>lactis</i> CNRZ226 (CIRM1030)	-	Effet anti-inflammatoire Intermédiaire	-
<i>Lactobacillus delbrueckii</i> ssp. <i>lactis</i> CNRZ700 (CIRM1533)	-	Effet anti-inflammatoire Intermédiaire	-
<i>Lactobacillus delbrueckii</i> ssp. <i>bulg</i> ATCC11842 (CIRM1033)	-	Effet anti-inflammatoire Intermédiaire	[167]

<i>Lactobacillus delbrueckii</i> ssp. <i>bulg</i> ATCC-BAA365 (CIRM1062)	-	Effet anti-inflammatoire Intermédiaire	[218]
<i>Lactobacillus delbrueckii</i> ssp. <i>bulg</i> VIB27 (CIRM1034)	-	Effet anti-inflammatoire Intermédiaire	-
<i>Lactobacillus delbrueckii</i> ssp. <i>bulg</i> VIB44 (CIRM1032)	-	Effet anti-inflammatoire Intermédiaire	-
<b>Plasmides</b>			
pDONR223	Spcr ccdB+	Vecteur d'entrée Gateway	Invitrogen
pDG148-GW	Kanr Ampr Cmr ccdB+	Vecteur d'expression, Gateway	[248]

**Méthodes protéomiques. Culture bactérienne.** Pour les analyses protéomiques, les bactéries ont été incubées à 42°C dans 100 ml de bouillon MRS, jusqu'à une densité optique comprise entre 0,3 et 0,8, et récoltées par centrifugation à 6000 g pendant 10 min à 4°C.

*Extraction des protéines de surface à la trypsine.* Pour l'extraction des protéines de surface par décapage à la trypsine, les bactéries sont lavées deux fois avec du PBS, et suspendues dans 0,5 ml de PBS (pH=8,5) avec 5mM dithiothreitol (DTT). La suspension bactérienne est incubée avec 20 µg de trypsine (Promega) pendant une heure à 37°C, sous agitation (120 rpm), et ensuite centrifugée (8000 g, 5 min). Les surnageants obtenus sont filtrés avec un filtre de 0,2 µm et incubés avec 1 µg de trypsine à 37°C, pendant la nuit et sous agitation (120 rpm). La digestion à la trypsine est arrêtée en ajoutant de l'acide trifluoroacétique (Sigma-Aldrich). Les surnageants sont ensuite séchés sous vide et stockés à -20°C en attendant d'être analysés par spectrométrie de masse.

*Extraction des protéines de surface par le chlorure de lithium.* Les bactéries sont lavées deux fois avec du PBS, resuspendues dans 2 ml d'une solution de chlorure de lithium 5 M (Sigma), et incubées à 0°C pendant 15 min. Après centrifugation (21000 g, 18 min, 4°C), les surnageants contenant les protéines de surface sont dialysés contre du SDS à 0,1% pendant une nuit. Les surnageants sont ensuite séchés sous vide et stockés à -20 ° C.

*SDS-PAGE.* Les échantillons protéiques extraits par chlorure de lithium sont dilués dans 15 µl du tampon Laemmli (SDS, β mercaptoéthanol et bleu de Bromophénol) et dénaturés à 100°C pendant 3 min. Les extraits protéiques sont ensuite séparés par SDS-PAGE sur un gel d'acrylamide 12% suivant les conditions décrites par Laemmli [250]. La migration est faite à 80 V pendant 2 h à température ambiante. Les gels ont été colorés avec du « BioSafe colloidal Coomassie » (Biorad). Les spots protéiques sont excisés et soumis à une digestion à la trypsine comme décrit précédemment [251].

*Analyse Nano-LC MS/MS.* L'identification des protéines a été réalisée par spectrométrie de masse tandem (MS/MS) comme décrit précédemment [252]. Toutes les données obtenues par spectrométrie de masse ont été traitées avec le logiciel MASCOT v.2.2 pour la recherche de protéines dans une base de données concaténée (concaténation entre la base de données Swiss-Prot (<http://www.expasy.org>) et une base de données locale contenant entre autre les protéines prédites à partir des génomes séquencés de *Lactobacillus delbrueckii*).

La tolérance de masse des peptides a été fixée à 0,2 Da pour l'analyse MS simple et à 0,15 Da pour l'analyse MS/MS. Une protéine est considérée comme identifiée si au moins 2 peptides homologues avec un score Mascot ayant une valeur p inférieure à 0,05 est obtenue. Pour la validation des peptides, la version 1.19.2 du logiciel IRMa a été utilisée [253].

**Transcriptomique.** L'ARN est extrait de cultures en phase exponentielle de croissance en utilisant le TRIzol® Reagent (Invitrogen) et purifié sur colonne avec le kit RNeasy (Qiagen). Les cultures bactériennes à partir desquelles les ARN ont été extraits, ont été testées en parallèle pour leur effet anti-inflammatoire *in vitro* avec des cellules HT-29 contenant un rapporteur pour le gène NF-κB afin de valider leur utilisation. Pour l'étape d'hybridation, des puces pangénomiques NimbleGen 4x72k (Roche) ont été conçues. L'étape d'hybridation a été sous-traitée à la plate-forme Biopuces (INSA Toulouse). Les résultats ont été normalisés par quantiles [254] avant analyse.

**Clonage des gènes potentiellement responsables pour l'effet anti-inflammatoire dans *B. subtilis*.** Les séquences codantes (CDS) sélectionnées ont été amplifiées et clonées sous contrôle d'un promoteur inductible à l'IPTG dans le plasmide pDG148-GW comme déjà décrit [248]. Brièvement, les CDS sont amplifiées par PCR nichée (nested-PCR) avec des amorces spécifiques à chaque CDS, et des amorces universelles contenant (i) un site *attB1.1* et un site de fixation du ribosome (RBS) (côté 5' de la CDS) ou (ii) un site *attB2.1* (côté 3' de

la CDS). Les codons « start » et « stop » utilisés dans les amorces sont systématiquement ATG et TAG (amber).

Les produits PCR ainsi obtenus sont clonés dans le plasmide pDONR223 (Invitrogen) par recombinaison site spécifique *in vitro* entre les sites *attB1.1* et *attB2.1* et les sites *attP1* et *attP2* présents dans le plasmide en utilisant la BP clonase II (Invitrogen). Le mélange ainsi obtenu est utilisé pour la transformation d'*E. coli* (OneShot<sup>R</sup> TOP10, Invitrogen). 25 à 50 des clones obtenus (clones entrée) sont « poolés ». La présence de l'insert dans le plasmide pDONR223 est vérifiée par digestion enzymatique avec *BsrGI* (New England Biolabs) et par amplification PCR en utilisant des amorces spécifiques au plasmide pDONR223 (5'-CCCAGTCACGACGTTGTAAAACG et 5'-GTAACATCAGAGATTTTGAGACAC). L'insert maintenant flanqué par les sites *attL1* et *attL2*, est ensuite transféré *in vitro* dans le plasmide pDG148-GW par recombinaison site spécifique sous l'action de la LR clonase II (Invitrogen), entre les sites *attL1* et *attL2*, et les sites *attR1* and *attR2* présents dans pDG148-GW. Le mélange obtenu est utilisé pour la transformation d'*E. coli* TOP10. Les colonies obtenues sont « poolées ». La présence de l'insert dans le plasmide pDG148-GW est vérifiée par digestion enzymatique avec *EcoRI* (New England Biolabs).

Le mélange de plasmides obtenu est ensuite transféré dans la souche TG1 d'*E. coli* qui permet d'obtenir après extraction d'ADN des formes monomériques et multimériques des plasmides. Seules les formes multimériques seront adaptées pour la transformation de *B. subtilis* [255]. Pour chaque CDS, les clones obtenus sont « poolés » et les plasmides extraits sont utilisés pour transformer *B. subtilis* VI7695 [248]. Trois clones de *B. subtilis* sont ensuite vérifiés par amplification par PCR sur colonies en utilisant des amorces spécifiques au vecteur (5'-CGCACCTGAAGAAGATTTA et 5'-GCCGACTCAAACATCAAATC). Un clone est gardé pour la suite de l'expérience.

**Expression hétérologue dans *B. subtilis* de gènes candidats.** L'expression des gènes clonés dans le pDG148-GW dans *B. subtilis* est induite par l'ajout de 2 mM d'Isopropyl β-D-

1-thiogalactopyranoside (IPTG). L'induction de l'expression des gènes d'intérêt est réalisée à partir de cultures en phase exponentielle de *B. subtilis* ( $DO_{600nm} = 0,4$ ) à 37°C en LB pendant 3 h. Une culture contrôle (sans induction du gène d'intérêt) est préparée en parallèle.

**Evaluation des effets immuno-modulateurs des souches de *B. subtilis* transformées.** Afin d'évaluer l'effet immuno-modulateur de *B. subtilis* transformé, une lignée de cellules épithéliales du colon humain, HT-29, portant un gène rapporteur (codant pour la luciférase) situé dans le chromosome sous le contrôle d'un promoteur inductible par le NF- $\kappa$ B (HT-29/NF- $\kappa$ B-luc-E) est utilisée. Les cellules HT-29/NF- $\kappa$ B-luc-E sont incubées à 37°C (5% CO<sub>2</sub>) dans le milieu RPMI 1640 (Lonza). Le milieu de culture est supplémenté avec 2 mM L-glutamine, 50 IU/ml de pénicilline, 50  $\mu$ g/ml de streptomycine et 10% de sérum fœtal de veau inactivé par la chaleur (FCS, Lonza). Les cellules HT-29/NF- $\kappa$ B-luc-E sont réparties dans une plaque à 96 puits, à hauteur de  $5 \times 10^4$  cellules/puit (3917 assay plate, Costar, Cambridge, MA), et incubées pendant 48 heures dans 100  $\mu$ l de milieu RPMI complémenté. Le milieu RPMI est changé avant la mise en contact avec les bactéries sans ou après induction par IPTG.

La quantité de culture bactérienne nécessaire pour obtenir 100  $\mu$ l de suspension bactérienne avec une  $DO_{600nm}$  de 0,5 est centrifugée. Le culot bactérien est ensuite lavé et resuspendu dans 100  $\mu$ l de RPMI. 10  $\mu$ L ( $5 \times 10^6$  bactéries) de cette suspension est ajoutée aux cellules HT-29/NF- $\kappa$ B-luc-E (MOI= 25) pour obtenir un volume final de 100  $\mu$ L.

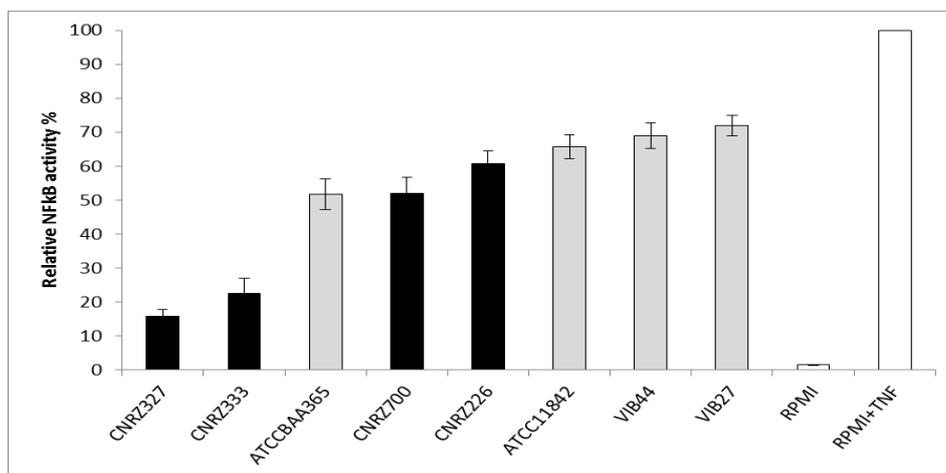
La cytokine TNF- $\alpha$  (Tumor necrosis factor alpha) (PeproTech, Rocky Hill, NJ) est ajoutée à une concentration finale de 10 ng/ml. Après 6 h d'incubation, l'activité luciférase est mesurée en utilisant le One-Glow Luciferase Assay System (Promega) en suivant les instructions du fabricant.

Pour chaque essai trois puits sont utilisés. Le résultat final est la moyenne des mesures des trois puits. Pour chaque clone d'intérêt, 3 essais indépendants sont réalisés.

## Résultats

L'objectif de cette partie de mon travail de thèse est d'identifier les effecteurs bactériens à l'origine de l'effet anti-inflammatoire de *L. delbrueckii*.

Nous avons sélectionné 8 souches parmi les 57 précédemment testées [219], appartenant aux deux catégories suivantes. D'une part, la catégorie de souches avec un effet anti-inflammatoire élevé et d'autre part la catégorie de souches avec un effet anti-inflammatoire intermédiaire à faible. L'effet inhibiteur de la voie NF- $\kappa$ B de ces 8 souches a été ré-évalué *in vitro* avec les cellules HT-29 à travers 14 expériences indépendantes pour chacune des souches (Figure 1). Les deux souches à effet anti-inflammatoire élevé, CNRZ327 et CNRZ333, appartiennent à la ssp. *lactis*. Les souches à effet anti-inflammatoire intermédiaire appartiennent à la ssp. *lactis* ou à la ssp. *bulgaricus*. Pour une partie des analyses comparatives, CNRZ226, la souche la moins anti-inflammatoire des souches de la ssp. *lactis* et VIB27, la souche la moins anti-inflammatoire des souches de la ssp. *bulgaricus*, ont été sélectionnées pour avoir un représentant de chaque ssp. dans la catégorie des effets immunomodulateurs intermédiaires.



**Figure 1. Effets anti-inflammatoires des 8 souches sélectionnées de *L. delbrueckii*.** Le niveau d'activation de NF- $\kappa$ B est mesuré avec les cellules HT-29/NF- $\kappa$ B-luc-E en présence de 10 ng/ml de la cytokine pro-inflammatoire TNF- $\alpha$ . Barres noires, *L. delbrueckii* ssp. *lactis*; barres grises, *L. delbrueckii* ssp. *bulgaricus*; barres blanches, contrôles sans ajout de bactéries. Moyennes et écarts types de 14 expériences indépendantes.

Les séquences complètes des génomes des souches ATCC11842 et ATCCBAA365 sont présents dans les bases de données publiques, les 6 autres génomes ont été séquencés pour cette étude (chapitre 1 et 2 des résultats).

Pour l'identification des effecteurs bactériens à l'origine de l'effet immuno-modulateur fort des souches CNRZ327 et CNRZ333, deux hypothèses alternatives ont été posées : une première hypothèse, qualitative, où l'effet serait dû à la présence de gènes codants pour des protéines exposées à la surface spécifiques des deux souches les plus anti-inflammatoires et absents des autres souches, et une deuxième hypothèse, quantitative, où l'effet dépendrait du taux d'expression de gènes présents aussi dans d'autres souches que les souches CNRZ327 et CNRZ333. La première hypothèse grâce aux approches de génomique comparative (présence ou absence de gènes spécifiques dans les différentes souches et corrélation avec le phénotype anti-inflammatoire), et de protéomique comparative (présence ou absence de protéines spécifiques dans le protéome de surface des différentes souches). La deuxième hypothèse est étudiée à travers l'approche transcriptomique.

Pour finir, les gènes candidats seront clonés, exprimés dans un système hétérologue et testés pour leurs effets sur les cellules épithéliales HT29 dotées du rapporteur luciférase permettant d'évaluer l'expression de NF- $\kappa$ B.

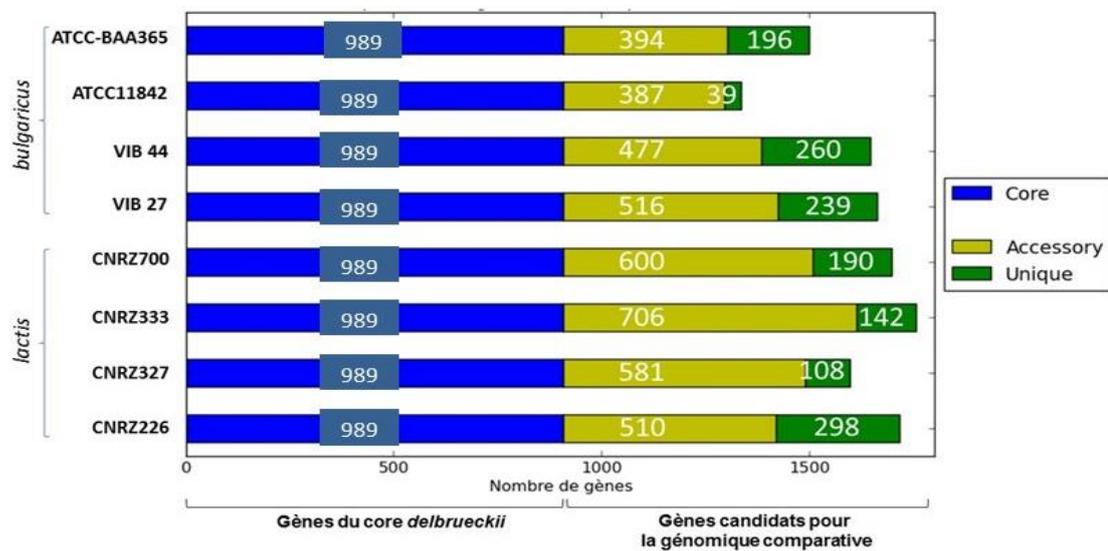
### **Génomique comparative et identification de gènes candidats pour l'effet anti-inflammatoire**

Les séquences complètes de deux génomes (*L. bulgaricus* ATCC 11842 [167] et *L. bulgaricus* ATCC-BAA 365 [218]) utilisées dans cette étude ont été récupérées d'NCBI. Les séquences des 6 autres souches de *L. delbrueckii* ont été établies dans mon travail de thèse (chapitre 1 et 2 des résultats). Selon notre analyse, ces 6 séquences génomiques sont quasiment complètes (mis à part les séquences répétées telles que les IS), permettant la comparaison des protéomes prédits (chapitre 1 et 2 des résultats).

A partir des séquences des 8 génomes de *L. delbrueckii*, les séquences protéiques codées par des non-pseudogènes ont été prédites et comparées par blastP afin de définir des groupes de protéines orthologues par « single linkage clustering » (chapitre 2 des résultats).

Le regroupement des gènes orthologues entre les 8 souches a permis de définir le core génome (les gènes qui codent pour les protéines partagées entre les 8 souches), les gènes accessoires (les gènes qui codent pour les protéines partagées par 2 à 7 souches) ainsi que les gènes uniques (gènes qui codent pour des protéines spécifiques à chacune des souches)

(  
Figure 2).



**Figure 2. Représentation du core génome de *L. delbrueckii* et des gènes dont la présence varie suivant les souches.** Core, les gènes qui codent pour les protéines partagées entre les 8 souches; Accessory, gènes partagés entre 2 souches à 7 souches de *L. delbrueckii*, Unique, gènes qui codent pour des protéines spécifiques à chacune des souches.

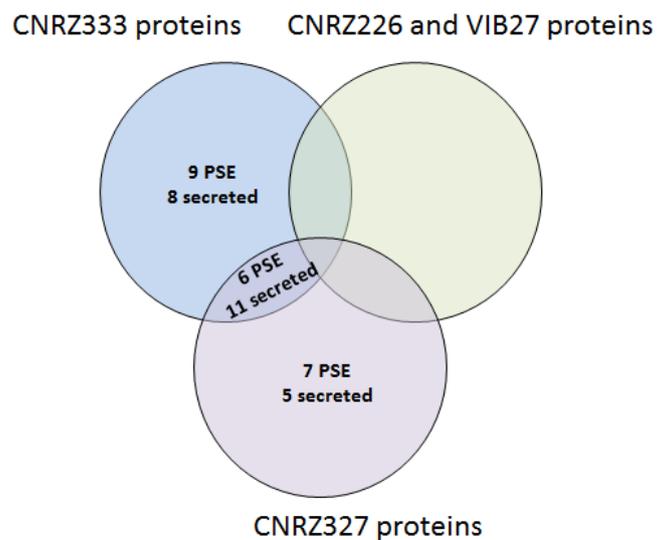
L'important nombre de gènes dans le core de *L. delbrueckii* montre l'importante similarité entre toutes les souches de *L. delbrueckii* (mis à part leur appartenance à une ssp. ou une autre). Pour l'étude comparative ce sont les protéines codées par les gènes accessoires et les gènes uniques qui nous intéressent car ce sont les gènes qui pourraient expliquer les différences entre souches du phénotype immunomodulateur.

Pour la sélection des protéines immunomodulatrices candidates par l'approche de génomique comparative deux critères ont été pris en compte :

1) seuls les gènes qui codent pour des protéines prédites comme exposées à la surface ou secrétées (en utilisant la prédiction de localisation de protéines par SurfG+ [128]) sont pris en considération pour l'analyse. Notons ici que les protéines secrétées sont prises en considération parce qu'elles pourraient être retenues à la surface bactérienne par des mécanismes non reconnus par SurfG+;

2) les protéines présentes chez l'une ou les deux souches les plus anti-inflammatoires (CNRZ327 et CNRZ333) et absentes des souches les moins anti-inflammatoires de la *ssp. lactis* (CNRZ226) et de la *ssp. bulgaricus* (VIB27) ont été sélectionnées comme les meilleures candidates.

Cette approche a permis de sélectionner 46 protéines candidates (Figure 3).



**Figure 3. Diagramme de Venn représentant les résultats de la comparaison génomique entre souches de *L. delbrueckii* ayant différents effets anti-inflammatoires.** Le contenu en protéines des souches de *L. delbrueckii* a été comparé en se limitant aux protéines de surface (PSE, potentially surface exposed) et aux protéines secrétées.

Le tableau 2 regroupe les protéines sélectionnées par l'approche de génomique comparative.

Les protéines sélectionnées sont majoritairement de fonctions inconnues (34 protéines sur les 46 protéines totales). 12 protéines sont partagées entre les deux souches à effet anti-inflammatoire élevé et absentes des 6 souches à effet intermédiaire. 13 protéines sont

spécifiques à l'une ou à l'autre des deux souches les plus anti-inflammatoires et absentes des 6 souches à effet intermédiaire.

Les lignes colorées en rouge représentent des protéines particulièrement intéressantes. Ces 2 protéines différentes, contiennent un (pour CNRZ327) ou deux (pour CNRZ333) domaines décrits comme «mucus binding protein domain» (MucBP domain, PFAM no.06458). Ces protéines sont prédites comme exposées à la surface bactérienne. Même si les fonctions de ces protéines restent inconnues, le fait qu'elles contiennent le domaine MucBP suggère une possible interaction avec l'environnement de la bactérie.

**Tableau 2. Protéines identifiées par l'approche de génomique comparative.**

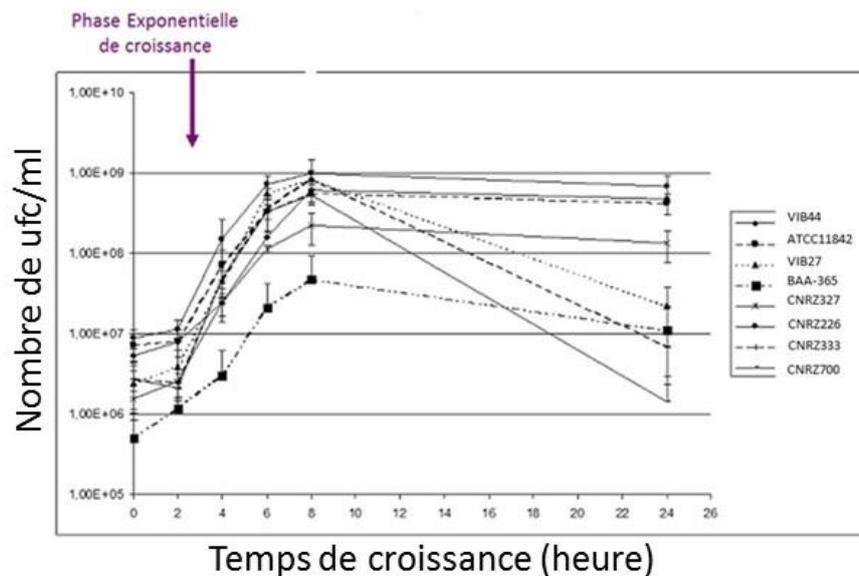
Cluster	Product	Function	Localization SurfG+	Present in CNRZ327	Present in CNRZ333	Absent from the 6 other strains
4793	Shikimate dehydrogenase	biosynthesis of aromatic amino acids	PSE	X	X	X
4976	Putative uncharacterized protein	proteins of unknown function that are similar to other proteins	PSE	X	X	X
3926	Putative uncharacterized protein	Unknown function	SECRETED	X	X	X
4774	No information	Unknown function	SECRETED	X	X	X
4973	Putative uncharacterized protein	Unknown function	SECRETED	X	X	X
3599	Hypothetical membrane protein	Protein of unknown function, without similarity to other proteins	PSE	X		X
3826	Methylated DNA cysteine S-methyltransferase	DNA repair	PSE	X		X
4144	hypothetical protein	Unknown function	PSE	X		X
4377	AbpK sensory transduction histidine kinase	Unknown function	PSE	X		X
4609	Mucus-binding protein, LPXTG-anchored	Unknown function	PSE	X		X
4677	Gram-positive signal peptide protein, YSIRK family	Unknown function	PSE	X		X
3810	No information	Unknown function	SECRETED	X		X
4175	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	X		X
7471	Putative uncharacterized protein	Unknown function	PSE		X	X
7517	ABC-type multidrug transport system,	ATPase activity, coupled to transmembrane movement of substances	PSE		X	X
7579	No information	Unknown function	PSE		X	X
8540	E4SVU4 (Cell surface protein)	Unknown function	PSE		X	X
8714	Hypothetical protein	Protein of unknown function, without similarity to other proteins	PSE		X	X
7851	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED		X	X
7855	Multicopper oxidase family protein	oxidoreductase activity /copper ion binding	SECRETED		X	X
7977	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED		X	X
1975	Putative uncharacterized protein	Unknown function	PSE	X	X	

4197	Putative Ser/Thr protein kinase	3.8 Protein modification	PSE	X	X
4246	AbpK sensory transduction histidine kinase	Unknown function	PSE	X	X
		ATPase activity, coupled to transmembrane movement of			
5085	ABC superfamily ATP binding cassette transporter	substances	PSE	X	X
1475	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	X	X
3593	Putative uncharacterized protein	Unknown function	SECRETED	X	X
3927	Putative uncharacterized protein	Unknown function	SECRETED	X	X
3947	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	X	X
4183	Putative uncharacterized protein	Unknown function	SECRETED	X	X
4222	Putative uncharacterized protein ORF00043	Unknown function	SECRETED	X	X
4268	Sugar-binding periplasmic protein	Carbohydrate transport and metabolism	SECRETED	X	X
5086	No information	Unknown function	SECRETED	X	X
4159	uncharacterized protein	Protein of unknown function, without similarity to other proteins	PSE	X	
1520	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	X	
3174	Putative uncharacterized protein	Unknown function	SECRETED	X	
4541	Putative uncharacterized protein	Unknown function	SECRETED	X	
7087	Putative uncharacterized protein	Unknown function	SECRETED		X
7910	hypothetical protein	Protein of unknown function, without similarity to other proteins	PSE		X
7911	Putative uncharacterized protein	Protein of unknown function, without similarity to other proteins	PSE		X
8007	Brp/Blh family beta-carotene 15,15 -monooxygenase	Unknown function	PSE		X
8690	Mucus binding protein	Unknown function	PSE		X
246	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED		X
7876	hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED		X
7909	No information	Unknown function	SECRETED		X
8017	Surface antigen repeat protein	Unknown function	SECRETED		X

Cluster, cluster de protéines homologues auquel appartient la protéine ; localisation, prédiction de la localisation cellulaire (PSE, potentiellement exposée à la surface; Secreted, sécrétée). Les deux « mucus binding proteins » identifiées sont indiquées en rouge.

## Protéomique comparative et identification des protéines candidates pour l'effet anti-inflammatoire

La composition du protéome de surface a été étudiée pour 7 souches (n'incluant pas la souche CNRZ700). Cette analyse a été faite à partir de cultures bactériennes en phase exponentielle de croissance (figure 4) puisque c'est dans cet état physiologique que l'effet anti-inflammatoire a été décrit *in vitro* [219].

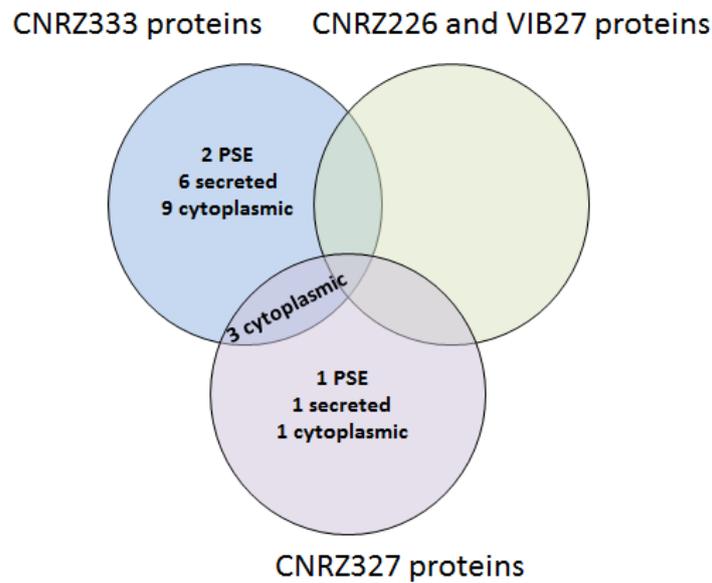


**Figure 4. Croissance de *L. delbrueckii* en milieu MRS.** La flèche indique le moment du prélèvement pour l'étude protéomique.

Pour l'étude du protéome, deux méthodes ont été utilisées : l'extraction de protéines de surface par traitement au chlorure de lithium et la digestion par la trypsine des protéines exposées à la surface. En partant des protéomes de surface et comme par la génomique comparative, nous recherchons une (des) protéine(s) présente(s) chez l'une ou les deux souches les plus anti-inflammatoires (CNRZ327 et CNRZ333) qui soit(en)t absente(s) des souches des ssp. *lactis* et *bulgaricus* les moins anti-inflammatoires (CNRZ226 et VIB27).

Au total 23 protéines candidates ont été identifiées par cette approche comparative (Figure 5). La prédiction *in silico* de la localisation de ces protéines révèle que seulement 10 protéines seraient exposées à la surface ou sécrétées. Les 13 restantes sont prédites comme

cytoplasmiques (Figure 5). Sur ces 13 protéines prédites comme étant cytoplasmiques, 9 sont décrites comme « moonlighting » protéines dans la littérature (Tableau 3).



**Figure 5. Diagramme de Venn représentant les résultats de la comparaison des protéomes de surface des souches de *L. delbrueckii* présentant différents effets anti-inflammatoires. PSE, protéines prédites comme exposées à la surface ; Secreted, protéines prédites comme sécrétées ; Cytoplasmic, protéines prédites comme cytoplasmiques.**

Même si nous ignorons comment ces protéines arrivent dans le protéome de surface, elles pourraient jouer un rôle dans le milieu extracellulaire en plus de leurs fonctions cytoplasmiques.

Soulignons ici que 5 protéines ont été identifiées comme candidates par les approches de génomique et protéomique comparatives (Tableau 3). Parmi ces 5 candidates, nous noterons la présence de la « mucus binding protein » identifiée chez CNRZ333 uniquement.

**Tableau 3. Protéines identifiées par l'approche de protéomique comparative.**

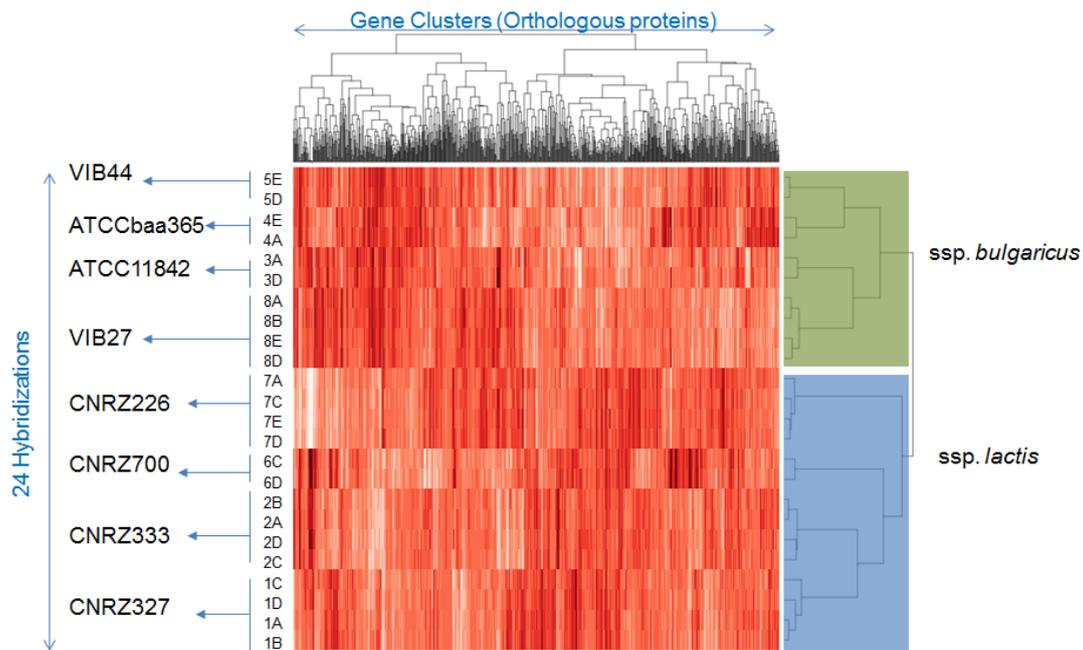
Cluster	Product	Function	Localization surfG+	protein already described as moonlighting	present in CNRZ327	present in CNRZ333	absent from the 6 other strains	Genes identified by Genomic approach	Selected genes for cloning step
cluster_603	Elongation factor Tu	Translation elongation	CYTOPLASMIC	[256]	X	X			X
cluster_114	Hypothetical protein	unknown function	PSE		X		X		X
cluster_8017	surface antigen repeat protein	No information	SECRETED			X	X	X	
cluster_6584	Glycerophosphoryl diester phosphodiesterase	unknown function	SECRETED			X	X		X
cluster_3715	Putative amidase	Cell wall	SECRETED			X	X		X
cluster_4677	Gram-positive signal peptide protein, YSIRK family	No information	PSE			X	X	X	
cluster_291	30S ribosomal protein S3	ribosomal protein	CYTOPLASMIC	[256]		X	X		
cluster_1289	PTS system, mannose-specific enzyme IIAB component	Transport/binding of carbohydrates	CYTOPLASMIC			X	X		
cluster_1832	hypothetical protein	unknown function	SECRETED			X	X		X
cluster_853	Proteinase precursor	Protein degradation	SECRETED			X	X		X
cluster_7909	No information	No information	SECRETED			X	X	X	
cluster_1037	50S ribosomal protein L27	Ribosomal proteins	CYTOPLASMIC	[256]		X	X		
cluster_4183	putative uncharacterized protein	unknown function	SECRETED		X			X	
cluster_8690	Mucus binding protein	No information	PSE			X	X	X	
cluster_924	Enolase	Main glycolytic pathways	CYTOPLASMIC	[257]	X	X			X
cluster_476	Glyceraldehyde 3-phosphate dehydrogenase	Main glycolytic pathways	CYTOPLASMIC	[257]	X	X			X
cluster_1204	50S Ribosomal protein L1	Ribosomal proteins	CYTOPLASMIC	[256]		X	X		
cluster_604	Trigger factor	Protein folding	CYTOPLASMIC			X			
cluster_1166	60 kDa chaperonin GroEL	Adaptation to atypical conditions	CYTOPLASMIC	[258]		X			X
cluster_287	50S ribosomal protein L23	Ribosomal proteins	CYTOPLASMIC	[256]	X				
cluster_1198	50S ribosomal protein L7/L12	Ribosomal proteins	CYTOPLASMIC	[256]		X			
cluster_398	Phosphoketolase	Main glycolytic pathways	CYTOPLASMIC			X			
cluster_167	Phosphoglycerate mutase	Main glycolytic pathways	CYTOPLASMIC			X			

Cluster\_number, cluster de protéines homologues auquel appartient la protéine ; localisation, prédiction de la localisation cellulaire (PSE, potentiellement exposée à la surface ; secreted, sécrétée ; cytoplasmic, cytoplasmique).

## **Transcriptomique comparative et identification de gènes candidats pour l'effet anti-inflammatoire**

Les différences d'effets anti-inflammatoires entre souches pourraient être dues non seulement à la présence ou l'absence de gènes spécifiques mais aussi, à une différence d'expression de gènes présents aussi bien dans les souches peu anti-inflammatoires que dans les souches fortement anti-inflammatoires. Pour cette raison, nous avons réalisé une étude transcriptomique visant à identifier les gènes dont le profil d'expression dans les 8 souches serait corrélé au phénotype anti-inflammatoire. Des hybridations sur puce à ADN avec l'ARN issu de 4 cultures indépendantes ont été réalisées pour les deux souches les plus anti-inflammatoires (CNRZ333 et CNRZ327) et la souche la moins anti-inflammatoire de chaque ssp. étudiée (VIB27 et CNRZ226). Pour les autres souches, deux hybridations sur puce à ADN ont été réalisées. Chacune des cultures ayant servi à l'extraction d'ARN a été testée pour son effet immuno-modulateur *in vitro* afin de confirmer qu'elle présentait le niveau d'immuno-modulation attendu.

Avant d'entamer l'étape d'identification des profils d'expressions corrélés au phénotype anti-inflammatoire, nous avons voulu avoir une vision globale du profil d'expression des gènes des 8 souches de *L. delbrueckii*. Pour cela une représentation graphique sous forme de « heatmap » a été réalisée pour les gènes appartenant au core génome des 8 souches de *L. delbrueckii* (Figure 6).



**Figure 6. Vue globale des profils d'expression des gènes du core génome de *L. delbrueckii*.** Sur l'axe horizontal, les gènes sont groupés selon la ressemblance de leurs profils d'expression dans les 8 souches. Sur l'axe vertical, les échantillons (chaque ligne représentant le résultat d'hybridation pour l'ARN d'une culture d'une souche) sont groupés selon leur ressemblance. L'intensité de la coloration du blanc vers le rouge foncé représente un taux d'expression croissant.

Cette « heatmap » montre une forte similarité entre les échantillons issus de cultures indépendantes de la même souche, ce qui indique une bonne reproductibilité des expériences. Au-delà de ce premier niveau de ressemblance d'échantillons, la « heatmap » montre une séparation claire entre les ssp. *bulgaricus* et *lactis* : les résultats pour les souches appartenant à la même ssp. se ressemblent plus que les résultats pour des souches appartenant à des ssp. différentes. De même, dans chacune des ssp., les résultats pour les souches les plus proches phylogénétiquement (chapitre 2 et 4 résultats) se ressemblent plus que les résultats des autres souches de cette même ssp..

### **Analyse des corrélations entre profil d'expression et phénotype anti-inflammatoire**

Dans un premier temps, nous avons utilisé un modèle de régression linéaire mixte pour estimer l'effet du niveau d'expression d'un gène sur l'inhibition de la voie NF-κB. Ce modèle peut être représenté par la formule :  $Y_{sr} = (a + Ar) X_{sg} + (b + Br + Bs) + \epsilon_{sr}$

où

- **Ysr** est le niveau d'inhibition de la voie NF- $\kappa$ B par la souche s pour le réplicat r ;
- **Xsg** est le niveau d'expression du gène g dans la souche s (exprimé en échelle logarithmique (base2)) ;
- **a** est la réponse de l'inhibition de la voie NF- $\kappa$ B à des changements d'expression du gène ;
- **Ar** est l'effet du réplicat sur la réponse ;
- **b** est le niveau de base de l'inhibition de la voie NF- $\kappa$ B (en l'absence totale d'expression du gène) ;
- **Br** est l'effet du réplicat sur le niveau de base ;
- **Bs** est l'effet de la souche sur le niveau de base ;
- **Esr** est l'erreur résiduelle (erreurs de mesures, effets non pris en compte dans le modèle).

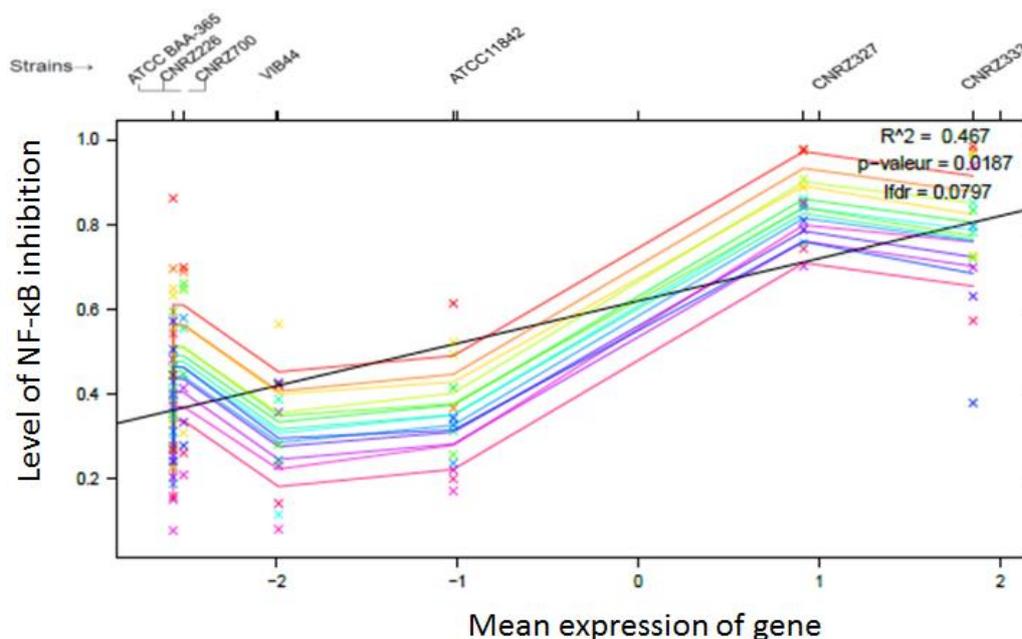
La corrélation a été calculée en utilisant les données sur l'inhibition de la voie NF- $\kappa$ B des 14 tests indépendants mentionnés plus haut (incluant ceux faits en parallèle de l'extraction d'ARN pour l'étude transcriptomique) (Figure 1). Une valeur d'expression par gène et par souche a été utilisée (représentant la moyenne de 2 ou 4 mesures indépendantes, selon les souches, comme décrit précédemment). Lorsque les gènes sont absents, une valeur par défaut équivalent au bruit de fond est assignée façon à pouvoir les prendre en compte dans l'analyse mathématique.

Comme le montre l'exemple de la figure 7, pour chaque gène une régression linéaire mixte est calculée (droite noire sur la Fig.) avec :

- un coefficient transcriptomique qui correspond à la pente de la droite de régression et qui indique la relation entre l'inhibition de NF- $\kappa$ B et le niveau de transcription ;
- un « local false discovery rate » [259] appelé fdr local indique la probabilité qu'il n'y ait pas d'association entre le niveau d'expression du gène étudié et l'inhibition de la voie NF- $\kappa$ B ;

- une fraction de variance,  $R^2$  qui renseigne sur la fraction de la variabilité de l'inhibition de la voie NF- $\kappa$ B qui peut être expliquée par le modèle ; sur l'exemple Fig. 7, environ 47% de la variabilité des mesures pour chaque souche est expliquée par le modèle.

Une valeur-p et surtout un fdr local faibles traduisent une association entre niveau d'expression et inhibition de la voie NF- $\kappa$ B incompatibles avec des fluctuations aléatoires.



**Figure 7. Corrélation entre expression et niveau d'inhibition de la voie NF- $\kappa$ B.** Exemple pour un cluster de gènes homologues, en utilisant un modèle linéaire mixte. Voir le texte pour les explications. Coeff transcriptomique : la pente de la droite de régression linéaire mixte; pval : la probabilité d'obtenir le même résultat si l'hypothèse nulle (pas d'association entre le taux d'expression du gène et l'inhibition de la voie NF- $\kappa$ B) était vraie ; lfd : p-valeur corrigée pour la multiplicité des tests ;  $R^2$  : fraction de la variance expliquée par le modèle. Le niveau d'expression du gène (axe X) est exprimé en échelle logarithmique (base 2). La droite noire représente la relation entre le niveau d'inhibition de NF- $\kappa$ B et le niveau d'expression du gène étudié selon le modèle. Les croix indiquent les mesures réelles (données expérimentales). Les lignes colorées indiquent le niveau d'inhibition de NF- $\kappa$ B prédit par le modèle pour un niveau d'expression donné, tenant compte des effets Ar, Br, et Bs (voir le texte pour les explications), où chaque couleur représente une expérience.

Pour qu'un cluster de gènes soit retenu pour effectuer cette analyse des résultats de transcriptomique (c'est-à-dire la corrélation entre le taux d'expression du gène et l'inhibition de NF- $\kappa$ B), il faut qu'il réponde aux critères suivant :

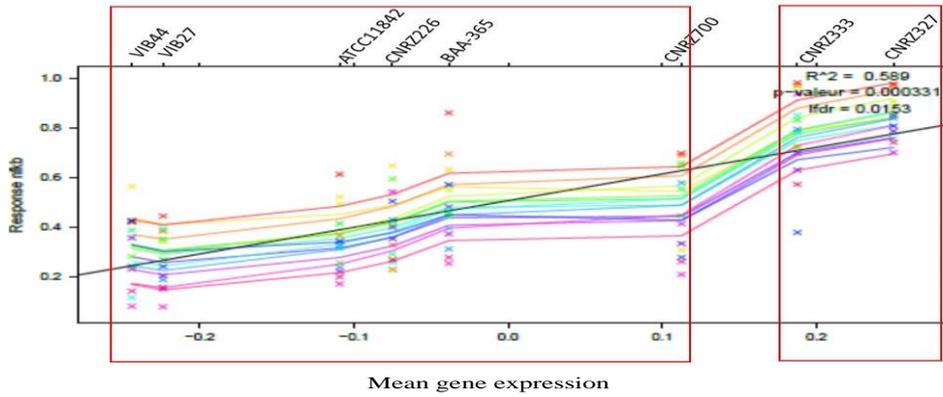
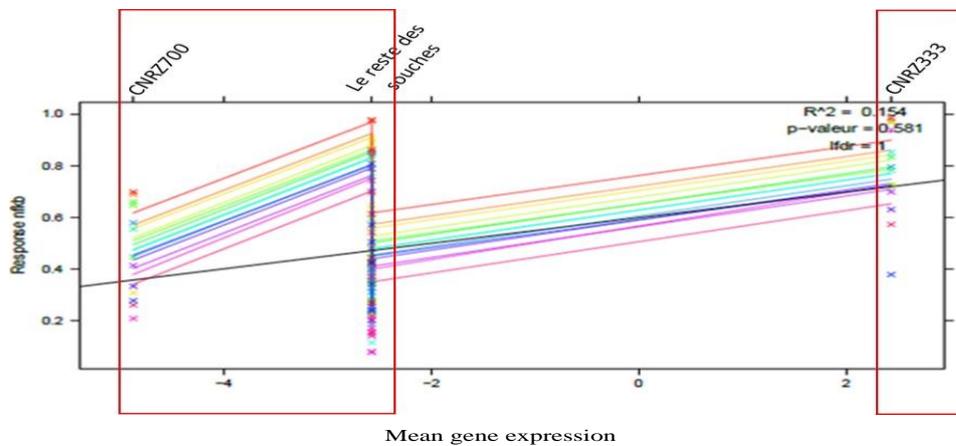
- Le cluster doit comporter au minimum deux gènes appartenant à deux souches différentes (dans une analyse relative l'étape de normalisation des résultats nécessite

l'existence d'au minimum deux conditions et par conséquent d'au minimum deux profils d'expression).

- Le R<sup>2</sup> des analyses de transcriptomique doit être supérieur à 0,1 (ou 10%).
- La valeur-p des analyses de transcriptomique doit être inférieure ou égale à 0,01.
- Le fdr local (lfdr) des analyses de transcriptomique doit être inférieur ou égal à 0,01.

En utilisant cette méthode plusieurs cluster de gènes répondaient bien aux critères statistiques de sélection choisis ( $R^2 > 0,1$  fdr local =  $< 0,01$ ) ; c'est-à-dire qu'ils présentaient une bonne corrélation entre le niveau d'expression et l'inhibition de NF- $\kappa$ B. Cependant certains clusters montrent une différence d'expression assez faible entre les souches les plus anti-inflammatoires (CNRZ327 et/ou CNRZ333) et les autres souches. Un exemple est présenté dans la figure 8A où cette différence d'expression est inférieure à 1,07 (2 puissance 0,1), c'est-à-dire en dessous de ce qui est habituellement considérée comme une différence d'expression biologiquement significative (généralement au moins un facteur 2).

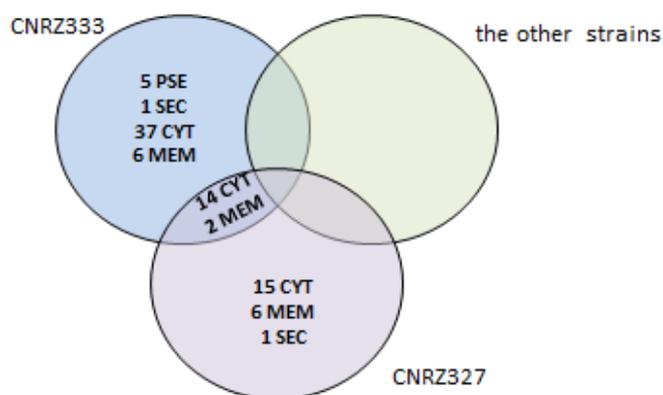
Dans ce cas, la question de la signification biologique de cette faible différence se pose. Nous avons également identifié des clusters de gènes où la différence d'expression entre les souches les plus anti-inflammatoires et celles présentant un effet intermédiaire, est très importante mais où la corrélation entre expression et effet sur NF- $\kappa$ B ne passe pas les critères statistiques (Figure 8B).

**A****B**

**Figure 8. Exemples de clusters.** A: Exemple du cluster de la « Glutamine-binding protein », avec une très bonne corrélation entre le profil d'expression et l'effet inhibiteur mais une faible différence d'expression du gène entre souches fortement et faiblement anti-inflammatoires. B : exemple du cluster d'une « mucus binding protein », avec une importante différence d'expression entre une des souches anti-inflammatoire et les autres souches, mais avec des contrôles statistiques très mauvais, notamment dû au fait que le gène ne soit présent que dans deux souches ( $R^2$ , lfr).

Suite à ces observations, nous avons décidé de sélectionner des gènes candidats sur la base de la différence d'expression entre les souches les plus anti-inflammatoires (CNRZ327 et/ou CNRZ333) et les autres souches. Nous nous sommes basés sur un coefficient d'expression différentielle (DE-coef) qui mesure pour chaque gène la plus faible différence d'expression entre une souche fortement anti-inflammatoire et une souche peu anti-inflammatoire. Les gènes candidats sont ceux dont le coefficient, exprimé en  $\log_2$ , est supérieur ou égal à 1 ce qui correspond à une différence de niveau d'expression d'au moins un facteur 2 entre les souches très et peu immuno-modulatrices.

Cette étape de sélection conduit à une liste de gènes des deux souches les plus anti-inflammatoires, CNRZ327 et CNRZ333, surexprimés ou sous-exprimés par rapport aux autres souches. La figure 9 illustre la distribution des gènes surexprimés chez CNRZ327, CNRZ333 ou les deux, et sélectionnés par l'approche de transcriptomique comparative.



**Figure 9. Diagramme de Venn représentant les résultats de la stratégie de transcriptomique comparative entre souches de *L. delbrueckii* ayant différents effets anti-inflammatoires.** Gènes surexprimés chez les souches les plus anti-inflammatoires par rapport aux autres souches. La localisation cellulaire prédite des protéines codées par les gènes sélectionnés est indiquée comme suit : PSE, potentiellement exposée à la surface ; SEC, sécrétée ; CYT, cytoplasmique ; MEM, membranaire.

Pour la liste des gènes candidats à cloner nous nous sommes limités aux gènes surexprimés dans CNRZ327 et/ou CNRZ333 et dont les produits sont prédits comme étant sécrétés ou potentiellement exposés à la surface bactérienne (tableau 4). Cette approche conduit à sélectionner un seul 1 gène en plus de ceux qui avaient été retenus par les analyses de génomique et protéomique comparatives.

**Tableau 4. Gènes sélectionnés par l'approche de transcriptomique comparative.**

cluster_number	protein ID	localization	product	function	327 transcr	333 transcr	genomics	proteomics
cluster_8690	CNRZ333#1409	PSE	Mucus binding protein	No information		x	x	x
cluster_7909	CNRZ333#150	SEC	No information	No information		x	x	x
cluster_1520	CNRZ327#193	SEC	Hypothetical protein	Protein of unknown function, without similarity to other proteins	x		x	
cluster_7910	CNRZ333#152	PSE	hypothetical protein	Protein of unknown function, without similarity to other proteins		x	x	
cluster_7911	CNRZ333#153	PSE	Putative uncharacterized protein	Protein of unknown function, without similarity to other proteins		x	x	
cluster_8007	CNRZ333#2106	PSE	Brp/Blh family beta-carotene 15,15'-monooxygenase	No information		x	x	
cluster_44	CNRZ333#2018	PSE	Glycosyl transferase	No information		x		

Cluster\_number, numéro du cluster de protéines homologues auquel appartient la protéine d'intérêt ; localization, prédiction de la localisation cellulaire (PSE, potentiellement exposée à la surface et SEC, sécrétée) ; le gène sélectionné par l'approche de transcriptomique comparée est en rouge.

## **Validation de l'effet immuno-modulateur des gènes candidats par expression hétérologue et test *in vitro***

Les approches de génomique, protéomique et transcriptomique comparatives ont permis de sélectionner un ensemble de gènes codants des protéines candidates pour le phénotype immuno-modulateur de CNRZ327 et CNRZ333.

L'approche de génomique comparative a permis d'identifier 46 gènes candidats. L'approche de protéomique comparative a permis d'identifier 23 gènes candidats. Parmi ces 23 gènes :

- 5 faisaient déjà partie des 46 candidats identifiés par génomique comparée ;
- 9 sont éliminés car ils codent pour des protéines prédites comme cytoplasmiques qui ne sont pas décrites comme « moonlighting » protéines.

Ainsi au final, 9 gènes candidats supplémentaires sont retenus pour clonage.

Pour la liste des gènes candidats identifiés initialement par transcriptomique comparative, nous nous sommes limités aux gènes surexprimés dans CNRZ327 et/ou CNRZ333 dont les produits sont prédits comme exposés à la surface ou sécrétés. Avec cette approche nous avons sélectionné 1 gène supplémentaire qui n'avait pas été identifié par les stratégies de génomique et protéomique comparatives.

Au total, 56 gènes de CNRZ333 et CNRZ327 ont été sélectionnés pour l'étape de clonage et de test *in vitro* pour confirmation de l'effet immuno-modulateur.

Parmi les 56 gènes sélectionnés, 41 gènes ont été clonés dans le système hétérologue *Bacillus subtilis*. A ce jour, 17 ont été testés *in vitro* pour leurs effets sur des cellules HT29 contenant un rapporteur reflétant le niveau d'expression de NF- $\kappa$ B (Tableau 5).

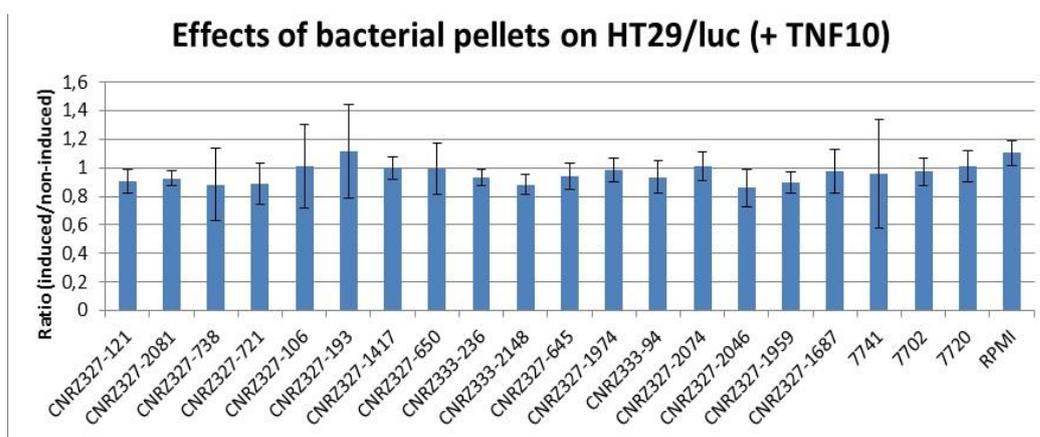
**Tableau 5. Liste des gènes sélectionnés pour l'étape de clonage dans *Bacillus subtilis*.**

Cluster	Protein ID	Product	Function	Localisation	Identified by genomic	Identified by proteomic	Identified by transcriptomic
cluster_8690	CNRZ333#1409	Mucus binding protein	No information	PSE	x	x	x
cluster_7909	CNRZ333#150	No information	No information	SECRETED	x	x	x
cluster_4677	CNRZ327#640	Gram-positive signal peptide protein, YSIRK family	No information	PSE	x	x	
cluster_4183	CNRZ327#12	Putative uncharacterized protein	proteins of unknown function that are similar to others proteins	SECRETED	x	x	
cluster_8017	CNRZ333#236	Surface antigen repeat protein	No information	SECRETED	x	x	
cluster_7910	CNRZ333#152	hypothetical protein	Protein of unknown function, without similarity to other proteins	PSE	x		x
cluster_7911	CNRZ333#153	Putative uncharacterized protein	Protein of unknown function, without similarity to other proteins	PSE	x		x
cluster_8007	CNRZ333#2106	Brp/Blh family beta-carotene 15,15'-monooxygenase	No information	PSE	x		x
cluster_1520	CNRZ327#193	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	x		x
cluster_4774	CNRZ327#1687	No information	No information	SECRETED	x		x
cluster_4973	CNRZ333#101	Putative uncharacterized protein	proteins of unknown function that are similar to others proteins	SECRETED	x		x
cluster_1975	CNRZ327#1296	Putative uncharacterized protein	No information	PSE	x		
cluster_3599	CNRZ327#1492	Hypothetical membrane protein	Protein of unknown function, without similarity to other proteins	PSE	x		
cluster_3826	CNRZ327#738	Methylated-DNA-[protein]-cysteine S-methyltransferase	DNA repair	PSE	x		
cluster_4144	CNRZ327#1959	hypothetical protein	No information	PSE	x		
cluster_4159	CNRZ327#955	uncharacterized protein	Protein of unknown function, without similarity to other proteins	PSE	x		
cluster_4197	CNRZ327#1967	Putative Ser/Thr protein kinase	Protein modification	PSE	x		
cluster_4246	CNRZ327#1974	AbpK sensory transduction histidine kinase	No information	PSE	x		
cluster_4377	CNRZ327#122	AbpK sensory transduction histidine kinase	No information	PSE	x		
cluster_4609	CNRZ327#432	Mucus-binding protein, LPXTG-anchored	Proteins of unknown function that are similar to others proteins	PSE	x		
cluster_4793	CNRZ327#1855	Shikimate dehydrogenase	Biosynthesis of aromatic amino acids	PSE	x		
cluster_4976	CNRZ327#2075	Putative uncharacterized protein	Proteins of unknown function that are similar to other proteins	PSE	x		
cluster_5065	CNRZ327#2081	ABC superfamily ATP binding cassette transporter	ATPase activity, coupled to transmembrane movement of substances	PSE	x		
cluster_7471	CNRZ333#1542	Putative uncharacterized protein	Proteins of unknown function that are similar to others proteins	PSE	x		
cluster_7517	CNRZ333#2045	ABC-type multidrug transport system, ATPase and permease component	ATPase activity, coupled to transmembrane movement of substances	PSE	x		
cluster_7579	CNRZ333#2060	No information	No information	PSE	x		
cluster_8540	CNRZ333#2148	E4SVU4 (Cell surface protein)	No information	PSE	x		
cluster_8714	CNRZ333#374	No information	No information	PSE	x		
cluster_1475	CNRZ327#106	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	x		
cluster_246	CNRZ333#1829	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	x		
cluster_3174	CNRZ327#121	Putative uncharacterized protein	proteins of unknown function that are similar to others proteins	SECRETED	x		
cluster_3593	CNRZ327#1485	Putative uncharacterized protein	No information	SECRETED	x		
cluster_3810	CNRZ327#721	No information	No information	SECRETED	x		
cluster_3926	CNRZ327#935	Putative uncharacterized protein	proteins of unknown function that are similar to others proteins	SECRETED	x		
cluster_3927	CNRZ327#936	Putative uncharacterized protein	Proteins of unknown function that are similar to others proteins	SECRETED	x		
cluster_3947	CNRZ327#1017	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	x		
cluster_4175	CNRZ327#1873	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	x		
cluster_4222	CNRZ327#56	Putative uncharacterized protein	Proteins of unknown function that are similar to others proteins	SECRETED	x		
cluster_4268	CNRZ327#1063	Sugar-binding periplasmic protein	Carbohydrate transport and metabolism	SECRETED	x		
cluster_4541	CNRZ327#645	Putative uncharacterized protein	Proteins of unknown function that are similar to others proteins	SECRETED	x		
cluster_5086	CNRZ327#2046	No information	No information	SECRETED	x		
cluster_7087	CNRZ333#1904	Putative uncharacterized protein	proteins of unknown function that are similar to others proteins	SECRETED	x		
cluster_7851	CNRZ333#90	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	x		
cluster_7855	CNRZ333#94	Multicopper oxidase family protein	Oxidoreductase activity /copper ion binding	SECRETED	x		
cluster_7876	CNRZ333#116	hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	x		
cluster_7977	CNRZ333#168	Hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED	x		
cluster_603	CNRZ327#1162	Elongation factor Tu	Translation elongation	CYTOPLASMIC		x	
cluster_924	CNRZ333#771	enolase	Main glycolytic pathways	CYTOPLASMIC		x	
cluster_1166	CNRZ333#643	60 kDa chaperonin GroEL	Adaptation to atypical conditions	CYTOPLASMIC		x	
cluster_476	CNRZ333#1353	Glyceraldehyde 3-phosphate dehydrogenase	Main glycolytic pathways	CYTOPLASMIC		x	
cluster_114	CNRZ327#1839	Hypothetical protein	Protein of unknown function, without similarity to other proteins	PSE	x		
cluster_1832	CNRZ327#1517	hypothetical protein	Protein of unknown function, without similarity to other proteins	SECRETED		x	
cluster_3715	CNRZ333#1713	Putative amidase	Cell wall	SECRETED		x	
cluster_6584	CNRZ333#233	Glycerophosphoryl diester phosphodiesterase	Protein of unknown function, without similarity to other proteins	SECRETED		x	
cluster_853	CNRZ333#882	Proteinase precursor  prtB	Protein degradation	SECRETED		x	
cluster_44	CNRZ333#2018	Glycosyl transferase	Peptidoglycan biosynthesis	PSE			x

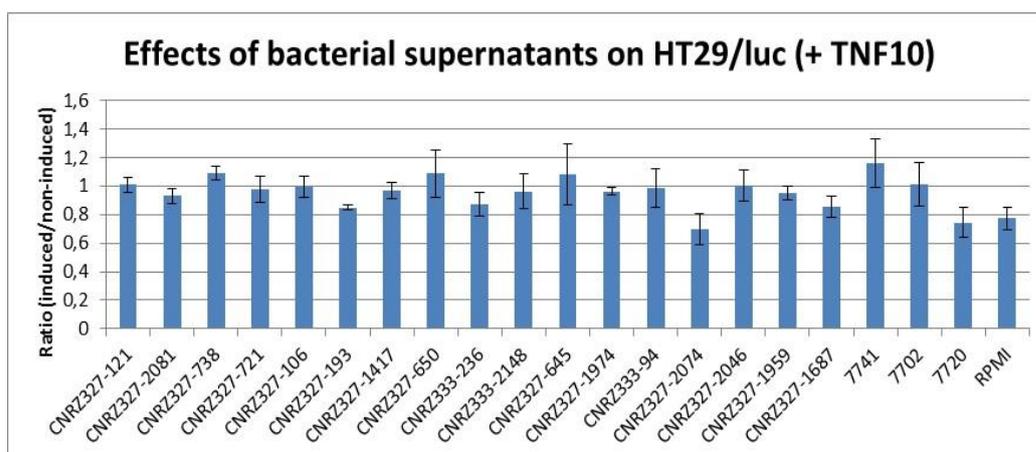
Cluster, numéro du cluster des protéines homologues auquel appartient la protéine d'intérêt ; protein ID, numéro identifiant de la protéine de CNRZ327 ou CNRZ333; localisation, prédiction de la localisation cellulaire (PSE, potentiellement exposée à la surface, SECRETED, sécrétée et CYTOPLASMIC, cytoplasmique); Identified by genomic/proteomic/transcriptomic, l'approche par laquelle la protéine a été sélectionnée.

L'activité immuno-modulatrice de 17 gènes candidats de *L. delbrueckii* a été testée (Figure 10) en mesurant le niveau d'expression de NF- $\kappa$ B après contact des cellules avec le clone de *B. subtilis* induit ou non-induit et en présence du TNF- $\alpha$ . Les rapports d'expression de NF- $\kappa$ B avec le clone induit sur celui mesuré avec le clone non induit révèlent que le surnageant du clone exprimant le gène CNRZ327#2074 (codant pour une protéine de fonction inconnue) a le plus faible rapport d'activité NF- $\kappa$ B. Ce rapport a une valeur légèrement inférieure à celles obtenues avec les contrôles 7702, 7720 et RPMI seul, mais nettement inférieure à la moyenne des valeurs obtenue pour l'ensemble des autres clones testés.

**A**



**B**



**Tableau 10. Test *in vitro* dans les cellules HT29/luc pour l'étude de l'inhibition de la voie NF- $\kappa$ B par les candidats protéiques bactériens.** A : effet immuno-modulateur du culot bactérien sur les cellules HT29/luc (cellules d'adénocarcinome du côlon humain transfectées avec le gène rapporteur de la luciférase en aval d'un promoteur inducible par le facteur transcriptionnel NF- $\kappa$ B) stimulées avec le TNF- $\alpha$ . B : effet immuno-modulateur du surnageant bactérien sur les cellules HT29/luc stimulées avec le TNF- $\alpha$ . 7741, contrôle avec *Bacillus subtilis* hétérologue exprimant une flagelline inflammatoire d'*E. coli* (induite par l'IPTG) [248]; 7702,

contrôle avec *Bacillus subtilis* contenant le plasmide pDG148 vide ; **7720**, contrôle avec *Bacillus subtilis* natif ; **RPMI**, contrôle négatif avec le milieu de culture RPMI seul. La moyenne du ratio de l'activité du NF- $\kappa$ B après induction de l'expression du gène dans *Bacillus subtilis* à l'IPTG sur l'activité du NF- $\kappa$ B sans induction a été calculée pour chaque gène à partir de 3 expériences indépendantes.

En observant séparément les résultats des trois tests *in vitro* pour chacun des 17 gènes candidats de *L. delbrueckii*, plusieurs culots et surnageants bactériens présentent de faibles rapports « clone induit sur non-induit » compatibles avec une inhibition de NF- $\kappa$ B dans deux expériences sur les trois réalisées (résultats non présentés). C'est le cas pour (i) les surnageants des clones exprimant les gènes CNRZ327#1687 et CNRZ327#193 codants pour des protéines de fonctions inconnues et pour (ii) les culots des clones exprimant les gènes CNRZ327#2046 (fonction inconnue), CNRZ327#738 (« Methylated DNA cysteine S-methyltransferase », impliquée dans la réparation de l'ADN) et CNRZ327#721 (fonction inconnue). Cette observation nous conduira à répéter le criblage fonctionnel pour ces 5 clones à minima.

## Discussion

Les approches de génomique, protéomique et transcriptomique comparatives ont permis de sélectionner un total de 56 gènes candidats pour le phénotype immuno-modulateur des souches CNRZ327 et CNRZ333. La localisation prédite pour les protéines codées par ces gènes correspond majoritairement à des protéines extracellulaires avec 28 sécrétées et 24 potentiellement exposées à la surface et pour une minorité (4) des protéines cytoplasmiques qui sont cependant déjà décrites dans la littérature comme « moonlighting proteins ». La sélection des protéines sécrétées pour l'étape de clonage, malgré le fait qu'on attribue l'effet immuno-modulateur à une protéine exposée à la surface, se justifie en considérant qu'on n'arrive peut-être pas à complètement discriminer entre les protéines exposées à la surface et sécrétées par les outils de prédiction de localisation cellulaire. Parmi les 56 gènes sélectionnés, 41 gènes ont pu être clonés dans le système hétérologue *Bacillus subtilis* dont, pour l'instant, 17 ont été testés *in vitro* pour leurs effets immuno-modulateurs. Malgré le fait

qu'on attendait un effet immuno-modulateur d'une protéine exposée à la surface, le culot mais aussi le surnageant bactériens ont été testés pour chaque clone d'intérêt. En effet, si la machinerie d'ancrage des protéines à la surface bactérienne de *B. subtilis* n'était pas appropriée pour les protéines de *L. delbrueckii*, les protéines d'intérêt pourraient être localisées dans le surnageant. La surproduction des protéines de *L. delbrueckii* pourrait aussi saturer la machinerie d'ancrage de *B. subtilis* ce qui conduirait aussi à une localisation dans le surnageant de culture de toute ou partie de la protéine d'intérêt.

Les résultats préliminaires (moyenne calculée à partir de trois expériences indépendantes) montrant l'activité inhibitrice de l'expression du régulon NF- $\kappa$ B par les 17 gènes candidats de *L. delbrueckii*, révèlent que le surnageant du clone exprimant le gène CNRZ327#2074, codant pour une protéine de fonction inconnue, donne les résultats les plus prometteurs avec le plus faible rapport entre l'activité inhibitrice du clone induit sur l'activité du clone non-induit en présence de la cytokine pro-inflammatoire TNF- $\alpha$ . En plus de ce gène, les surnageants (CNRZ327#1687 et CNRZ327#193) et les culots (CNRZ327#2046 et CNRZ327#738) d'autres clones montrent des rapports d'activité NF- $\kappa$ B induit sur non induit faibles pour deux essais sur trois. Cela montre l'importance de refaire le test d'évaluation de l'effet immuno-modulateur *in vitro* pour pouvoir trancher. Les fonctions de ces gènes sont toutefois inconnues à l'exception de CNRZ327#738 qui code pour une «Methylated DNA cysteine S-methyltransferase» impliquée dans la réparation de l'ADN et prédite comme exposée à la surface cellulaire.

Parmi les autres gènes candidats, certains ont des domaines qui pourraient être compatibles avec une interaction avec l'hôte conduisant à un effet immuno-modulateur. Ainsi, deux gènes codent pour des « mucus binding protein » identifiés spécifiquement dans les souches immuno-modulatrices CNRZ327 (CNRZ327#432) et CNRZ333 (CNRZ333#1409).

Plusieurs gènes ont déjà été décrits comme ayant des fonctions importantes dans l'adhésion de lactobacilles au niveau du tractus digestif [260]. Des protéines de lactobacilles médiant

l'adhésion à la couche de mucus et aux cellules épithéliales ont pu être caractérisées [261]. Les adhésines au mucus des lactobacilles identifiées et fonctionnellement caractérisées incluent la « mucus binding protein » (Mub) extracellulaire de *Lactobacillus reuteri* 1063 [262], ainsi que la protéine Mub de *Lactobacillus acidophilus* NCFM [263]. Ces deux protéines partagent un « mucus binding domain » (MUB, PFAM no.06458) similaire [229]. L'analyse des protéines de surface a montré que le domaine MUB est l'un des domaines protéiques les plus abondants chez les lactobacilles retrouvés dans l'intestin [229], ce qui suggère fortement qu'il pourrait jouer un rôle clef dans l'environnement intestinal. Chez *Streptococcus pneumoniae*, une « mucus binding protein » de 21 kDa a déjà été décrite comme impliquée dans l'adhésion et la colonisation des poumons et des voies respiratoires humains. Le « mucus binding domain » isolé de cette protéine possède *in vitro* une affinité de liaison à la mucine comparable à celle observée pour la forme mature de la protéine (obtenue après modifications post-traductionnelles) [264].

Toutes ces indications vont dans le sens d'un rôle des « mucus binding proteins » identifiées chez CNRZ327 et CNRZ333 (et plus précisément le domaine « mucus binding domain » commun à ces deux protéines), dans l'interaction avec le mucus des cellules épithéliales. Toutefois, cela ne permet pas encore de conclure que ces protéines sont responsables de l'effet anti-inflammatoire des deux souches sur les cellules épithéliales intestinales.

Parmi les protéines sélectionnées par protéomique comparative, certaines comme le facteur d'élongation Tu (CNRZ327#1162) et la protéine GroEL (CNRZ333#643), semblent être des candidats plus prometteurs pour un rôle immuno-modulateur. En effet bien que ces protéines soient prédites comme cytoplasmiques et que les gènes correspondant soient présents dans toutes les souches, elles ont seulement été identifiées dans les protéomes de surface des souches les plus anti-inflammatoires CNRZ333 et CNRZ327.

Le facteur d'élongation Tu (EFTu) est une protéine G (« guanosine nucleotide binding protein ») qui joue un rôle central dans la synthèse des protéines au niveau du cytoplasme. La protéine GroEL est une protéine de choc thermique appartenant à la classe des protéines chaperons Hsp60. En plus des rôles qu'elles jouent dans le cytoplasme bactérien, la protéine EFTu et la protéine GroEL sont toutes deux décrites comme associées à la surface bactérienne chez *Lactobacillus johnsonii* La1, une bactérie ayant des propriétés immuno-modulatrices et pouvant adhérer à la fois aux cellules épithéliales intestinales et au mucus [265, 266].

Etudiées chez *Lactobacillus johnsonii* La1, ces deux protéines se sont avérées pro-inflammatoires [265, 266]. Bien que ces deux protéines ne soient actuellement décrites que pour des effets inflammatoires, il reste intéressant d'étudier leur potentiel immuno-modulateur également pour *Lactobacillus delbrueckii*. Les 4 gènes candidats correspondant à ces protéines (CNRZ327#432, CNRZ333#1409, CNRZ333#643 et CNRZ327#1162) sont encore en cours de clonage.

## **V.D *Lactobacillus delbrueckii* ssp. *bulgaricus* and ssp. *lactis*: two subspecies, two evolution histories.**

### Introduction

Une des caractéristiques les plus intrigantes observée chez *L. bulgaricus* ATCC11842 [167] est la présence d'une séquence inversée répétée de 47,5 Kb dans la région du terminus de réplication. Cette répétition inclut le site *dif* impliqué dans la résolution du chromosome sous sa forme dimérique en fin de réplication. Par conséquent, *L. bulgaricus* contient deux sites *dif* d'orientations opposées à une distance de 2,5 Kb. Ces observations dans le génome de *L. bulgaricus* ATCC 11842 nous ont incité à rechercher une telle structure dans d'autres génomes de *L. bulgaricus* ainsi que dans les génomes de souches de la ssp. *lactis*. Le but ultime étant de savoir si ce genre de structure peut nous renseigner sur l'histoire évolutive des deux ssp. . Pour ce faire l'analyse de séquence de la région du terminus a été faite dans tous les génomes complets de *L. delbrueckii* présents dans les bases de données ainsi que dans les génomes que nous venons de séquencer. Pour finir, nous avons étendu la recherche de structure répétée dans la région terminus à un nombre élevé de génomes complets d'autres espèces bactériennes afin de déterminer la fréquence de ce type de structure.

***Lactobacillus delbrueckii* ssp. *bulgaricus* and ssp. *lactis*: two subspecies, two evolution histories.**

Hela El Kafsi<sup>1,2</sup>, Mahendra Mariadassou<sup>3</sup>, Camille Blin<sup>3</sup>, Valentin Loux<sup>3</sup>, H  l  ne Chiapello<sup>3\*</sup>, Anne-Laure Abraham<sup>1,2</sup>, Emmanuelle Maguin<sup>1,2</sup>, and Maarten van de Guchte<sup>1,2,#</sup>

<sup>1</sup>INRA, UMR1319 Micalis, Jouy en Josas, France

<sup>2</sup>AgroParisTech, UMR Micalis, Jouy en Josas, France

<sup>3</sup>INRA, UR1077 Math  matique Informatique et G  nome, Jouy en Josas, France

#Corresponding author

## Summary

The *Lactobacillus delbrueckii* subspecies *lactis* and *bulgaricus* are lactic acid producing bacteria that are largely used in dairy industries, respectively in cheese-making and yogurt production. Both show clear signs of adaptation to the milk environment through the loss of superfluous functions in relation to carbohydrate metabolism and amino acid biosynthesis.

In this work, two new aspects of the genome structure were studied. First, we show that within the common core genome, recombination events contributed more importantly to the diversity of the ssp. *lactis* than to the diversity of the ssp. *bulgaricus*, relative to the contribution of point mutations. Second, the replication terminus region turns out to be a highly variable region of the *L. delbrueckii* genome. In the ssp. *bulgaricus* this region appears to contain an unusual large inverted repeat (IR), the size of which varies from 37 to 76 kbp among the 5 strains studied. These IRs most probably derive from a common ancestral IR of at least 100 kbp. In contrast, no sign of such an IR was found in the 5 ssp. *lactis* strains studied. Nonwithstanding this major difference, the corresponding region in the ssp. *lactis* genomes too appears to be the scene of extensive recombination.

## Introduction

The thermophilic lactic acid producing bacterium *Lactobacillus delbrueckii* is mainly known from its application in dairy fermentations, notably in the production of yogurt (*L. delbrueckii* ssp. *bulgaricus*) and cheese (*L. delbrueckii* ssp. *lactis*). Phylogenetically, *L. delbrueckii* belongs to the "acidophilus group" [1], a group of closely related lactobacilli of which several other members (*L. acidophilus*, *L. johnsonii*, *L. gasseri*) are commensals of the human gastrointestinal (GI) tract with alleged probiotic properties. Although *L. delbrueckii* is generally not considered as a commensal bacterium, the ssp. *bulgaricus* was reported to survive in the human GI tract after its ingestion with yogurt [2, 3]. These features instigated a renewed interest in the species *L. delbrueckii* addressing the question of its probiotic potential, a hypothesis originally put forward for yogurt bacteria (*L. delbrueckii* ssp. *bulgaricus* (*L.*

*bulgaricus*) and *Streptococcus thermophilus*) by Metchnikof [4]. Both *L. delbrueckii* ssp. *bulgaricus* and *L. delbrueckii* ssp. *lactis* were shown to down-regulate the activation of NF- $\kappa$ B, a key factor in the establishment of inflammatory responses, in human gut epithelial cells *in vitro* [5]. A selected strain of the ssp. *lactis* also improved the symptoms of DSS (dextran sodium sulfate)-induced colitis in mice, a model of human ulcerative colitis [5, 6].

From an evolutionary viewpoint, the first complete genome sequence of *L. bulgaricus* revealed several interesting features [7]. A relatively high GC content at third codon positions and the presence of an exceptionally high number of pseudogenes and incomplete metabolic pathways indicated that the *L. bulgaricus* genome is in a state of ongoing reductive evolution, in what appears to be an adaptation to the protein-rich milk environment through the loss of superfluous functions. The nature of a number of the pseudogenes, remnants of genes involved in the transport and metabolism of various carbohydrates, points to an ancestor that most probably evolved in an environment where plant-carbohydrates were readily available. We recently extended these observations to several other *L. bulgaricus* genomes, and revealed a similar, but less advanced, adaptation in the ssp. *lactis* [8].

A very intriguing characteristic of the *L. bulgaricus* ATCC11842 genome is the presence of a 47.5 kbp inverted repeat [8] in the replication terminus region. The repeated sequence includes the putative *dif* site, the recombination site involved in the resolution of chromosome dimers [9], and, consequently, *L. bulgaricus* contains two putative *dif* sites in opposite orientation at a distance of 2.5 kbp (0.55 kbp of repeated sequence at either side of the unique central sequence of 1.4 kbp) instead of the unique site found in other microbial genomes.

These observations in the *L. bulgaricus* ATCC 11842 genome were a reason to investigate if similar IRs are found in other *L. bulgaricus* genomes and in the related *L. delbrueckii* ssp. *lactis*. Here we present an analysis of 10 *L. delbrueckii* genome sequences, representing 5 strains of the ssp. *lactis* and 5 strains of the ssp. *bulgaricus*, and explore if these sequences can tell us more about the evolution history of the two subspecies.

## Materials and methods

**Genome sequences.** The genome sequences of 2 *L. delbrueckii* ssp. *bulgaricus* strains (Vib27 and Vib44) and 4 ssp. *lactis* strains (CNRZ327, CNRZ333, CNRZ226, and CNRZ700), were recently established to near completion as described in [8]. The earlier published complete genome sequences of the ssp. *bulgaricus* strains ATCC11842 [7], ATCC BAA365 [10], and 2038 [11], and the ssp. *lactis* strain NDO2 (originally classified as belonging to the ssp. *bulgaricus* [12], but in fact belonging to the ssp. *lactis* [8]) were retrieved from NCBI.

**Core-genomes.** Complete genomes retrieved from NCBI and ordered sets of scaffolds [8] from newly sequenced genomes were aligned using MOSAIC with default parameters [13, 14] to determine core-genomes, containing DNA sequence regions (> 1 kb, including coding and non-coding sequences) shared by all *L. delbrueckii* ssp. *bulgaricus* strains, by all *L. delbrueckii* ssp. *lactis* strains, or by all strains of the two ssp..

**Phylogenetic reconstruction.** Strain phylogenetic trees were computed on the basis of core genome alignments using two reconstruction methods. PhyML 3.0 [15] was used to reconstruct maximum likelihood trees (substitution model GTR +I +G selected using modelgenerator [16], 100 bootstrap replications). UPGMA trees were reconstructed using the PHYLIP package (<http://evolution.genetics.washington.edu/phylip.html>). Distance matrices were computed with DNAdist (Jukes-Cantor distance) after which Neighbor was used to construct the UPGMA tree (100 bootstrap replications). FigTree 1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>) and the R package ape [17] were used to visualize phylogenies.

**Determination of substitution and recombination rates within subspecies.** Nucleotide substitution and recombination rates within each ssp. were determined using ClonalFrame [18], in the core-genome sequences (as defined above). Since both the UPGMA and maximum likelihood tree had a congruent topology with high bootstrap values, we fixed

UPGMA as the starting tree and updated only the branch lengths, not the topology. The mutation rate  $\theta$  was fixed to the value given by the Watterson [19] estimate  $\widehat{\theta}_w$ . In the presence of recombination,  $\widehat{\theta}_w$  overestimates the mutation rate, which should lead to a conservative calling of recombination events. All other parameters were set to default values and estimated by ClonalFrame. Three independent runs with different priors were run for 150,000 generations and samples were recorded every 100 generations after a burn-in of 50,000 generations. All runs converged to the same posterior distributions as evidenced by visual inspection of the posterior distributions and more objectively by high Effective Sample Sizes (ESS > 220 for all parameters) and low potential scale reduction factor [20] (<1.12 for all parameters in all runs).

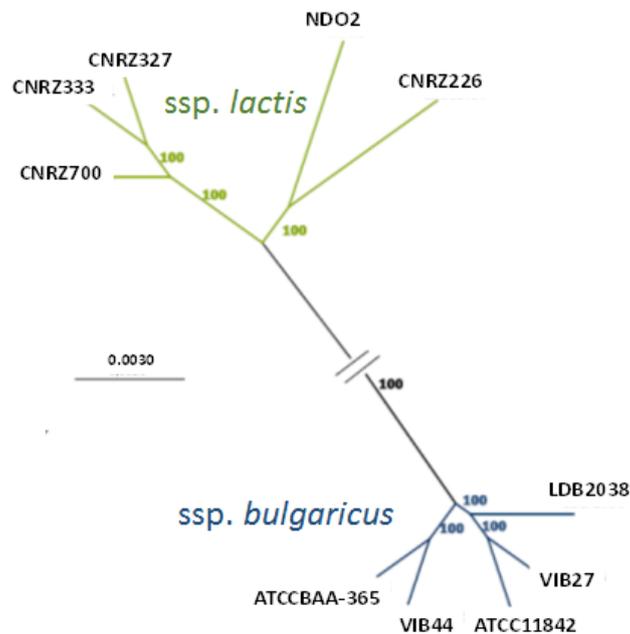
**Identification of large IRs in replication terminus regions of completely sequenced genomes.** 1368 complete bacterial genome sequences were retrieved from Genbank and analyzed for the presence of large inverted repeats comparable to those found in the *L. delbrueckii* ssp. *bulgaricus* ATCC11842 genome [7] using Repseek [21] with the following parameters: -L 1000 (repetition > 1000 bp), -c (circular chromosome), -i (inverted repeats). The results were filtered to keep only inverted repeats where the size of the repeated sequence was  $\geq 8$  kbp, the distance between the two copies of the repeated sequence comprised between 1 and 10 kbp, and where the two repeats were  $\geq 98\%$  identical. For the genomes where a large IR meeting these criteria was identified, an intra-chromosome homology dot plot was generated using Gepard [22] to visualize inverted repeats (i.e. the same chromosome is represented on the X-axis and on the Y-axis of the plot, and homology between the two copies of the chromosome indicated by dots. This results in an ascending diagonal line representing the identity between the same positions on the X-axis copy of the chromosome and the Y-axis copy of the chromosome. Inverted repeats become visible as additional descending diagonal lines that cross the ascending diagonal.). The origin of replication was

assumed to be situated near the position of the *dnaA* gene as is the case in most circular bacterial genomes [23]. A 200 kbp region at the opposite side of the chromosome was analyzed to verify the occurrence of a change of sign in the GC-skew characteristic for the replication terminus [24], and inverted repeats in this region manually inspected.

## Results

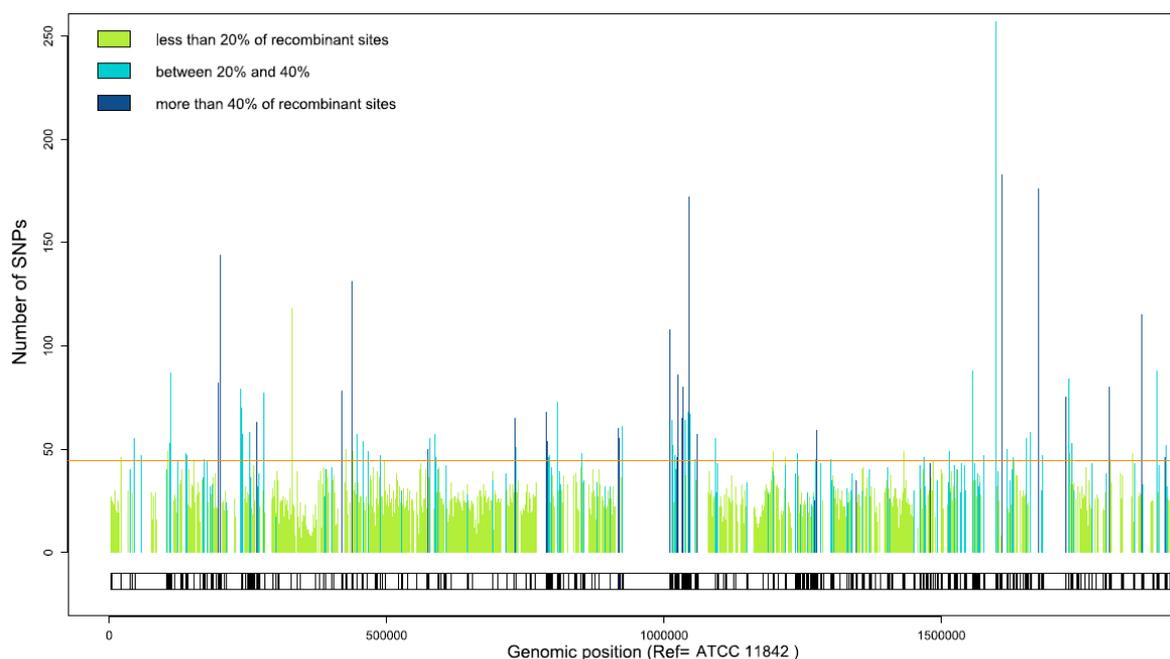
**Strain diversity is non-evenly distributed over the *L. delbrueckii* core-genome.** In this study we used 4 complete and 6 nearly complete *L. delbrueckii* genome sequences, together representing 5 *L. delbrueckii* ssp. *bulgaricus* strains and 5 *L. delbrueckii* ssp. *lactis* strains. We aligned the different genomes to determine the core-genome, consisting of the sequence intervals shared by all the strains studied.

The genealogy of the strains reconstructed on the basis of the core genomes shows a clear separation between the ssp. *bulgaricus* and the ssp. *lactis* clades (Fig. 1), consolidating the 16S rRNA and MLST based strain classifications of the two subspecies [8]. When the closely related *Lactobacillus acidophilus* is used as an external root, the root is placed on the branch separating the two clades (results not shown).



**Figure 1. Maximum likelihood tree obtained with phyML using core genome sequences of the ten strains in this study.** The numbers on nodes indicate bootstrap values. The scale represents the number of mutations per site.

The 10 strains are strongly related with at least 97% of identity, but diversity between the 10 strains appears to be non-evenly distributed over the core genome (Fig. 2), which may be indicative of recombination that occurred during the evolution of the clade. Remarkably, a relatively high amount of diversity is concentrated in a region (to the right of position 1000,000 in Fig. 2) close to the replication terminus, just next to a region that is not included in the core genome (i.e. not conserved among strains).



**Figure 2. SNPs density along the core genome of the 10 *L. delbrueckii* strains.** The density of SNPs in non-overlapping 1 kbp windows is indicated. The white zones correspond to regions not included in the core genome. The orange threshold line represents the amount of SNPs above which a region is significantly enriched in SNPs (binomial test,  $p < 0.05$  after *fd*r correction). The bar color indicates the fraction of SNPs inferred as resulting from recombination within a window. The black bars at the bottom of the figure indicate the actual position of recombination events detected by ClonalFrame.

### **Higher contribution of recombination to core genome diversity in *L. delbrueckii* ssp.**

***lactis* than in ssp. *bulgaricus*.** In order to get more insight in the evolution of the two *L.*

*delbrueckii* ssp., we used ClonalFrame [18], a method developed to infer the clonal relationships of bacteria. ClonalFrame distinguishes recombination events from point mutations assuming that recombination events introduce a constant and elevated rate of substitution in a contiguous sequence region while point mutations are scattered and rare.

Under this assumption, a high local density of SNPs is the footprint of past recombination events. When applied to the core-genomes of the 10 *L. delbrueckii* strains, it appears that 35% of the observed diversity can be attributed to recombination events (Table 1). Within the ssp. *lactis* (5 strains), the corresponding figure is 40%, and within the ssp. *bulgaricus* 36%. These results show that recombination contributed more to diversity in the ssp. *lactis* strains than in the ssp. *bulgaricus* strains, relative to the contribution of point mutations. The size of the recombinant tracts ranged from 5 to 1688bp, with an average between 189 and 291 bp, depending on the dataset (Table1).

**Table 1. Datasets used in this study and principal results obtained using ClonalFrame.**

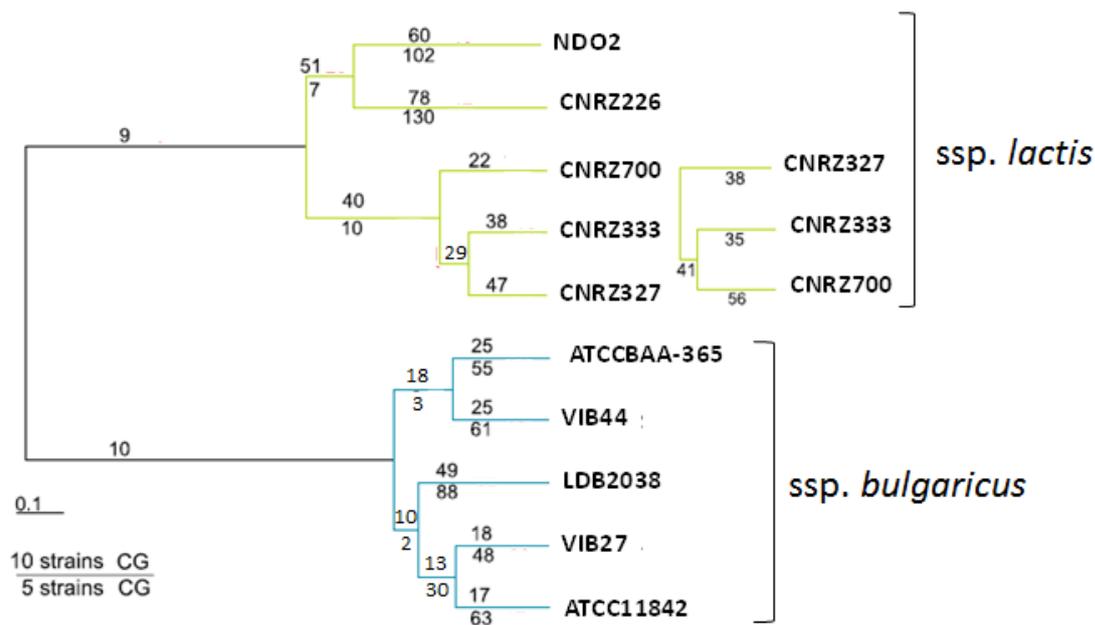
Core genome alignments

	10 strains	5 strains <i>ssp. lactis</i>	5 strains <i>ssp. bulgaricus</i>
<b>Datasets</b>			
Total length	1,169,615	1,377,054	1,532,043
Number of contigs	322	308	210
Polymorphic sites	34,720 (3.00%)	22,703 (1.65%)	15,390 (1.00%)
<b>Clonal Frame Results</b>			
Recombination events	559	419	350
r/m (1)	0.537 (35%)	0.679 (40%)	0.563 (36%)
Tract Size (median-mean) (2)	203-291	141-204	114-189

(1) r/m, ratio of the number of polymorphic sites resulting from recombination and the number of polymorphic sites resulting from point mutations; percentages indicate the fraction of observed diversity that can be attributed to recombination; (2) tract size, length of recombined sequence (nucleotides).

Fig. 3 shows the distribution of detected recombination tracts over the strain genealogy. The number of recombination events appears to be similar among ssp. *bulgaricus* strains,

while among the *ssp. lactis* strains there are much more recombination events for strains CNRZ226 and NDO2 than for the other strains. This may be explained by the longer branch lengths separating these strains from the other *ssp. lactis* strains (Fig. 1). Likewise, the fact that in the *ssp. lactis* more recombination events are observed than in the *ssp. bulgaricus* may be explained by the higher diversity among the *ssp. lactis* strains than among the *ssp. bulgaricus* strains (Fig. 1).



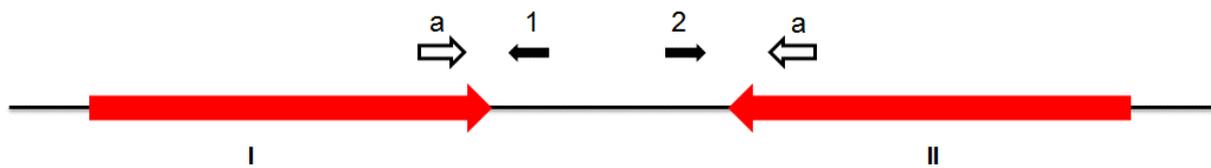
**Figure 3. Strain genealogy reconstructed with ClonalFrame.** Each branch of the genealogy is labeled with 2 values corresponding to the number of detected recombination events affecting the branch according to the different datasets : Complete Genomes (CG) alignments, 10 strains dataset over the branch, 5 strains dataset under the branch. For the *ssp. lactis*, the topology represented on the left results from the analysis of the 10 strains dataset. The alternative topology obtained with the 5 *ssp. lactis* genome dataset is represented on the right.

***L. delbrueckii ssp. bulgaricus* genomes, but not *L. delbrueckii ssp. lactis* genomes, contain a unique large inverted repeat in the replication terminus region.** An intriguing observation made in the first complete genome sequence of *L. delbrueckii ssp. bulgaricus* concerns the presence of a large, 47.5 kbp, inverted repeat in the replication terminus region (consisting of repeats of 23 kbp each that enclose a unique central region of 1.4 kbp) [7]. In bacterial genomes, inverted repeats of this size are extremely rare and have to our knowledge only been reported in the replication terminus region of certain, normally linear, *Streptomyces*

chromosomes that had circularized after telomere deletion [25]. In the present study, we analyzed the replication termination regions of 1368 complete bacterial genomes present in the public databases to see if similar structures, which may not have been reported in literature, could be detected in other bacteria. We identified 5 genomes containing IRs with repeats > 8 kbp separated by a unique central region < 10 kbp (not shown). For 4 of these genomes, the IR was situated in the replication terminus region. These included the 3 earlier published *L. delbrueckii* ssp. *bulgaricus* genomes (strains ATCC11842, ATCC BAA365, and 2038), with sizes varying between strains from 23 to 37.6 kbp for the repeated sequence and from 1 to 8.8 kbp for the unique central region between the repeats (see below). Only one other genome, of *Magnetospirillum magneticum* AMB-1, contained an IR in the replication terminus region, albeit with a much smaller sized repeat (9.5 kbp) than in the *L. delbrueckii* genomes. The repeats were separated by a unique region of 2.9 kbp (not shown). Among completely sequenced genomes, large size IRs in the replication terminus like the one originally identified in *L. delbrueckii* ssp. *bulgaricus* ATCC11842 thus appear to be present only in *L. delbrueckii*.

In the *L. bulgaricus* ATCC11842 genome, this IR is situated at positions 918952–966484, which coincides with the region that is not covered by the core-genome of the 10 strains in the present study (Fig. 2), signifying that important differences between strains exist in this region. We therefore analyzed the replication terminus regions of our recently sequenced *L. delbrueckii* genomes for the presence of this type of inverted repeats. Even if the genome sequences are not complete due to the presence of repeated sequences [8], several lines of evidence indicate the presence of large IRs in the replication termination regions of both newly sequenced *L. delbrueckii* ssp. *bulgaricus* genomes, in agreement with their presence in the earlier established complete ssp. *bulgaricus* genome sequences mentioned above: 1) sequence assembly ends (i.e. reaches the end of a scaffold) near the replication terminus, as expected in the case of a repeated sequence (only one copy of the repeated sequence is

expected to be assembled); 2) the average sequencing depth of a 27 (strain Vib27) or 18 kbp sequence (strain Vib44) near the replication terminus appears to be twice as high as in the adjacent regions, as expected if this sequence is repeated in the genome (results not shown). PCR amplification analysis corroborated these indications (Fig. 4). Remarkably, the newly sequenced *ssp. lactis* genomes don't show either of these signs, and the presence of a large IR in the replication terminus region of the chromosome thus appears to be an exclusive property of the *ssp. bulgaricus*. The complete genome sequence of *L. delbrueckii* NDO2, originally classified as belonging to the *ssp. bulgaricus* [12], lacks an IR, in agreement with its reclassification in the *ssp. lactis* [8].



**Figure 4. Inverted repeat (IR) in the replication terminus region of the *ssp. bulgaricus* genomes.** Schematic representation of the IR (not to scale) in the *L. delbrueckii ssp. bulgaricus* strains, showing the repeated sequence (red arrows I and II). The unique sequence between the repeats is represented by a black line, as are adjacent chromosomal sequences at either side of the IR. Arrows a, 1 and 2 represent primers used in PCR amplification. Successful PCR amplification was obtained using only primer a, using primers 1 and a, and using primers 2 and a, for both strain Vib27 and Vib44 (not shown), indicating the existence of an IR. Primers were designed on the basis of the ATCC11842 sequence: a, CTCAGCTTGCACGTCATGAACG; 1, GGCTACTTTACTGCCAGGACAG; 2, GGTGTTTGGACCAGCATTAGCC.

**Important size variation of the conserved IR structure foretells its disappearance.** A more detailed analysis of the replication terminus regions for the 3 complete *L. delbrueckii ssp. bulgaricus* genomes (ATCC11842, ATCC BAA365, and 2038) reveals similar, largely homologous, structures, albeit with different sizes of the repeated and unique central sequences (Fig. 5). The deduced structures for the nearly complete sequences of strains Vib27 and Vib44 follow the same pattern.

The repeat size varies from 18 kbp for strain Vib44 to 37.6 kbp for strain ATCC BAA365, where the smaller repeats appear to be derived from the repeat represented in strain ATCC BAA365 by internal deletions and deletions from the righthand side of the ATCC BAA365 IR. In some strains, additional sequences with respect to the ATCC BAA365 sequence, but just

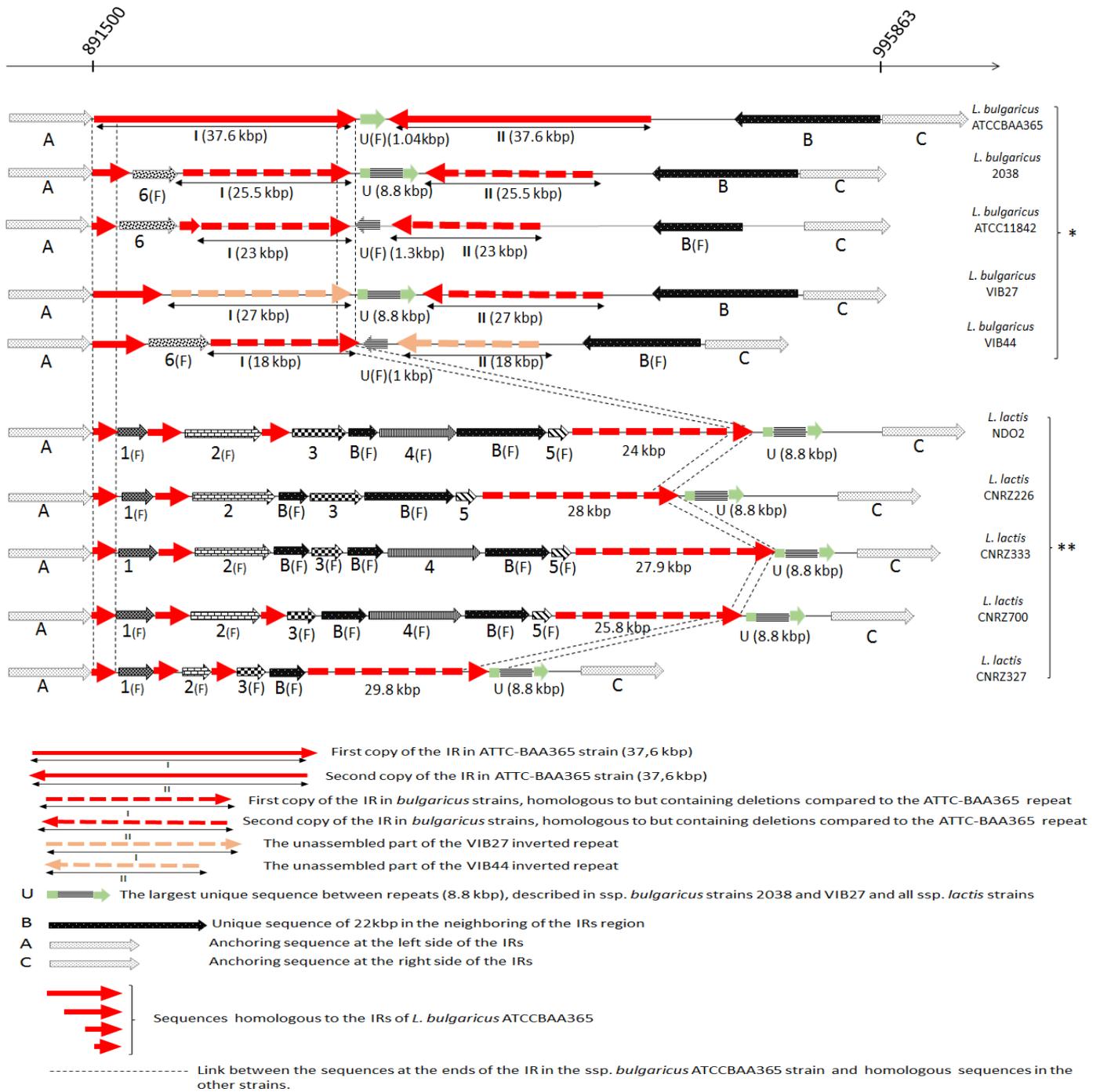
outside of the repeat in the strain concerned, are present (Fig. 5, sequence 6 or 6(F)). The central sequence in between the repeats varies from 1 to 8.8 kbp (Fig. 5), where the smaller sequences appear to be derived from the largest sequence represented in strains 2038 and Vib27.

The simplest hypothesis to explain these observations may be to assume that an ancestral IR consisted of repeats resembling those present in strain ATCC BAA365 but containing the sequence 6 insertion, and the central sequence present in strains 2038 and Vib27, and that all present IRs have derived from this ancestral repeat by differential deletions.

In the *ssp. lactis* strains, a non-repeated region with homology to one copy of the *ssp. bulgaricus* repeat can be found, of which the beginning and end correspond to the beginning and the end of the ATCC BAA365 repeat and are conserved in all *L. delbrueckii* strains examined, *ssp. bulgaricus* and *ssp. lactis* alike (Fig. 5). Within these *ssp. lactis* regions, several parts are found that are not present in the *ssp. bulgaricus* genomes (Fig. 5, sequences 1, 2, 3, 4, 5) and, intriguingly, also parts of a sequence that is exterior to the repeats in the *ssp. bulgaricus* strains (Fig. 5 sequence B). The *ssp. lactis* strains all contain the longest version of the sequence that separates the inverted repeats in the *ssp. bulgaricus* strains (8.8 kbp; Fig. 5 sequence U). Analogous to the hypothesis put forward for the *ssp. bulgaricus* genomes, and taking the status of sequence B in the two *ssp.* into account, one might speculate that a common ancestor of the *ssp. lactis* strains contained the sequences depicted in red in Fig. 5 as well as the sequences 1, 2, 5 and U. Sequence B, or part of it, may then have been inserted between sequences 2 and 5, followed by insertion of sequences 3 and 4 in sequence B to yield a strain similar to strain CNRZ333. The *ssp. lactis* strains studied here would then be deletion derivatives of the latter strain.

Together, the observations indicate that extensive recombination, duplication and deletion formation have taken place in the replication terminus region of *L. delbrueckii*, to produce a hallmark difference between the *ssp. lactis* and *bulgaricus*, where the latter contains a very

large IR that is absent from the former. High sequence similarity between the IRs in the different *ssp. bulgaricus* strains accompanied by important variability in size furthermore suggests that in this *ssp.* the IR is in the course of elimination. The latter conclusion is further backed by the observation that many genes in the IR of *L. delbrueckii ssp. bulgaricus* ATCC11842 are fragmented pseudogenes (Table 3).



**Figure 5. Organisation of the replication terminus regions of *L. delbrueckii* strains (not to scale).** \*: replication terminus region in the *L. delbrueckii* *ssp. bulgaricus* strains. \*\*: replication terminus region in the *L. delbrueckii* *ssp. lactis* strains. Sequence 1, 2, 3, 4 and 5: additional sequences of, respectively 6kbp, 13kbp, 9 kbp, 17 kbp, 2 kbp, in the only copy of the repeat in the *ssp. lactis* strains compared to the ATCC-BAA365 strain. Sequence 6: additional sequences of 8 kbp neighboring the ATCC11842 IR region, compared to strain ATCC-BAA365. (F): Fragmented. The scale above the figure represents the position between sequence A and C on the ATCC-BAA365 genome.

**Table 2. Details of the replication terminus region in *L. delbrueckii* genomes**

	<i>ssp. bulgaricus</i> ATCCBAA365	<i>ssp. bulgaricus</i> 2038	<i>ssp. bulgaricus</i> ATCC11842	<i>ssp. bulgaricus</i> VIB27	<i>ssp. bulgaricus</i> VIB44	<i>ssp. lactis</i> NDO2	<i>ssp. lactis</i> CNRZ226	<i>ssp. lactis</i> CNRZ333	<i>ssp. lactis</i> CNRZ700	<i>ssp. lactis</i> CNRZ327
Complete IR size (kbp) (repeats and the unique central sequence they enclose)	76.24	59.8	47.3	62.8*	37*	NA	NA	NA	NA	NA
Size (kbp) of region between end of sequence A and start of sequence U (Fig. 5)	37.6	37.5	37	37.7*	37.6*	64.2	64.1	69.5	63.8	32.5
Size (kbp) of region between anchoring sequences A and C	104.24	95.8	84.3	97.8*	82.6 *	88,8	90.7	90.4	86.2	53.89

\*Estimated size including the unassembled repeat of the IR.

**Table 3. Genes in repeats of completely sequenced *L. delbrueckii* ssp. *bulgaricus* genomes.**

<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> ATCCBAA365		<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> ATCC11842*		<i>L. delbrueckii</i> ssp. <i>bulgaricus</i> 2038	
<b>Genes in the Inverted repeat region</b>					
<b>locus tag</b>	<b>Function</b>	<b>locus tag</b>	<b>Function</b>	<b>locus tag</b>	<b>Function</b>
LB952	Hypothetical protein	ldb1064	Conserved hypothetical protein (fragment)	LBU903	Hypothetical conserved protein
LB953	Hypothetical protein	ldb1065	Conserved hypothetical protein (fragment)	LBU904	Oligopeptide ABC transporter substrate binding protein
LB954	Hypothetical protein	ldb1066	ABC transporter, ATP-binding protein (fragment)	LBU905	pheromone-peptide binding protein
LB955	Hypothetical protein	ldb1067	Carbamoyl-phosphate synthase, large chain (fragment1)	LBU906	Methyltransferase
LB956	Saccharopine dehydrogenase related protein	ldb1068	Carbamoyl-phosphate synthase, large chain (fragment2)	LBU907	ABC transporter, transmembrane region:ABC transporter
LB957	Hypothetical protein	ldb1069	Carbamoyl-phosphate synthase, large chain (fragment3)	LBU908	Carbamoyl-phosphate synthase, large subunit, glutamine-dependent
LB959	Hypothetical protein	ldb1070	Carbamoyl-phosphate synthase small chain	LBU909	Carbamoyl-phosphate synthase, small subunit
LB960	Hypothetical protein	ldb1071	Conserved hypothetical protein (fragment1)	LBU910	Carbamoyl-phosphate synthase , small chain
LB961	Metal-dependent amidase/aminoacylase/ Carboxypeptidase	ldb1072	Conserved hypothetical protein (fragment2)	LBU911	Hypothetical conserved protein
LB962	Phosphoribosylanthranilate isomerase	ldb1073	Conserved hypothetical protein (fragment3)	LBU912	Hypothetical protein
LB963	Hypothetical proteinn (transposase)	ldb1074	Putative multidrug efflux protein	LBU913	Na <sup>+</sup> driven multidrug efflux pump
LB964	Hypothetical protein	ldb1075	Putative transcriptional regulator of multidrug efflux pump gene (MerR family)	LBU914	Hypothetical conserved protein
LB965	Peptide methionine sulfoxide reductase	ldb1076	Hypothetical protein	LBU915	Hypothetical protein
LB966	Hypothetical protein	ldb1077	Putative transcriptional regulator (fragment1)	LBU916	Hypothetical conserved protein
LB967	Hypothetical protein	ldb1078	Putative transcriptional regulator (fragment2)	LBU917	site-specific recombinase, phage integrase family
LB969	Hypothetical protein	ldb1079	Putative integrase/recombinase	LBU918	Putative acetyltransferase

LB971	Hypothetical protein	ldb1080	Putative acetyltransferase (fragment)	LBU919	ABC transporter, intermembrane subunit precursor
LB975	Carbamoyl-phosphate synthase small subunit		CDS	LBU920	Hypothetical protein
LB978	Na <sup>+</sup> -driven multidrug efflux pump	ldb1081	Putative ABC transporter, permease protein (fragment1)	LBU921	Hypothetical protein
LB979	Predicted transcriptional regulator	ldb1082	Putative ABC transporter, permease protein (fragment2)	LBU922	Arsenate reductase, arsC, protein-tyrosine-phosphatase family enzyme
LB980	hypothetical protein	ldb1083	Hypothetical protein	LBU923	Hypothetical protein
LB981	Predicted transcriptional regulator	ldb1084	Hypothetical membrane protein	LBU924	ABC transporter ATP-binding protein
LB982	Integrase	ldb1085	Hypothetical protein	LBU925	ABC transporter, permease protein
LB984	Hypothetical protein	ldb1086	Putative phosphotyrosine protein phosphatase	LBU926	Hypothetical conserved protein
LB985	Hypothetical protein		CDS	LBU927	Putative flavoprotein subunit of a reductase
LB986	Hypothetical protein	ldb1087	Hypothetical protein	LBU928	Ribosomal-protein-serine acetyltransferase, putative
LB987	Protein-tyrosine-phosphatase	ldb1088	ABC transporter, ATP-binding protein		
LB988	Hypothetical protein	ldb1089	ABC transporter, permease protein		
LB989	ABC-type antimicrobial peptide transport system, ATPase component	ldb1090	Hypothetical protein		
LB990	ABC-type antimicrobial peptide transport system, permease component	ldb1091	Hypothetical protein		
LB991	Hypothetical protein	ldb1092	Putative transcriptional regulator (fragment)		
LB992	LysR substrate binding domain	ldb1093	Putative transcriptional regulator (fragment)		
LB996	Uncharacterized phage-associated protein	ldb1094	Putative transcriptional regulator (fragment)		
LB997	Uncharacterized phage-associated protein	ldb1095	Putative fumarate reductase (flavoprotein)		
LB998	Hypothetical protein	ldb1096	Putative phage protein (fragment)		
LB999	Hypothetical protein	ldb1097	Putative phage protein (fragment)		
LB1000	Acetyltransferase, including N-acetylase of ribosomal protein	ldb1098	Conserved hypothetical protein (fragment)		
LB1001	Hypothetical protein	ldb1099	Hypothetical protein		

LB1002	Hypothetical protein		CDS		
		ldb1100	Hypothetical protein		
			CDS		
		ldb1101	Putative acetyltransferase		
		ldb1102	Putative surface layer protein (fragment)		
<b>genes in the unique central sequence</b>					
LB1003	hypothetical protein	Ldb1103	aspartate amino transferase (fragment)	LBU929	Hypothetical conserved protein
LB1004	peptidyl-prolyl cis-trans isomerase	Ldb1104	peptidyl-prolyl cis-trans isomerase	LBU930	peptidyl-prolyl cis-trans isomerase
LB1005	hypothetical protein	Ldb1105	Hypothetical conserved protein (fragment)	LBU931	Aspartate aminotransferase
				LBU932	Dihydrodipicolinate reductase
				LBU933	Dihydrodipicolinate synthase
				LBU934	Amino acid amidohydrolase
				LBU935	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase
				LBU936	Aspartokinase-homoserinedehydrogenase
				LBU937	Diaminopimelate epimerase

\*Gene annotations are taken from the respective GenBank files. In the annotation of the ATCC11842 genome, particular attention was given to the annotation of pseudogenes (fragmented genes) [7]. This may not have been the case for the other genomes shown.

## Discussion

Both *L. delbrueckii* ssp. *bulgaricus* and ssp. *lactis* are largely used in dairy fermentations, and their respective genome sequences show clear signs of reductive evolution in adaptation to the milk environment [7, 8]. The ssp. *bulgaricus* appears to be more advanced in this adaptation, as becomes clear from a more extensive inactivation and elimination of genes involved in the metabolism of various carbohydrate sources or the biosynthesis of amino acids, which have become superfluous in the new environment [8].

In the present work, we highlight two other aspects of the genome structure of these bacteria. First, we show that within the common core genome, recombination events, i.e. genetic exchange, contributed more importantly to the diversity of the ssp. *lactis* than to the diversity of the ssp. *bulgaricus*, relative to the contribution of point mutations. It is hard to say what the reason for this difference may be as many factors can influence effective recombination (e.g. availability of source DNA, relatedness of source DNA, efficiency of DNA uptake, efficiency of recombination, ....). It is tempting to believe that the difference may be caused by differences in the microbiological environment, where cheese starter cultures in which the ssp. *lactis* occurs are often less defined, containing a higher variety of strains, than yogurt starters that contain the ssp. *bulgaricus*.

The same study showed that diversity was not evenly distributed over the common core genome, with a concentration of SNPs close to the replication terminus region that had earlier been identified as containing an unusual large IR in *L. delbrueckii* ssp. *bulgaricus* ATCC11842 [7]. In the present study, similar IRs are shown to be present in the other ssp. *bulgaricus* strains investigated, but not in the ssp. *lactis* strains. The replication terminus region turns out to be a highly variable region of the *L. delbrueckii* genome. As a consequence, the region does not make part of the core genome.

Our observations in the replication terminus regions of the *ssp. bulgaricus* genomes lead us to postulate that the IRs all derive from an ancestral IR of at least 100 kbp (repeats of at least 45.6 kbp enclosing a unique central region of at least 8.8 kbp), which is gradually eliminated through multiple deletion events. This hypothesis is consistent with the large size variation observed between the IRs of different strains while at the same time they share a high degree of homology, and with the fact that, at least in strain ATCC11842, many genes in the IR appear to be pseudogenes (fragmented genes in the course of elimination). It is also consistent with the earlier hypothesis that the maintenance of the IR would be disadvantageous because it contains a duplicated *dif* site, a feature that could interfere with the efficient segregation of dimeric chromosomes that may be formed during replication [7]. While it is known that small IRs can be entirely excised during DNA replication, at least *in vitro* (through slippage of the DNA polymerase on small direct repeats at the base of the stem-loop structure formed by the IR [26]), there is no obvious way for the bacterial cell to remove large inverted repeats like those present in *L. delbrueckii ssp. bulgaricus* as they will probably never form a stem-loop structure. Even if possible, complete one-time removal of the IR would probably not be desirable as it would leave the genome without any *dif* site. Gradual deletion involving other mechanisms would thus be the only possibility to clear the genome of such a large IR.

Intriguingly, the size of the whole replication terminus region (Fig. 5) is very similar in all *ssp. bulgaricus* and *ssp. lactis* strains studied except CNRZ327 (Table 2), in spite of the presence of the large sequence duplication (IR) in the *ssp. bulgaricus*. This is explained by the fact that the replication terminus region in the *ssp. lactis* strains contains large stretches of sequence that are not found in the *ssp. bulgaricus*. Part of these may have been present in the common ancestor of the two *ssp.* (Fig. 5, sequences 1, 2, 5), while others appear to have been inserted in the *ssp. lactis* lineage. Analogous to the *ssp. bulgaricus*, the *ssp. lactis* region

appears to be subject to extensive deletion and recombination events, starting from a common *ssp. lactis* ancestor.

From our analysis of 1368 complete circular bacterial genomes it appears that IRs of the size described in this study are unique to *L. delbrueckii ssp. bulgaricus*. We can only speculate on how such a large IR may have originated in the *L. delbrueckii* genome. In human cancer cells, where the formation of large inverted repeats has been studied as an initiating event of extensive gene amplification, and other eukaryotes, intrastrand annealing and illegitimate repair following a double-strand DNA break near to a short IR appear to be key steps of the process [27, 28]. This mechanism has also been evoked to explain the formation of IRs in an *E. coli sbcCD recA* mutant [29]. A similar event may have taken place in *L. delbrueckii*, especially as in the replication termination region double strand breaks may be provoked by stalling replication forks [30] or by the accidental cleavage of incompletely segregated chromosomes at septum closure [31]. In spite of the expected rarity of large IR formation and a high probability of counterselection of such structures, the IR may have been fixed in the population through positive selection if, at the time of formation, the duplication of genes in the repeat represented an advantage outweighing the inconvenience. Alternatively, even if disadvantageous, the IR may have been fixed by chance if the population went through a population bottleneck. As discussed above, once fixed in the population, IRs appear to be difficult to remove.

## References

1. Nicolas P, Bessieres P, Ehrlich SD, Maguin E, van de Guchte M: **Extensive horizontal transfer of core genome genes between two *Lactobacillus* species found in the gastrointestinal tract.** *BMC evolutionary biology* 2007, **7**:141.
2. Elli M, Callegari ML, Ferrari S, Bessi E, Cattivelli D, Soldi S, Morelli L, Goupil Feuillerat N, Antoine JM: **Survival of yogurt bacteria in the human gut.** *Appl Environ Microbiol* 2006, **72**(7):5113-5117.
3. Mater DD, Bretigny L, Firmesse O, Flores MJ, Mogenet A, Bresson JL, Corthier G: **Streptococcus thermophilus and Lactobacillus delbrueckii subsp. bulgaricus survive gastrointestinal transit of healthy volunteers consuming yogurt.** *FEMS Microbiol Lett* 2005, **250**(2):185-187.
4. Metchnikoff E: **The prolongation of life.** London: Heinemann; 1907.
5. Santos Rocha C, Lakhdari O, Blottiere HM, Blugeon S, Sokol H, Bermudez-Humaran LG, Azevedo V, Miyoshi A, Dore J, Langella P *et al*: **Anti-inflammatory properties of dairy lactobacilli.** *Inflamm Bowel Dis* 2012, **18**(4):657-666.
6. Santos Rocha C, Gomes-Santos AC, Garcias Moreira T, de Azevedo M, Diniz Luerce T, Mariadassou M, Longaray Delamare AP, Langella P, Maguin E, Azevedo V *et al*: **Local and systemic immune mechanisms underlying the anti-colitis effects of the dairy bacterium *Lactobacillus delbrueckii*.** *PLoS One* 2014, **In press**.
7. van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, Robert C, Oztas S, Mangenot S, Couloux A *et al*: **The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(24):9274-9279.
8. El Kafsi H, Binesse J, Loux V, Buratti J, Boudebouze S, Dervyn R, Maguin E, van de Guchte M: ***Lactobacillus delbrueckii* ssp. lactis and ssp. bulgaricus: a chronicle of evolution in action.** *submitted for publication* 2014.
9. Kuempel PL, Henson JM, Dircks L, Tecklenburg M, Lim DF: **dif, a recA-independent recombination site in the terminus region of the chromosome of *Escherichia coli*.** *New Biol* 1991, **3**(8):799-811.
10. 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.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(42):15611-15616.
11. Hao P, Zheng H, Yu Y, Ding G, Gu W, Chen S, Yu Z, Ren S, Oda M, Konno T *et al*: **Complete sequencing and pan-genomic analysis of *Lactobacillus delbrueckii* subsp. bulgaricus reveal its genetic basis for industrial yogurt production.** *PLoS One* 2011, **6**(1):e15964.
12. Sun Z, Chen X, Wang J, Zhao W, Shao Y, Guo Z, Zhang X, Zhou Z, Sun T, Wang L *et al*: **Complete genome sequence of *Lactobacillus delbrueckii* subsp. bulgaricus strain ND02.** *J Bacteriol* 2011, **193**(13):3426-3427.
13. Chiapello H, Gendrait A, Caron C, Blum J, Petit MA, El Karoui M: **MOSAIC: an online database dedicated to the comparative genomics of bacterial strains at the intra-species level.** *BMC Bioinformatics* 2008, **9**:498.
14. Devillers H, Chiapello H, Schbath S, Karoui ME: **Robustness assessment of whole bacterial genome segmentations.** *J Comput Biol* 2011, **18**(9):1155-1165.
15. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Systematic biology*, **59**(3):307-321.
16. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO: **Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad**

- hoc assumptions for choice of matrix are not justified.** *BMC evolutionary biology* 2006, **6**:29.
17. Paradis E, Claude J, Strimmer K: **APE: Analyses of Phylogenetics and Evolution in R language.** *Bioinformatics (Oxford, England)* 2004, **20**(2):289-290.
  18. Didelot X, Falush D: **Inference of bacterial microevolution using multilocus sequence data.** *Genetics* 2007, **175**(3):1251-1266.
  19. Watterson GA: **On the number of segregating sites in genetical models without recombination.** *Theoretical population biology* 1975, **7**(2):256-276.
  20. Gelman A, Rubin DB: **Inference from iterative simulation using multiple sequences.** *Statistical Science* 1992, **7**(4):457-472.
  21. Achaz G, Boyer F, Rocha EP, Viari A, Coissac E: **Repseek, a tool to retrieve approximate repeats from large DNA sequences.** *Bioinformatics (Oxford, England)* 2007, **23**(1):119-121.
  22. Krumsiek J, Arnold R, Rattei T: **Gepard: a rapid and sensitive tool for creating dotplots on genome scale.** *Bioinformatics (Oxford, England)* 2007, **23**(8):1026-1028.
  23. Mackiewicz P, Zakrzewska-Czerwinska J, Zawilak A, Dudek MR, Cebzat S: **Where does bacterial replication start? Rules for predicting the oriC region.** *Nucleic acids research* 2004, **32**(13):3781-3791.
  24. Lobry JR: **Asymmetric substitution patterns in the two DNA strands of bacteria.** *Molecular biology and evolution* 1996, **13**(5):660-665.
  25. Uchida T, Ishihara N, Zenitani H, Hiratsu K, Kinashi H: **Circularized chromosome with a large palindromic structure in *Streptomyces griseus* mutants.** *J Bacteriol* 2004, **186**(11):3313-3320.
  26. Canceill D, Ehrlich SD: **Copy-choice recombination mediated by DNA polymerase III holoenzyme from *Escherichia coli*.** *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(13):6647-6652.
  27. Tanaka H, Cao Y, Bergstrom DA, Kooperberg C, Tapscott SJ, Yao MC: **Intrastrand annealing leads to the formation of a large DNA palindrome and determines the boundaries of genomic amplification in human cancer.** *Molecular and cellular biology* 2007, **27**(6):1993-2002.
  28. Tanaka H, Tapscott SJ, Trask BJ, Yao MC: **Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(13):8772-8777.
  29. Darmon E, Eykelenboom JK, Lincker F, Jones LH, White M, Okely E, Blackwood JK, Leach DR: ***E. coli* SbcCD and RecA control chromosomal rearrangement induced by an interrupted palindrome.** *Molecular cell* 2010, **39**(1):59-70.
  30. Michel B, Flores MJ, Viguera E, Grompone G, Seigneur M, Bidnenko V: **Rescue of arrested replication forks by homologous recombination.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(15):8181-8188.
  31. Lesterlin C, Barre FX, Cornet F: **Genetic recombination and the cell cycle: what we have learned from chromosome dimers.** *Molecular microbiology* 2004, **54**(5):1151-1160.

## Conclusion

A travers ce travail, nous avons mis en évidence deux nouvelles différences de structure génomique entre les ssp. *lactis* et ssp. *bulgaricus* de *L. delbrueckii*.

D'une part au sein du core génome de l'espèce *L. delbrueckii*, les évènements d'échanges et de recombinaisons génétiques ont beaucoup plus contribué à la diversité au sein de la ssp. *lactis* qu'au sein de la ssp. *bulgaricus*.

D'autre part, le génome de la ssp. *bulgaricus* est caractérisé par une structure inversée répétée de grande taille, très rarement observée chez les bactéries. La raison de l'apparition de cette structure chez la ssp. *bulgaricus* reste inconnue. Cette structure aurait pu se fixer dans la population par un phénomène de « bottleneck » populationnel ou par sélection positive ce qui impliquerait qu'au moment de son apparition la duplication des gènes de la région du terminus était bénéfique pour la bactérie.

## VI. Discussion générale et perspectives

*L. delbrueckii* est un membre des lactobacilles du groupe acidophilus. Un groupe de bactéries proches de *L. acidophilus*. Ce groupe comprend un nombre important de bactéries isolées du tractus digestif et présentant des propriétés probiotiques, particulièrement d'immuno-modulation (*L. acidophilus*, *L. johnsonii*, *L. gasseri*).

*L. delbrueckii* est depuis longtemps utilisée pour la fermentation de produits laitiers où la sous-espèce *bulgaricus* est principalement connue pour son utilisation dans la fermentation du yaourt alors que la sous-espèce *lactis* est traditionnellement utilisée dans la production des fromages à pâte dure de type parmesan et emmental.

L'étude du génome de *L. delbrueckii* ssp. *bulgaricus* ATCC11842 a mis en évidence l'inactivation d'un important nombre de gènes principalement impliqués dans la biosynthèse d'acides aminés. Certains de ces gènes sont retrouvés sous forme de fragments dans le génome alors que d'autres ont totalement disparus [167]. La présence de fragments de gènes impliqués dans le transport et le métabolisme de carbohydrates de plantes dans le génome de *L. delbrueckii* ssp. *bulgaricus* ATCC11842 a permis d'émettre l'hypothèse d'un ancêtre qui aurait évolué dans un écosystème où les carbohydrates d'origine végétale seraient disponibles. En considérant les premières archives sur le yaourt datant de 3200 avant Jésus-Christ [267] on pourrait supposer qu'à travers le temps *L. delbrueckii* ssp. *bulgaricus* se serait adapté au milieu riche du lait (milieu riche en protéine et en lactose) par la perte de certaines fonctions devenues superflues [167].

Malgré le fait que *L. delbrueckii* ssp. *lactis* a pu être distinguée phénotypiquement de la ssp. *bulgaricus* par le métabolisme d'un plus large spectre de carbohydrates (principalement des carbohydrates de nature végétale [267]) elle reste moins étudiée au niveau génomique.

En plus de l'importance d'étudier *L. delbrueckii* pour son rôle biotechnologique dans la fermentation de certains produits laitiers l'ingestion quotidienne de grandes quantités de cette

bactérie (à travers la consommation de fromages et de yaourt) sans avoir un recul par rapport à ses effets sur la santé humaine a été l'une des principales raisons pour laquelle sa potentielle propriété probiotique a été étudiée. Ainsi les effets probiotiques longtemps utilisés comme argument de vente pour certains produits complémentés par des bactéries probiotiques sélectionnées pourraient être depuis toujours retrouvés dans certains produits laitiers.

Plusieurs études montrent des effets anti-inflammatoires de *L. delbrueckii in vitro* [213, 214, 219] ou un rôle bénéfique de protection contre l'inflammation induit chimiquement dans des modèles de maladies inflammatoires de l'intestin chez la souris [219, 220]. Un travail réalisé précédemment au sein de notre laboratoire a permis d'étudier *in vitro* l'effet anti-inflammatoire de 57 souches de *L. delbrueckii* montrant que l'ampleur de cet effet est souche-dépendant et qu'il est dû à l'action de protéines de la surface bactérienne [219]. Ce qui a été aussi intrigant dans les résultats obtenus est que les 57 souches de *L. delbrueckii* testées peuvent être divisées en deux catégories immuno-modulatrices : catégorie de souches avec un effet anti-inflammatoire élevé d'une part et catégorie de souches avec un effet intermédiaire à faible d'autre part. Ce qui a aussi été intéressant à noter est que les souches appartenant à la catégorie fortement anti-inflammatoire appartiennent à la sous-espèce *lactis* alors que la catégorie intermédiaire à faible est un mélange de souches des deux sous-espèces.

La mise en place des approches génomique, protéomique et transcriptomique comparatives entre souches des ssp. *lactis* et *bulgaricus* a été la stratégie choisie afin d'identifier le ou les facteurs potentiellement à l'origine des différences phénotypiques, principalement le phénotype anti-inflammatoire, entre les deux sous-espèces.

Le choix des souches pour l'étude comparative s'est fait à partir des 57 souches initialement étudiées *in vitro* pour leurs effets anti-inflammatoires. La sélection de 8 souches de *L. delbrueckii* comprend les deux souches les plus anti-inflammatoires de la ssp. *lactis*,

CNRZ327 et CNRZ333, ainsi que 6 autres souches appartenant aux deux ssp. et présentant des effets anti-inflammatoires intermédiaires à faibles.

La première partie de mon travail de thèse a été la finition du génome de la souche la plus anti-inflammatoire, *L. delbrueckii* ssp. *lactis* CNRZ327, permettant d'avoir toute l'information génétique nécessaire à la recherche des déterminants bactériens immuno-modulateurs. A savoir qu'au début de ces études aucun génome de la sous-espèce *lactis* n'avait été séquencé. C'est seulement en parallèle de nos études que le génome de *L. delbrueckii* NDO2, dont nous avons démontré l'appartenance à la ssp. *lactis*, est apparu dans les bases de données mais avec une attribution erronée à la ssp. *bulgaricus*.

A partir des 8 génomes de *L. delbrueckii* initialement sélectionnées pour leurs effets anti-inflammatoires différentiels et des 2 génomes complets restants de *L. delbrueckii* présents dans les bases de données (*L. delbrueckii* ssp. *lactis* NDO2 et *L. delbrueckii* ssp. *bulgaricus* 2038) une étude de génomique comparative a été réalisée dans le but d'explorer les différences génomiques pouvant expliquer les distinctions phénotypiques de ces deux sous-espèces de grand intérêt économique dans l'industrie laitière.

Malgré le nombre plus élevé d'IS chez les souches de la ssp. *lactis* (habituellement connues pour leur rôle dans le réarrangement génomique et l'inactivation de gènes) *L. delbrueckii* ssp. *lactis* possède une capacité métabolique indéniablement plus importante que celle de la ssp. *bulgaricus*. Toutefois une importante variabilité dans les capacités métaboliques est observée entre souches.

Les capacités fermentaires différentielles entre les deux ssp., pour la plupart confirmées génotypiquement, semblent en incohérence avec ce qui a pu être décrit précédemment [267].

En effet la distinction phénotypique entre ces deux ssp. étant préalablement basée sur la production d'acide par fermentation du maltose, mannose, saccharose, salicine, tréhalose, amygdaline ainsi que par l'hydrolyse de l'esculine, où les souches de la ssp. *lactis* sont les

seules décrites comme pouvant métaboliser ces carbohydrates, n'a pu être confirmé par notre étude.

Seules les fermentations du N-acetylglucosamine (sucre simple) et de l'amidon (carbohydrate complexe) permettent de distinguer entre les deux ssp. (sucres exclusivement métabolisés par les souches de la ssp. *lactis*). Ces deux sucres n'ont pas été décrits préalablement comme différenciellement métabolisés par les deux ssp..

Malgré le fait qu'habituellement la fermentation du D-mannose est considérée comme un des critères distinctifs entre la ssp. *lactis* et la ssp. *bulgaricus* ce travail montre qu'il s'agit du seul carbohydrate métabolisé par toutes les souches étudiées.

Ainsi cette étude a pu encore une fois montrer les limites de l'approche de classification des sous-espèces suivant leur profil fermentaire.

Comme décrit précédemment l'analyse *in silico* du premier génome complet de la ssp. *bulgaricus* [167] suggère que cette souche dérive d'un ancêtre qui serait capable de fermenter un certain nombre de carbohydrates d'origine végétale. Suite à des événements d'inactivation et élimination de gènes *L. bulgaricus* se serait adapté au milieu lait par une réduction en nombre de gènes impliqués dans des voies métaboliques devenues superflues. Ce travail de thèse a permis à la fois d'étendre cette observation à 4 autres génomes de *L. bulgaricus* et de montrer qu'une perte de gènes différentielle est à l'origine de la variabilité dans les capacités fermentaires entre souches.

La même analyse réalisée pour une première fois pour des souches de la ssp. *lactis* a permis de révéler un phénomène similaire, mais beaucoup moins avancé, d'adaptation au milieu lait par la perte souche dépendante de certaines voies de métabolisme de carbohydrates.

Une seconde importante adaptation au milieu lait consiste en la perte de voies de biosynthèse d'acides aminés, précédemment décrit pour *L. bulgaricus* ATCC11842 [167], est confirmée pour la ssp. *lactis*, mais toujours à un stade moins avancé.

Une des observations les plus intrigantes de cette analyse comparative est que les ssp. *bulgaricus* et *lactis* ont acquis toutes deux un système anti-porteur lactose-galactose pour l'import du lactose du milieu extracellulaire alors que l'ancêtre en commun possède un système PTS qui reste encore présent seulement chez la ssp. *lactis*.

Sachant que le système anti-porteur est plus adapté à un milieu riche en lactose alors que le système PTS l'est à un milieu pauvre en lactose l'environnement dans lequel a évolué l'ancêtre en commun serait un milieu où le lactose est présent mais à faible concentration.

En intégrant cette indication à celle d'un environnement ancestral contenant des carbohydrates d'origine végétale cela peut suggérer que l'ancêtre de *L. delbrueckii* aurait évolué dans le tractus digestif de mammifères (environnement où les deux conditions coexistent). Une hypothèse cohérente avec le fait que la majeure partie des lactobacilles du groupe acidophilus auquel appartient *L. delbrueckii* sont isolés de l'intestin.

Cette observation centrale dans l'étude comparative entre ssp. *lactis* et ssp. *bulgaricus* peut aussi être une indication clé dans la compréhension de l'origine de l'effet anti-inflammatoire de *L. delbrueckii*: En considérant le fait que l'effet fortement anti-inflammatoire est exclusivement observé chez certaines souches de la ssp. *lactis* et que cette ssp. est plus proche de l'ancêtre en commun cela laisse penser que l'effet anti-inflammatoire serait un phénotype ancestral. Cela est cohérent avec la présence de ce phénotype chez plusieurs autres bactéries proches du groupe acidophilus auquel appartient *L. delbrueckii*. L'absence de ce phénotype chez certaines souches de la ssp. *lactis* peut s'expliquer par le fait que la ssp. est en cours d'évolution vers une adaptation au milieu lait.

En ce qui concerne les déterminants bactériens à l'origine de l'effet anti-inflammatoire de *L. delbrueckii* les approches génomique, transcriptomique et protéomique comparatives ont permis de sélectionner un total de 56 gènes candidats. Parmi ces gènes candidats, certains semblent avoir des fonctions en adéquation avec une éventuelle interaction avec l'hôte pour

l'induction de l'effet immune-modulateur dont, les plus pertinents, deux différentes « mucus binding proteins » identifiées chez les deux souches fortement anti-inflammatoires de la ssp. *lactis*. Les protéines comme les « mucus binding protein » médiant l'adhésion des lactobacilles à la couche de mucus et aux cellules épithéliales ont pu être étudiées et fonctionnellement caractérisées [260–263]. Malgré le fait que les séquences des deux « mucus binding proteins » sont différentes entre les deux souches hautement anti-inflammatoires elles partagent un même « mucus binding domain » (présent en une seule copie dans la séquence de la « mucus binding protein » de CNRZ327 et en deux copies chez celle de CNRZ333). Ce domaine protéique est l'un des plus abondants décrit chez les lactobacilles isolés de l'intestin [229], ce qui appuie le rôle potentiel qu'il pourrait jouer dans l'interaction avec l'hôte. Bien que le rôle éventuel de ces « mucus binding protein » dans le pouvoir anti-inflammatoire de certaines souches de *L. delbrueckii* reste à établir elles constituent une indication supplémentaire suggérant l'appartenance de l'ancêtre de *L. delbrueckii* aux bactéries du tractus digestif. En regardant dans la liste des pseudogènes identifiés pour chacun des 10 génomes des souches de *L. delbrueckii* étudiées aucunes des souches de la ssp. *bulgaricus* ne possède des fragments de gènes annotés comme « mucus binding protein » alors que toutes les souches de la ssp. *lactis* (à l'exception de la souche *lactis* la moins anti-inflammatoire CNRZ226) contiennent un fragment de gène annoté comme « mucus binding protein ». Ces résultats appuient l'hypothèse que le « mucus binding protein » serait un gène ancestral de l'ancêtre intestinal qui serait totalement perdu dans les souches de la ssp. *bulgaricus* (plus avancée dans le processus d'adaptation au milieu lait) et en cours de perte dans la ssp. *lactis*. Ainsi, les gènes potentiellement impliqués dans l'effet anti-inflammatoire décrits chez CNRZ327 et CNRZ333 (mais totalement absent chez CNRZ226) peuvent aussi être des gènes ancestraux, maintenus seulement dans certaines souches de la ssp. *lactis*. Cet effet anti-

inflammatoire serait ainsi une propriété ancestrale de la bactérie, probablement importante dans son maintien dans le tractus digestif humain.

Les résultats préliminaires du test *in vitro* pour l'évaluation de l'effet anti-inflammatoire des gènes de *L. delbrueckii* sélectionnés montrent que le surnageant du clone exprimant le gène CNRZ327#2074, codant pour une protéine de fonction inconnue, donne les résultats les plus prometteurs. La recherche de la séquence génomique de cette protéine dans les autres génomes de *L. delbrueckii* présents dans les bases de données publiques, montrent qu'elle existe seulement chez *L. delbrueckii* ssp. *lactis* DSM 20072 isolée du tractus digestif humain. Cette observation appuierait l'idée d'un effet anti-inflammatoire hérité de l'ancêtre intestinal de *L. delbrueckii*. Le choix de souches de *L. delbrueckii* ssp. *lactis*, avec à la fois de bonnes propriétés fermentaires et probiotiques, peut constituer un grand avantage économique et un bon argument de vente pour les industriels.

Seul une partie des gènes candidats sélectionnés pour l'effet anti-inflammatoire ont pu être clonés et tester *in vitro* pour leurs effets immuno-modulateurs. Malgré les résultats préliminaires plutôt prometteurs obtenus pour certains gènes le test devrait être refait pour pouvoir confirmer ces effets.

Comme étapes ultérieures à l'identification des protéines à effets anti-inflammatoires il serait important de pouvoir confirmer ces effets en inactivant les gènes correspondants dans *L. delbrueckii* et testant le potentiel immuno-modulateur des bactéries mutées *in vitro* et *in vivo*.

Pour pouvoir comprendre le mécanisme d'action par lequel le déterminant bactérien interagit avec les cellules eucaryotes pour abaisser l'inflammation il est aussi important d'identifier les différents intervenants cellulaires ainsi que les voies de signalisation impliquées.

## **VII. Références**

## **Bibliographiques**

1. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464: 59–65.
2. Sekirov I, Russell SL, Antunes LCM, Finlay BB (2010) Gut microbiota in health and disease. *Physiol Rev* 90: 859–904.
3. Holmes E, Li J V, Marchesi JR, Nicholson JK (2012) Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab* 16: 559–564.
4. Xu J, Gordon JI (2003) Honor thy symbionts. *Proc Natl Acad Sci U S A* 100: 10452–10459.
5. Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122: 107–118.
6. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, et al. (2011) Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* 331: 337–341.
7. Gaboriau-Routhiau V, Rakotobe S, Lécuyer E, Mulder I, Lan A, et al. (2009) The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* 31: 677–689.
8. Kaser A, Zeissig S, Blumberg RS (2010) Inflammatory bowel disease. *Annu Rev Immunol* 28: 573–621. doi:10.1146/annurev-immunol-030409-101225.
9. Wu GD, Bushman FD, Lewis JD (2013) Diet, the human gut microbiota, and IBD. *Anaerobe*. doi:10.1016/j.anaerobe.2013.03.011.
10. Adolfsson O, Meydani SN, Russell RM (2004) Yogurt and gut function. *Am J Clin Nutr* 80: 245–256.
11. Natividad JMM, Verdu EF (2013) Modulation of intestinal barrier by intestinal microbiota: pathological and therapeutic implications. *Pharmacol Res* 69: 42–51. doi:10.1016/j.phrs.2012.10.007.
12. Chung H, Kasper DL (2010) Microbiota-stimulated immune mechanisms to maintain gut homeostasis. *Curr Opin Immunol* 22: 455–460. doi:10.1016/j.coi.2010.06.008.
13. Wendy S. Garrett, Jeffrey I. Gordon LHG (2010) Homeostasis and Inflammation in the Intestine. *Cell* 140: 859–870.
14. Harrison LC, Honeyman MC, Morahan G, Wentworth JM, Elkassaby S, et al. (2008) Type 1 diabetes: lessons for other autoimmune diseases? *J Autoimmun* 31: 306–310. doi:10.1016/j.jaut.2008.04.026.
15. Riesenfeld CS, Schloss PD, Handelsman J (2004) Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* 38: 525–552. doi:10.1146/annurev.genet.38.072902.091216.

16. Von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, et al. (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* 315: 1126–1130. doi:10.1126/science.1133420.
17. Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* 6: 805–814. doi:10.1038/nrg1709.
18. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. *Nature* 489: 220–230. doi:10.1038/nature11550.
19. Zhao L (2013) The gut microbiota and obesity: from correlation to causality. *Nat Rev Microbiol* 11: 639–647. doi:10.1038/nrmicro3089.
20. De Vos WM, de Vos EAJ (2012) Role of the intestinal microbiome in health and disease: from correlation to causation. *Nutr Rev* 70 Suppl 1: S45–56. doi:10.1111/j.1753-4887.2012.00505.x.
21. Kelly T, Yang W, Chen C-S, Reynolds K, He J (2008) Global burden of obesity in 2005 and projections to 2030. *Int J Obes (Lond)* 32: 1431–1437. doi:10.1038/ijo.2008.102.
22. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42: 937–948. doi:10.1038/ng.686.
23. Bäckhed F, Ding H, Wang T, Hooper L V, Koh GY, et al. (2004) The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* 101: 15718–15723. doi:10.1073/pnas.0407076101.
24. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, et al. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444: 1027–1031. doi:10.1038/nature05414.
25. Bäckhed F, Manchester JK, Semenkovich CF, Gordon JI (2007) Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc Natl Acad Sci U S A* 104: 979–984. doi:10.1073/pnas.0605374104.
26. Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Microbial ecology: human gut microbes associated with obesity. *Nature* 444: 1022–1023. doi:10.1038/4441022a.
27. Duncan SH, Belenguer A, Holtrop G, Johnstone AM, Flint HJ, et al. (2007) Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl Environ Microbiol* 73: 1073–1078. doi:10.1128/AEM.02340-06.
28. Furet J-P, Kong L-C, Tap J, Poitou C, Basdevant A, et al. (2010) Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes* 59: 3049–3057. doi:10.2337/db10-0253.

29. Schwartz A, Taras D, Schäfer K, Beijer S, Bos NA, et al. (2010) Microbiota and SCFA in lean and overweight healthy subjects. *Obesity (Silver Spring)* 18: 190–195. doi:10.1038/oby.2009.167.
30. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, et al. (2009) A core gut microbiome in obese and lean twins. *Nature* 457: 480–484. doi:10.1038/nature07540.
31. Bäckhed F, Ding H, Wang T, Hooper L V, Koh GY, et al. (2004) The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* 101: 15718–15723. doi:10.1073/pnas.0407076101.
32. Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI (2008) Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host Microbe* 3: 213–223.
33. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, et al. (2013) Richness of human gut microbiome correlates with metabolic markers. *Nature* 500: 541–546. doi:10.1038/nature12506.
34. Vaarala O, Atkinson MA, Neu J (2008) The “perfect storm” for type 1 diabetes: the complex interplay between intestinal microbiota, gut permeability, and mucosal immunity. *Diabetes* 57: 2555–2562. doi:10.2337/db08-0331.
35. Guidelines for the Evaluation of Probiotics in Food (n.d.).
36. Calcinaro F, Dionisi S, Marinaro M, Candeloro P, Bonato V, et al. (2005) Oral probiotic administration induces interleukin-10 production and prevents spontaneous autoimmune diabetes in the non-obese diabetic mouse. *Diabetologia* 48: 1565–1575. doi:10.1007/s00125-005-1831-2.
37. Matsuzaki T, Nagata Y, Kado S, Uchida K, Kato I, et al. (1997) Prevention of onset in an insulin-dependent diabetes mellitus model, NOD mice, by oral feeding of *Lactobacillus casei*. *APMIS* 105: 643–649.
38. Yadav H, Jain S, Sinha PR (2007) Antidiabetic effect of probiotic dahi containing *Lactobacillus acidophilus* and *Lactobacillus casei* in high fructose fed rats. *Nutrition* 23: 62–68. doi:10.1016/j.nut.2006.09.002.
39. Roesch LFW, Lorca GL, Casella G, Giongo A, Naranjo A, et al. (2009) Culture-independent identification of gut bacteria correlated with the onset of diabetes in a rat model. *ISME J* 3: 536–548. doi:10.1038/ismej.2009.5.
40. Valladares R, Sankar D, Li N, Williams E, Lai K-K, et al. (2010) *Lactobacillus johnsonii* N6.2 mitigates the development of type 1 diabetes in BB-DP rats. *PLoS One* 5: e10507. doi:10.1371/journal.pone.0010507.
41. Giongo A, Gano KA, Crabb DB, Mukherjee N, Novelo LL, et al. (2011) Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J* 5: 82–91. doi:10.1038/ismej.2010.92.

42. Dandona P, Aljada A, Bandyopadhyay A (2004) Inflammation: the link between insulin resistance, obesity and diabetes. *Trends Immunol* 25: 4–7.
43. Cani PD, Rottier O, Guiot Y, Neyrinck AM, Geurts L, et al. (2008) Changes in gut microbiota control intestinal permeability-induced inflammation in obese and diabetic mice through unexpected dependent mechanisms. 44th EASD Annu Meet og Eur Assoc Study Diabetes 51: S34–S35.
44. Wu X, Ma C, Han L, Nawaz M, Gao F, et al. (2010) Molecular characterisation of the faecal microbiota in patients with type II diabetes. *Curr Microbiol* 61: 69–78. doi:10.1007/s00284-010-9582-9.
45. McCormick DA, Horton LW, Mee AS (1990) Mucin depletion in inflammatory bowel disease. *J Clin Pathol* 43: 143–146.
46. Scaldaferri F, Gerardi V, Lopetuso LR, Del Zompo F, Mangiola F, et al. (2013) Gut Microbial Flora, Prebiotics, and Probiotics in IBD: Their Current Usage and Utility. *Biomed Res Int* 2013: 435268. doi:10.1155/2013/435268.
47. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, et al. (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55: 205–211. doi:10.1136/gut.2005.073817.
48. Martinez C, Antolin M, Santos J, Torrejon A, Casellas F, et al. (2008) Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. *Am J Gastroenterol* 103: 643–648. doi:10.1111/j.1572-0241.2007.01592.x.
49. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464: 59–65. doi:10.1038/nature08821.
50. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, et al. (2008) Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 105: 16731–16736. doi:10.1073/pnas.0804812105.
51. Halme L, Paavola-Sakki P, Turunen U, Lappalainen M, Farkkila M, et al. (2006) Family and twin studies in inflammatory bowel disease. *World J Gastroenterol* 12: 3668–3672.
52. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, et al. (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411: 599–603. doi:10.1038/35079107.
53. Zhou Z, Lin X-Y, Akolkar PN, Gulwani-Akolkar B, Levine J, et al. (2002) Variation at NOD2/CARD15 in familial and sporadic cases of Crohn's disease in the Ashkenazi Jewish population. *Am J Gastroenterol* 97: 3095–3101. doi:10.1111/j.1572-0241.2002.07105.x.

54. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, et al. (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411: 603–606. doi:10.1038/35079114.
55. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, et al. (2001) Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 357: 1925–1928. doi:10.1016/S0140-6736(00)05063-7.
56. Podolsky DK (2002) The current future understanding of inflammatory bowel disease. *Best Pract Res Clin Gastroenterol* 16: 933–943. doi:10.1053/bega.2002.0354.
57. Round JL, Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 9: 313–323. doi:10.1038/nri2515.
58. Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, et al. (2010) A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139: 1844–1854.e1. doi:10.1053/j.gastro.2010.08.049.
59. Lepage P, Häslér R, Spehlmann ME, Rehman A, Zvirbliene A, et al. (2011) Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* 141: 227–236. doi:10.1053/j.gastro.2011.04.011.
60. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, et al. (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 105: 16731–16736. doi:10.1073/pnas.0804812105.
61. Taurog JD, Richardson JA, Croft JT, Simmons WA, Zhou M, et al. (1994) The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. *J Exp Med* 180: 2359–2364.
62. Chiodini RJ, Van Kruiningen HJ, Thayer WR, Merkal RS, Coutu JA (1984) Possible role of mycobacteria in inflammatory bowel disease. I. An unclassified Mycobacterium species isolated from patients with Crohn's disease. *Dig Dis Sci* 29: 1073–1079.
63. Hultén K, Almashhrawi A, El-Zaatari FA, Graham DY (2000) Antibacterial therapy for Crohn's disease: a review emphasizing therapy directed against mycobacteria. *Dig Dis Sci* 45: 445–456.
64. Selby W, Pavli P, Crotty B, Florin T, Radford-Smith G, et al. (2007) Two-year combination antibiotic therapy with clarithromycin, rifabutin, and clofazimine for Crohn's disease. *Gastroenterology* 132: 2313–2319. doi:10.1053/j.gastro.2007.03.031.
65. Chassaing B, Darfeuille-Michaud A (2011) The commensal microbiota and enteropathogens in the pathogenesis of inflammatory bowel diseases. *Gastroenterology* 140: 1720–1728. doi:10.1053/j.gastro.2011.01.054.

66. Nagalingam NA, Lynch S V (2012) Role of the microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* 18: 968–984. doi:10.1002/ibd.21866.
67. Ohkusa T, Sato N, Ogihara T, Morita K, Ogawa M, et al. (2002) *Fusobacterium varium* localized in the colonic mucosa of patients with ulcerative colitis stimulates species-specific antibody. *J Gastroenterol Hepatol* 17: 849–853.
68. Ohkusa T, Okayasu I, Ogihara T, Morita K, Ogawa M, et al. (2003) Induction of experimental ulcerative colitis by *Fusobacterium varium* isolated from colonic mucosa of patients with ulcerative colitis. *Gut* 52: 79–83.
69. Sokol H, Lepage P, Seksik P, Doré J, Marteau P (2006) Temperature gradient gel electrophoresis of fecal 16S rRNA reveals active *Escherichia coli* in the microbiota of patients with ulcerative colitis. *J Clin Microbiol* 44: 3172–3177. doi:10.1128/JCM.02600-05.
70. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, et al. (2012) Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 142: 46–54.e42; quiz e30. doi:10.1053/j.gastro.2011.10.001.
71. Lashner BA (1995) Epidemiology of inflammatory bowel disease. *Gastroenterol Clin North Am* 24: 467–474.
72. Hou JK, Abraham B, El-Serag H (2011) Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. *Am J Gastroenterol* 106: 563–573. doi:10.1038/ajg.2011.44.
73. O’Mahony SM, Marchesi JR, Scully P, Codling C, Ceolho A-M, et al. (2009) Early life stress alters behavior, immunity, and microbiota in rats: implications for irritable bowel syndrome and psychiatric illnesses. *Biol Psychiatry* 65: 263–267. doi:10.1016/j.biopsych.2008.06.026.
74. Khor B, Gardet A, Xavier RJ (2011) Genetics and pathogenesis of inflammatory bowel disease. *Nature* 474: 307–317. doi:10.1038/nature10209.
75. Smith AM, Rahman FZ, Hayee B, Graham SJ, Marks DJB, et al. (2009) Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn’s disease. *J Exp Med* 206: 1883–1897. doi:10.1084/jem.20091233.
76. Shanahan F (2002) Crohn’s disease. *Lancet* 359: 62–69. doi:10.1016/S0140-6736(02)07284-7.
77. Steidler L, Hans W, Schotte L, Neiryneck S, Obermeier F, et al. (2000) Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. *Science* 289: 1352–1355.
78. Van Wijngaarden P, Meijssen MA (2001) Tuberculous pleurisy: an unusual complication during treatment of Crohn disease with azathioprine. *Scand J Gastroenterol* 36: 1004–1007.

79. Sandhu BK, Fell JME, Beattie RM, Mitton SG, Wilson DC, et al. (2010) Guidelines for the management of inflammatory bowel disease in children in the United Kingdom. *J Pediatr Gastroenterol Nutr* 50 Suppl 1: S1–13. doi:10.1097/MPG.0b013e3181c92c53.
80. Caprilli R, Gassull MA, Escher JC, Moser G, Munkholm P, et al. (2006) European evidence based consensus on the diagnosis and management of Crohn's disease: special situations. *Gut* 55 Suppl 1: i36–58. doi:10.1136/gut.2005.081950c.
81. Takagi S, Utsunomiya K, Kuriyama S, Yokoyama H, Takahashi S, et al. (2006) Effectiveness of an “half elemental diet” as maintenance therapy for Crohn's disease: A randomized-controlled trial. *Aliment Pharmacol Ther* 24: 1333–1340. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17059514>. Accessed 20 August 2013.
82. Zachos M, Tondeur M, Griffiths AM (2007) Enteral nutritional therapy for induction of remission in Crohn's disease. *Cochrane database Syst Rev*: CD000542. doi:10.1002/14651858.CD000542.pub2.
83. Leach ST, Mitchell HM, Eng WR, Zhang L, Day AS (2008) Sustained modulation of intestinal bacteria by exclusive enteral nutrition used to treat children with Crohn's disease. *Aliment Pharmacol Ther* 28: 724–733.
84. Lionetti P, Callegari ML, Ferrari S, Cavicchi MC, Pozzi E, et al. (n.d.) Enteral nutrition and microflora in pediatric Crohn's disease. *JPEN J Parenter Enteral Nutr* 29: S173–5; discussion S175–8, S184–8.
85. Kruis W, Schütz E, Fric P, Fixa B, Judmaier G, et al. (1997) Double-blind comparison of an oral *Escherichia coli* preparation and mesalazine in maintaining remission of ulcerative colitis. *Aliment Pharmacol Ther* 11: 853–858.
86. Rembacken BJ, Snelling AM, Hawkey PM, Chalmers DM, Axon AT (1999) Non-pathogenic *Escherichia coli* versus mesalazine for the treatment of ulcerative colitis: a randomised trial. *Lancet* 354: 635–639.
87. Fitzpatrick LR, Hertzog KL, Quatse AL, Koltun WA, Small JS, et al. (2007) Effects of the probiotic formulation VSL#3 on colitis in weanling rats. *J Pediatr Gastroenterol Nutr* 44: 561–570.
88. Gionchetti P, Rizzello F, Venturi A, Brigidi P, Matteuzzi D, et al. (2000) Oral bacteriotherapy as maintenance treatment in patients with chronic pouchitis: a double-blind, placebo-controlled trial. *Gastroenterology* 119: 305–309.
89. Gionchetti P, Rizzello F, Helwig U, Venturi A, Lammers KM, et al. (2003) Prophylaxis of pouchitis onset with probiotic therapy: a double-blind, placebo-controlled trial.
90. Malin M, Suomalainen H, Saxelin M, Isolauri E (1996) Promotion of IgA immune response in patients with Crohn's disease by oral bacteriotherapy with *Lactobacillus GG*. *Ann Nutr Metab* 40: 137–145.
91. Kaila M, Isolauri E, Soppi E, Virtanen E, Laine S, et al. (1992) Enhancement of the circulating antibody secreting cell response in human diarrhea by a human

- Lactobacillus strain. *Pediatr Res* 32: 141–144. doi:10.1203/00006450-199208000-00002.
92. Gibson GR, Roberfroid MB (1995) Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J Nutr* 125: 1401–1412.
  93. Mazmanian SK, Round JL, Kasper DL (2008) A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* 453: 620–625. doi:10.1038/nature07008.
  94. Kelly D, Campbell JI, King TP, Grant G, Jansson EA, et al. (2004) Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA. *Nat Immunol* 5: 104–112. doi:10.1038/ni1018.
  95. Choi J-M, Bothwell ALM (2012) The nuclear receptor PPARs as important regulators of T-cell functions and autoimmune diseases. *Mol Cells* 33: 217–222. doi:10.1007/s10059-012-2297-y.
  96. Are A, Aronsson L, Wang S, Greicius G, Lee YK, et al. (2008) Enterococcus faecalis from newborn babies regulate endogenous PPARgamma activity and IL-10 levels in colonic epithelial cells. *Proc Natl Acad Sci U S A* 105: 1943–1948. doi:10.1073/pnas.0711734105.
  97. Waidmann M, Bechtold O, Frick J-S, Lehr H-A, Schubert S, et al. (2003) Bacteroides vulgatus protects against Escherichia coli-induced colitis in gnotobiotic interleukin-2-deficient mice. *Gastroenterology* 125: 162–177.
  98. Kamada N, Seo S-U, Chen GY, Núñez G (2013) Role of the gut microbiota in immunity and inflammatory disease. *Nat Rev Immunol* 13: 321–335. doi:10.1038/nri3430.
  99. Maynard CL, Elson CO, Hatton RD, Weaver CT (2012) Reciprocal interactions of the intestinal microbiota and immune system. *Nature* 489: 231–241. doi:10.1038/nature11551.
  100. Gourbeyre P, Denery S, Bodinier M (2011) Probiotics, prebiotics, and synbiotics: impact on the gut immune system and allergic reactions. *J Leukoc Biol* 89: 685–695. doi:10.1189/jlb.1109753.
  101. Metchnikoff E, Metchnikoff II (1908) *The Prolongation of Life: Optimistic Studies* (Google eBook). Springer Publishing Company.
  102. Tissier H (1900) *Recherches sur la flore intestinale des nourrissons: état normal et pathologique*.
  103. Neu J, Caicedo R (2005) Probiotics: protecting the intestinal ecosystem? *J Pediatr* 147: 143–146. doi:10.1016/j.jpeds.2005.05.033.
  104. Stiles ME, Holzapfel WH (1997) Lactic acid bacteria of foods and their current taxonomy. *Int J Food Microbiol* 36: 1–29.

105. Claesson MJ, van Sinderen D, O'Toole PW (2008) Lactobacillus phylogenomics--towards a reclassification of the genus. *Int J Syst Evol Microbiol* 58: 2945–2954. doi:10.1099/ijs.0.65848-0.
106. Guarner F, Malagelada J-R (2003) Gut flora in health and disease. *Lancet* 361: 512–519. doi:10.1016/S0140-6736(03)12489-0.
107. Trebichavsky I, Rada V, Splichalova A, Splichal I (2009) Cross-talk of human gut with bifidobacteria. *Nutr Rev* 67: 77–82. doi:10.1111/j.1753-4887.2008.00141.x.
108. Delgado S, O'Sullivan E, Fitzgerald G, Mayo B (2008) In vitro evaluation of the probiotic properties of human intestinal Bifidobacterium species and selection of new probiotic candidates. *J Appl Microbiol* 104: 1119–1127. doi:10.1111/j.1365-2672.2007.03642.x.
109. Kerckhoffs APM, Samsom M, van der Rest ME, de Vogel J, Knol J, et al. (2009) Lower Bifidobacteria counts in both duodenal mucosa-associated and fecal microbiota in irritable bowel syndrome patients. *World J Gastroenterol* 15: 2887–2892.
110. Martins FS, Silva AA, Vieira AT, Barbosa FHF, Arantes RME, et al. (2009) Comparative study of Bifidobacterium animalis, Escherichia coli, Lactobacillus casei and Saccharomyces boulardii probiotic properties. *Arch Microbiol* 191: 623–630. doi:10.1007/s00203-009-0491-x.
111. Klijn A, Mercenier A, Arigoni F (2005) Lessons from the genomes of bifidobacteria. *FEMS Microbiol Rev* 29: 491–509. doi:10.1016/j.femsre.2005.04.010.
112. Gueimonde M, Garrigues C, van Sinderen D, de los Reyes-Gavilán CG, Margolles A (2009) Bile-inducible efflux transporter from Bifidobacterium longum NCC2705, conferring bile resistance. *Appl Environ Microbiol* 75: 3153–3160. doi:10.1128/AEM.00172-09.
113. Alp G, Aslim B (2010) Relationship between the resistance to bile salts and low pH with exopolysaccharide (EPS) production of Bifidobacterium spp. isolated from infants feces and breast milk. *Anaerobe* 16: 101–105. doi:10.1016/j.anaerobe.2009.06.006.
114. Bron PA, Marco M, Hoffer SM, Van Mullekom E, de Vos WM, et al. (2004) Genetic characterization of the bile salt response in Lactobacillus plantarum and analysis of responsive promoters in vitro and in situ in the gastrointestinal tract. *J Bacteriol* 186: 7829–7835. doi:10.1128/JB.186.23.7829-7835.2004.
115. Sheehan VM, Sleator RD, Hill C, Fitzgerald GF (2007) Improving gastric transit, gastrointestinal persistence and therapeutic efficacy of the probiotic strain Bifidobacterium breve UCC2003. *Microbiology* 153: 3563–3571. doi:10.1099/mic.0.2007/006510-0.
116. Favier C, Neut C, Mizon C, Cortot A, Colombel JF, et al. (1997) Fecal beta-D-galactosidase production and Bifidobacteria are decreased in Crohn's disease. *Dig Dis Sci* 42: 817–822.

117. Venturi A, Gionchetti P, Rizzello F, Johansson R, Zucconi E, et al. (1999) Impact on the composition of the faecal flora by a new probiotic preparation: preliminary data on maintenance treatment of patients with ulcerative colitis. *Aliment Pharmacol Ther* 13: 1103–1108.
118. Gosselink MP, Schouten WR, van Lieshout LMC, Hop WCJ, Laman JD, et al. (2004) Delay of the first onset of pouchitis by oral intake of the probiotic strain *Lactobacillus rhamnosus* GG. *Dis Colon Rectum* 47: 876–884. doi:10.1007/s10350-004-0525-z.
119. Kleerebezem M, Vaughan EE (2009) Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity. *Annu Rev Microbiol* 63: 269–290. doi:10.1146/annurev.micro.091208.073341.
120. Candela M, Perna F, Carnevali P, Vitali B, Ciati R, et al. (2008) Interaction of probiotic *Lactobacillus* and *Bifidobacterium* strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production. *Int J Food Microbiol* 125: 286–292. doi:10.1016/j.ijfoodmicro.2008.04.012.
121. Lan J-G, Cruickshank S-M, Singh J-C-I, Farrar M, Lodge J-P-A, et al. (2005) Different cytokine response of primary colonic epithelial cells to commensal bacteria. *World J Gastroenterol* 11: 3375–3384.
122. García-Urkiá N, Asensio AB, Zubillaga Azpiroz I, Zubillaga Huici P, Vidales C, et al. (2002) [Beneficial effects of *Bifidobacterium lactis* in the prevention of bacterial translocation in experimental short bowel syndrome]. *Cir Pediatr* 15: 162–165.
123. Chen C-C, Louie S, Shi HN, Walker WA (2005) Preinoculation with the probiotic *Lactobacillus acidophilus* early in life effectively inhibits murine *Citrobacter rodentium* colitis. *Pediatr Res* 58: 1185–1191. doi:10.1203/01.pdr.0000183660.39116.83.
124. Helwig U, Lammers KM, Rizzello F, Brigidi P, Rohleder V, et al. (2006) Lactobacilli, bifidobacteria and *E. coli* nissle induce pro- and anti-inflammatory cytokines in peripheral blood mononuclear cells. *World J Gastroenterol* 12: 5978–5986.
125. Imaoka A, Shima T, Kato K, Mizuno S, Uehara T, et al. (2008) Anti-inflammatory activity of probiotic *Bifidobacterium*: enhancement of IL-10 production in peripheral blood mononuclear cells from ulcerative colitis patients and inhibition of IL-8 secretion in HT-29 cells. *World J Gastroenterol* 14: 2511–2516.
126. Matsumoto S, Hara T, Hori T, Mitsuyama K, Nagaoka M, et al. (2005) Probiotic *Lactobacillus*-induced improvement in murine chronic inflammatory bowel disease is associated with the down-regulation of pro-inflammatory cytokines in lamina propria mononuclear cells. *Clin Exp Immunol* 140: 417–426. doi:10.1111/j.1365-2249.2005.02790.x.
127. Van Baarlen P, Wells JM, Kleerebezem M (2013) Regulation of intestinal homeostasis and immunity with probiotic lactobacilli. *Trends Immunol* 34: 208–215. doi:10.1016/j.it.2013.01.005.

128. Barinov A, Loux V, Hammani A, Nicolas P, Langella P, et al. (2009) Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. *Proteomics* 9: 61–73. doi:10.1002/pmic.200800195.
129. Schell MA, Karmirantzou M, Snel B, Vilanova D, Berger B, et al. (2002) The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* 99: 14422–14427. doi:10.1073/pnas.212527599.
130. Heuvelin E, Lebreton C, Grangette C, Pot B, Cerf-Bensussan N, et al. (2009) Mechanisms involved in alleviation of intestinal inflammation by *bifidobacterium breve* soluble factors. *PLoS One* 4: e5184. doi:10.1371/journal.pone.0005184.
131. Jijon H, Backer J, Diaz H, Yeung H, Thiel D, et al. (2004) DNA from probiotic bacteria modulates murine and human epithelial and immune function. *Gastroenterology* 126: 1358–1373.
132. Yan F, Cao H, Cover TL, Whitehead R, Washington MK, et al. (2007) Soluble proteins produced by probiotic bacteria regulate intestinal epithelial cell survival and growth. *Gastroenterology* 132: 562–575. doi:10.1053/j.gastro.2006.11.022.
133. Yan F, Cao H, Cover TL, Washington MK, Shi Y, et al. (2011) Colon-specific delivery of a probiotic-derived soluble protein ameliorates intestinal inflammation in mice through an EGFR-dependent mechanism. *J Clin Invest* 121: 2242–2253. doi:10.1172/JCI44031.
134. Claes IJJ, Schoofs G, Regulski K, Courtin P, Chapot-Chartier M-P, et al. (2012) Genetic and biochemical characterization of the cell wall hydrolase activity of the major secreted protein of *Lactobacillus rhamnosus* GG. *PLoS One* 7: e31588. doi:10.1371/journal.pone.0031588.
135. Konstantinov SR, Smidt H, de Vos WM, Bruijns SCM, Singh SK, et al. (2008) S layer protein A of *Lactobacillus acidophilus* NCFM regulates immature dendritic cell and T cell functions. *Proc Natl Acad Sci U S A* 105: 19474–19479. doi:10.1073/pnas.0810305105.
136. Macho Fernandez E, Fernandez EM, Valenti V, Rockel C, Hermann C, et al. (2011) Anti-inflammatory capacity of selected lactobacilli in experimental colitis is driven by NOD2-mediated recognition of a specific peptidoglycan-derived muropeptide. *Gut* 60: 1050–1059. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21471573>. Accessed 24 November 2013.
137. Kaci G, Lakhdari O, Doré J, Ehrlich SD, Renault P, et al. (2011) Inhibition of the NF- $\kappa$ B pathway in human intestinal epithelial cells by commensal *Streptococcus salivarius*. *Appl Environ Microbiol* 77: 4681–4684. doi:10.1128/AEM.03021-10.
138. Collier-Hyams LS, Neish AS (2005) Innate immune relationship between commensal flora and the mammalian intestinal epithelium. *Cell Mol Life Sci* 62: 1339–1348. doi:10.1007/s00018-005-5038-y.

139. Tien M-T, Girardin SE, Regnault B, Le Bourhis L, Dillies M-A, et al. (2006) Anti-inflammatory effect of *Lactobacillus casei* on *Shigella*-infected human intestinal epithelial cells. *J Immunol* 176: 1228–1237.
140. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, et al. (2011) Enterotypes of the human gut microbiome. *Nature* 473: 174–180. doi:10.1038/nature09944.
141. Pinget M (2008) [Epidemiology of type 2 diabetes, is it a problem of bad nutrition or malnutrition?]. *Soins*: S4–8.
142. Van Baak MA (2013) Nutrition as a link between obesity and cardiovascular disease: how can we stop the obesity epidemic? *Thromb Haemost* 110. doi:10.1160/TH13-01-0045.
143. Chapman-Kiddell CA, Davies PSW, Gillen L, Radford-Smith GL (2010) Role of diet in the development of inflammatory bowel disease. *Inflamm Bowel Dis* 16: 137–151. doi:10.1002/ibd.20968.
144. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO (2007) Development of the human infant intestinal microbiota. *PLoS Biol* 5: e177. doi:10.1371/journal.pbio.0050177.
145. Schwartz S, Friedberg I, Ivanov I V, Davidson LA, Goldsby JS, et al. (2012) A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. *Genome Biol* 13: r32. doi:10.1186/gb-2012-13-4-r32.
146. Roger LC, McCartney AL (2010) Longitudinal investigation of the faecal microbiota of healthy full-term infants using fluorescence in situ hybridization and denaturing gradient gel electrophoresis. *Microbiology* 156: 3317–3328. doi:10.1099/mic.0.041913-0.
147. Fallani M, Amarri S, Uusijarvi A, Adam R, Khanna S, et al. (2011) Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology* 157: 1385–1392. doi:10.1099/mic.0.042143-0.
148. Fallani M, Young D, Scott J, Norin E, Amarri S, et al. (2010) Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *J Pediatr Gastroenterol Nutr* 51: 77–84. doi:10.1097/MPG.0b013e3181d1b11e.
149. Kunz C, Rudloff S, Baier W, Klein N, Strobel S (2000) Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu Rev Nutr* 20: 699–722. doi:10.1146/annurev.nutr.20.1.699.
150. Thurl S, Muller-Werner B, Sawatzki G (1996) Quantification of individual oligosaccharide compounds from human milk using high-pH anion-exchange chromatography. *Anal Biochem* 235: 202–206. doi:10.1006/abio.1996.0113.

151. Morrow AL, Ruiz-Palacios GM, Altaye M, Jiang X, Guerrero ML, et al. (2004) Human milk oligosaccharides are associated with protection against diarrhea in breast-fed infants. *J Pediatr* 145: 297–303. doi:10.1016/j.jpeds.2004.04.054.
152. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, et al. (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334: 105–108. doi:10.1126/science.1208344.
153. Miller TL, Wolin MJ (1996) Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. *Appl Environ Microbiol* 62: 1589–1592.
154. Langlands SJ, Hopkins MJ, Coleman N, Cummings JH (2004) Prebiotic carbohydrates modify the mucosa associated microflora of the human large bowel. *Gut* 53: 1610–1616. doi:10.1136/gut.2003.037580.
155. Videla S, Vilaseca J, Antolín M, García-Lafuente A, Guarner F, et al. (2001) Dietary inulin improves distal colitis induced by dextran sodium sulfate in the rat. *Am J Gastroenterol* 96: 1486–1493. doi:10.1111/j.1572-0241.2001.03802.x.
156. Winkler J, Butler R, Symonds E (2007) Fructo-oligosaccharide reduces inflammation in a dextran sodium sulphate mouse model of colitis. *Dig Dis Sci* 52: 52–58. doi:10.1007/s10620-006-9224-z.
157. Cherbut C, Michel C, Lecannu G (2003) The prebiotic characteristics of fructooligosaccharides are necessary for reduction of TNBS-induced colitis in rats. *J Nutr* 133: 21–27.
158. Camuesco D, Peran L, Comalada M, Nieto A, Di Stasi LC, et al. (2005) Preventative effects of lactulose in the trinitrobenzenesulphonic acid model of rat colitis. *Inflamm Bowel Dis* 11: 265–271.
159. Casellas F, Borruel N, Torrejón A, Varela E, Antolin M, et al. (2007) Oral oligofructose-enriched inulin supplementation in acute ulcerative colitis is well tolerated and associated with lowered faecal calprotectin. *Aliment Pharmacol Ther* 25: 1061–1067. doi:10.1111/j.1365-2036.2007.03288.x.
160. Lindsay J, Whelan K, Stagg A, Gobin P, Al-Hassi H, et al. (2006) Clinical, microbiological, and immunological effects of fructo-oligosaccharide in patients with Crohn's disease. *Gut* 55: 348–355.
161. Kanauchi O, Suga T, Tochiara M, Hibi T, Naganuma M, et al. (2002) Treatment of ulcerative colitis by feeding with germinated barley foodstuff: first report of a multicenter open control trial. *J Gastroenterol* 37 Suppl 1: 67–72.
162. Fox PF, editor (1993) *Cheese: Chemistry, Physics and Microbiology*. Boston, MA: Springer US. doi:10.1007/978-1-4615-2650-6.
163. Pasteur L (1995) *Mémoire sur la fermentation appelée lactique (Extrait par l'auteur)*. *Mol Med* 1: 599.

164. Lister J (1873) a further contribution to the natural history of bacteria and the germ theory of fermentative changes. *Q Microbiol Sci* 13: 380–408.
165. Bibel DJ (1988) Elie Metchnikoff's bacillus of long life. *ASM News* 54: 661–665.
166. Prescott, L. M., Harley, J. P., Klein DA (1999) *Microbiology*, 4th edn.
167. Van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, et al. (2006) The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc Natl Acad Sci U S A* 103: 9274–9279. doi:10.1073/pnas.0603024103.
168. Bourlioux P, Pochart P (1988) Nutritional and health properties of yogurt. *World Rev Nutr Diet* 56: 217–258.
169. Elli M, Callegari ML, Ferrari S, Bessi E, Cattivelli D, et al. (2006) Survival of yogurt bacteria in the human gut. *Appl Environ Microbiol* 72: 5113–5117. doi:10.1128/AEM.02950-05.
170. Shahani KM, Chandan RC (1979) Nutritional and healthful aspects of cultured and culture-containing dairy foods. *J Dairy Sci* 62: 1685–1694. doi:10.3168/jds.S0022-0302(79)83481-5.
171. Rašić J (1978) *Yoghurt : scientific grounds, technology, manufacture and preparations*. [S.l.] ;Copenhagen: Raišić and Kurmann ;;Distributed by Technical Dairy Pub. House.
172. Lactose digestion from unmodified, low- - PubMed Mobile (n.d.).
173. Vesa TH, Marteau P, Korpela R (2000) Lactose intolerance. *J Am Coll Nutr* 19: 165S–175S.
174. Kolars JC, Levitt MD, Aouji M, Savaiano DA (1984) Yogurt--an autodigesting source of lactose. *N Engl J Med* 310: 1–3. doi:10.1056/NEJM198401053100101.
175. Jiang J, Wolk A, Vessby B (1999) Relation between the intake of milk fat and the occurrence of conjugated linoleic acid in human adipose tissue. *Am J Clin Nutr* 70: 21–27.
176. Park Y, McGuire MK, Behr R, McGuire MA, Evans MA, et al. (1999) High-fat dairy product consumption increases delta 9c,11t-18:2 (rumenic acid) and total lipid concentrations of human milk. *Lipids* 34: 543–549.
177. Whigham LD, Cook ME, Atkinson RL (2000) Conjugated linoleic acid: implications for human health. *Pharmacol Res* 42: 503–510. doi:10.1006/phrs.2000.0735.
178. Kemp MQ, Jeffy BD, Romagnolo DF (2003) Conjugated linoleic acid inhibits cell proliferation through a p53-dependent mechanism: effects on the expression of G1-restriction points in breast and colon cancer cells. *J Nutr* 133: 3670–3677.

179. Buttriss J (1997) Nutritional properties of fermented milk products. *Int J Dairy Technol* 50: 21–27. doi:10.1111/j.1471-0307.1997.tb01731.x.
180. Kneifel W, Kaufmann M, Fleischer A, Ulberth F (1992) Screening of Commercially Available Mesophilic Dairy Starter Cultures: Biochemical, Sensory, and Microbiological Properties. *J Dairy Sci* 75: 3158–3166. doi:10.3168/jds.S0022-0302(92)78080-1.
181. Crittenden RG, Martinez NR, Playne MJ (2003) Synthesis and utilisation of folate by yoghurt starter cultures and probiotic bacteria. *Int J Food Microbiol* 80: 217–222.
182. WIGERTZ K, SVENSSON UK, JÄGERSTAD M (1997) Folate and folate-binding protein content in dairy products. *J Dairy Res* 64: 239–252. doi:10.1017/S002202999700215X.
183. King T (1996) Comparison of symptoms after the consumption of milk or lactose-hydrolysed milk by people with self-reported severe lactose intolerance. *Clin Nutr* 15: 97–98.
184. Martini MC, Lerebours EC, Lin WJ, Harlander SK, Berrada NM, et al. (1991) Strains and species of lactic acid bacteria in fermented milks (yogurts): effect on in vivo lactose digestion. *Am J Clin Nutr* 54: 1041–1046.
185. Lerebours E, N'Djitoyap Ndam C, Lavoine A, Hellot MF, Antoine JM, et al. (1989) Yogurt and fermented-then-pasteurized milk: effects of short-term and long-term ingestion on lactose absorption and mucosal lactase activity in lactase-deficient subjects. *Am J Clin Nutr* 49: 823–827.
186. Thoreux K, Balas D, Bouley C, Senegas-Balas F (n.d.) Diet supplemented with yoghurt or milk fermented by *Lactobacillus casei* DN-114 001 stimulates growth and brush-border enzyme activities in mouse small intestine. *Digestion* 59: 349–359.
187. Beshkova DM, Simova ED, Frengova GI, Simov ZI, Adilov EF (1998) Production of amino acids by yogurt bacteria. *Biotechnol Prog* 14: 963–965. doi:10.1021/bp980082j.
188. Conway PL, Gorbach SL, Goldin BR (1987) Survival of lactic acid bacteria in the human stomach and adhesion to intestinal cells. *J Dairy Sci* 70: 1–12. doi:10.3168/jds.S0022-0302(87)79974-3.
189. Pedrosa MC, Golner BB, Goldin BR, Barakat S, Dallal GE, et al. (1995) Survival of yogurt-containing organisms and *Lactobacillus gasseri* (ADH) and their effect on bacterial enzyme activity in the gastrointestinal tract of healthy and hypochlorhydric elderly subjects. *Am J Clin Nutr* 61: 353–359.
190. Bouhnik Y, Pochart P, Marteau P, Arlet G, Goderel I, et al. (1992) Fecal recovery in humans of viable *Bifidobacterium* sp ingested in fermented milk. *Gastroenterology* 102: 875–878.

191. Maragkoudakis PA, Zoumpopoulou G, Miaris C, Kalantzopoulos G, Pot B, et al. (2006) Probiotic potential of *Lactobacillus* strains isolated from dairy products. *Int Dairy J* 16: 189–199.
192. Liu Y-W, Su Y-W, Ong W-K, Cheng T-H, Tsai Y-C (2011) Oral administration of *Lactobacillus plantarum* K68 ameliorates DSS-induced ulcerative colitis in BALB/c mice via the anti-inflammatory and immunomodulatory activities. *Int Immunopharmacol* 11: 2159–2166. doi:10.1016/j.intimp.2011.09.013.
193. Schultz M, Veltkamp C, Dieleman LA, Grenther WB, Wyrick PB, et al. (2002) *Lactobacillus plantarum* 299V in the treatment and prevention of spontaneous colitis in interleukin-10-deficient mice. *Inflamm Bowel Dis* 8: 71–80.
194. Naruszewicz M, Johansson M-L, Zapolska-Downar D, Bukowska H (2002) Effect of *Lactobacillus plantarum* 299v on cardiovascular disease risk factors in smokers. *Am J Clin Nutr* 76: 1249–1255.
195. Duary RK, Bhausahab MA, Batish VK, Grover S (2012) Anti-inflammatory and immunomodulatory efficacy of indigenous probiotic *Lactobacillus plantarum* Lp91 in colitis mouse model. *Mol Biol Rep* 39: 4765–4775. doi:10.1007/s11033-011-1269-1.
196. Xia Y, Chen H-Q, Zhang M, Jiang Y-Q, Hang X-M, et al. (2011) Effect of *Lactobacillus plantarum* LP-Only on gut flora and colitis in interleukin-10 knockout mice. *J Gastroenterol Hepatol* 26: 405–411. doi:10.1111/j.1440-1746.2010.06498.x.
197. Taverniti V, Guglielmetti S (2012) Health-Promoting Properties of *Lactobacillus helveticus*. *Front Microbiol* 3: 392. doi:10.3389/fmicb.2012.00392.
198. Latvala S, Miettinen M, Kekkonen RA, Korpela R, Julkunen I (2011) *Lactobacillus rhamnosus* GG and *Streptococcus thermophilus* induce suppressor of cytokine signalling 3 (SOCS3) gene expression directly and indirectly via interleukin-10 in human primary macrophages. *Clin Exp Immunol* 165: 94–103. doi:10.1111/j.1365-2249.2011.04408.x.
199. Rodríguez C, Medici M, Rodríguez A V, Mozzi F, Font de Valdez G (2009) Prevention of chronic gastritis by fermented milks made with exopolysaccharide-producing *Streptococcus thermophilus* strains. *J Dairy Sci* 92: 2423–2434. doi:10.3168/jds.2008-1724.
200. Foligné B, Deutsch S-M, Breton J, Cousin FJ, Dewulf J, et al. (2010) Promising immunomodulatory effects of selected strains of dairy propionibacteria as evidenced in vitro and in vivo. *Appl Environ Microbiol* 76: 8259–8264. doi:10.1128/AEM.01976-10.
201. Thierry A, Deutsch S-M, Falentin H, Dalmaso M, Cousin FJ, et al. (2011) New insights into physiology and metabolism of *Propionibacterium freudenreichii*. *Int J Food Microbiol* 149: 19–27. doi:10.1016/j.ijfoodmicro.2011.04.026.
202. Okada Y, Tsuzuki Y, Narimatsu K, Sato H, Ueda T, et al. (2013) 1,4-Dihydroxy-2-naphthoic acid from *Propionibacterium freudenreichii* reduces inflammation in

- interleukin-10-deficient mice with colitis by suppressing macrophage-derived proinflammatory cytokines. *J Leukoc Biol* 94: 473–480. doi:10.1189/jlb.0212104.
203. Okada Y, Tsuzuki Y, Miyazaki J, Matsuzaki K, Hokari R, et al. (2006) *Propionibacterium freudenreichii* component 1,4-dihydroxy-2-naphthoic acid (DHNA) attenuates dextran sodium sulphate induced colitis by modulation of bacterial flora and lymphocyte homing. *Gut* 55: 681–688. doi:10.1136/gut.2005.070490.
  204. Oozeer R, Goupil-Feuillerat N, Alpert CA, van de Guchte M, Anba J, et al. (2002) *Lactobacillus casei* is able to survive and initiate protein synthesis during its transit in the digestive tract of human flora-associated mice. *Appl Environ Microbiol* 68: 3570–3574.
  205. Perdígón G, Vintiñi E, Alvarez S, Medina M, Medici M (1999) Study of the possible mechanisms involved in the mucosal immune system activation by lactic acid bacteria. *J Dairy Sci* 82: 1108–1114. doi:10.3168/jds.S0022-0302(99)75333-6.
  206. Borruel N, Carol M, Casellas F, Antolín M, de Lara F, et al. (2002) Increased mucosal tumour necrosis factor alpha production in Crohn's disease can be downregulated ex vivo by probiotic bacteria. *Gut* 51: 659–664.
  207. Borruel N, Casellas F, Antolín M, Llopis M, Carol M, et al. (2003) Effects of nonpathogenic bacteria on cytokine secretion by human intestinal mucosa. *Am J Gastroenterol* 98: 865–870. doi:10.1111/j.1572-0241.2003.07384.x.
  208. Perdígón G, de Macias ME, Alvarez S, Oliver G, de Ruiz Holgado AP (1988) Systemic augmentation of the immune response in mice by feeding fermented milks with *Lactobacillus casei* and *Lactobacillus acidophilus*. *Immunology* 63: 17–23.
  209. Wallace TD, Bradley S, Buckley ND, Green-Johnson JM (2003) Interactions of lactic acid bacteria with human intestinal epithelial cells: effects on cytokine production. *J Food Prot* 66: 466–472.
  210. Ministère de l'agriculture, de l'agroalimentaire et de la forêt - Agreste - La statistique, l'évaluation et la prospective agricole (n.d.).
  211. Sood A, Midha V, Makharia GK, Ahuja V, Singal D, et al. (2009) The probiotic preparation, VSL#3 induces remission in patients with mild-to-moderately active ulcerative colitis. *Clin Gastroenterol Hepatol* 7: 1202–9, 1209.e1. doi:10.1016/j.cgh.2009.07.016.
  212. Tursi A, Brandimarte G, Papa A, Giglio A, Elisei W, et al. (2010) Treatment of relapsing mild-to-moderate ulcerative colitis with the probiotic VSL#3 as adjunctive to a standard pharmaceutical treatment: a double-blind, randomized, placebo-controlled study. *Am J Gastroenterol* 105: 2218–2227. doi:10.1038/ajg.2010.218.
  213. Galdeano CM, de Moreno de LeBlanc A, Vinderola G, Bonet MEB, Perdígón G (2007) Proposed model: mechanisms of immunomodulation induced by probiotic bacteria. *Clin Vaccine Immunol* 14: 485–492. doi:10.1128/CVI.00406-06.

214. Takamura T, Harama D, Fukumoto S, Nakamura Y, Shimokawa N, et al. (2011) *Lactobacillus bulgaricus* OLL1181 activates the aryl hydrocarbon receptor pathway and inhibits colitis. *Immunol Cell Biol* 89: 817–822. doi:10.1038/icb.2010.165.
215. Hao P, Zheng H, Yu Y, Ding G, Gu W, et al. (2011) Complete sequencing and pan-genomic analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* reveal its genetic basis for industrial yogurt production. *PLoS One* 6: e15964. doi:10.1371/journal.pone.0015964.
216. Ben Zakour N, Grimaldi C, Gautier M, Langella P, Azevedo V, et al. (2006) Testing of a whole genome PCR scanning approach to identify genomic variability in four different species of lactic acid bacteria. *Res Microbiol* 157: 386–394. doi:10.1016/j.resmic.2005.09.006.
217. Germond J-E, Lapierre L, Delley M, Mollet B, Felis GE, et al. (2003) Evolution of the bacterial species *Lactobacillus delbrueckii*: a partial genomic study with reflections on prokaryotic species concept. *Mol Biol Evol* 20: 93–104.
218. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, et al. (2006) Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci U S A* 103: 15611–15616. doi:10.1073/pnas.0607117103.
219. Santos Rocha C, Lakhdari O, Blotti re HM, Blugeon S, Sokol H, et al. (2012) Anti-inflammatory properties of dairy lactobacilli. *Inflamm Bowel Dis* 18: 657–666. doi:10.1002/ibd.21834.
220. Santos Rocha C, Gomes-Santos AC, Garcias Moreira T, de Azevedo M, Diniz Luerce T, et al. (2014) Local and Systemic Immune Mechanisms Underlying the Anti-Colitis Effects of the Dairy Bacterium *Lactobacillus delbrueckii*. *PLoS One* 9: e85923. doi:10.1371/journal.pone.0085923.
221. Wilke RA, Mareedu RK, Moore JH (2008) The Pathway Less Traveled: Moving from Candidate Genes to Candidate Pathways in the Analysis of Genome-Wide Data from Large Scale Pharmacogenetic Association Studies. *Curr Pharmacogenomics Person Med* 6: 150–159.
222. Fleischmann R, Adams M, White O, Clayton R, Kirkness E, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* (80-) 269: 496–512. doi:10.1126/science.7542800.
223. Lagesen K, Ussery DW, Wassenaar TM (2010) Genome update: the 1000th genome--a cautionary tale. *Microbiology* 156: 603–608. doi:10.1099/mic.0.038257-0.
224. Callanan M, Kaleta P, O’Callaghan J, O’Sullivan O, Jordan K, et al. (2008) Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion. *J Bacteriol* 190: 727–735. doi:10.1128/JB.01295-07.

225. Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, et al. (2003) Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci U S A* 100: 1990–1995. doi:10.1073/pnas.0337704100.
226. Lambert JM, Siezen RJ, de Vos WM, Kleerebezem M (2008) Improved annotation of conjugated bile acid hydrolase superfamily members in Gram-positive bacteria. *Microbiology* 154: 2492–2500. doi:10.1099/mic.0.2008/016808-0.
227. Siezen R, Boekhorst J, Muscariello L, Molenaar D, Renckens B, et al. (2006) *Lactobacillus plantarum* gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria. *BMC Genomics* 7: 126. doi:10.1186/1471-2164-7-126.
228. Boekhorst J, Wels M, Kleerebezem M, Siezen RJ (2006) The predicted secretome of *Lactobacillus plantarum* WCFS1 sheds light on interactions with its environment. *Microbiology* 152: 3175–3183. doi:10.1099/mic.0.29217-0.
229. Boekhorst J, Helmer Q, Kleerebezem M, Siezen RJ (2006) Comparative analysis of proteins with a mucus-binding domain found exclusively in lactic acid bacteria. *Microbiology* 152: 273–280. doi:10.1099/mic.0.28415-0.
230. Lebeer S, Claes I, Tytgat HLP, Verhoeven TLA, Marien E, et al. (2012) Functional analysis of *Lactobacillus rhamnosus* GG pili in relation to adhesion and immunomodulatory interactions with intestinal epithelial cells. *Appl Environ Microbiol* 78: 185–193. doi:10.1128/AEM.06192-11.
231. Tripathi P, Beaussart A, Alsteens D, Dupres V, Claes I, et al. (2013) Adhesion and nanomechanics of pili from the probiotic *Lactobacillus rhamnosus* GG. *ACS Nano* 7: 3685–3697. doi:10.1021/nn400705u.
232. Pridmore RD, Berger B, Desiere F, Vilanova D, Barretto C, et al. (2004) The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc Natl Acad Sci U S A* 101: 2512–2517.
233. Denou E, Pridmore RD, Berger B, Panoff J-M, Arigoni F, et al. (2008) Identification of genes associated with the long-gut-persistence phenotype of the probiotic *Lactobacillus johnsonii* strain NCC533 using a combination of genomics and transcriptome analysis. *J Bacteriol* 190: 3161–3168. doi:10.1128/JB.01637-07.
234. Denou E, Berger B, Barretto C, Panoff J-M, Arigoni F, et al. (2007) Gene expression of commensal *Lactobacillus johnsonii* strain NCC533 during in vitro growth and in the murine gut. *J Bacteriol* 189: 8109–8119. doi:10.1128/JB.00991-07.
235. Marco ML, Peters THF, Bongers RS, Molenaar D, van Hemert S, et al. (2009) Lifestyle of *Lactobacillus plantarum* in the mouse caecum. *Environ Microbiol* 11: 2747–2757. doi:10.1111/j.1462-2920.2009.02001.x.
236. Marco ML, de Vries MC, Wels M, Molenaar D, Mangell P, et al. (2010) Convergence in probiotic *Lactobacillus* gut-adaptive responses in humans and mice. *ISME J* 4: 1481–1484. doi:10.1038/ismej.2010.61.

237. Siciliano RA, Mazzeo MF (2012) Molecular mechanisms of probiotic action: a proteomic perspective. *Curr Opin Microbiol* 15: 390–396. doi:10.1016/j.mib.2012.03.006.
238. Van de Guchte M, Chaze T, Jan G, Mistou M-Y (2012) Properties of probiotic bacteria explored by proteomic approaches. *Curr Opin Microbiol* 15: 381–389. doi:10.1016/j.mib.2012.04.003.
239. Sanchez B, Ruiz L, de los Reyes-Gavilan CG, Margolles A (2008) Proteomics of stress response in *Bifidobacterium*. *Front Biosci* 13: 6905–6919.
240. Ruiz L, Ruas-Madiedo P, Gueimonde M, de Los Reyes-Gavilán CG, Margolles A, et al. (2011) How do bifidobacteria counteract environmental challenges? Mechanisms involved and physiological consequences. *Genes Nutr* 6: 307–318. doi:10.1007/s12263-010-0207-5.
241. Koponen J, Laakso K, Koskenniemi K, Kankainen M, Savijoki K, et al. (2012) Effect of acid stress on protein expression and phosphorylation in *Lactobacillus rhamnosus* GG. *J Proteomics* 75: 1357–1374. doi:10.1016/j.jprot.2011.11.009.
242. Savijoki K, Lietzén N, Kankainen M, Alatossava T, Koskenniemi K, et al. (2011) Comparative proteome cataloging of *Lactobacillus rhamnosus* strains GG and Lc705. *J Proteome Res* 10: 3460–3473. doi:10.1021/pr2000896.
243. Gilad O, Svensson B, Viborg AH, Stuer-Lauridsen B, Jacobsen S (2011) The extracellular proteome of *Bifidobacterium animalis* subsp. *lactis* BB-12 reveals proteins with putative roles in probiotic effects. *Proteomics* 11: 2503–2514. doi:10.1002/pmic.201000716.
244. Tamura 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. doi:10.1093/molbev/msr121.
245. Sun Z, Chen X, Wang J, Zhao W, Shao Y, et al. (2011) Complete genome sequence of *Lactobacillus delbrueckii* subsp. *bulgaricus* strain ND02. *J Bacteriol* 193: 3426–3427. doi:10.1128/JB.05004-11.
246. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* Chapter 2: Unit 2.3. doi:10.1002/0471250953.bi0203s00.
247. Tanigawa K, Watanabe K (2011) Multilocus sequence typing reveals a novel subspeciation of *Lactobacillus delbrueckii*. *Microbiology* 157: 727–738. doi:10.1099/mic.0.043240-0.
248. Dobrijevic D, Di Liberto G, Tanaka K, de Wouters T, Dervyn R, et al. (2013) High-throughput system for the presentation of secreted and surface-exposed proteins from gram-positive bacteria in functional metagenomics studies. *PLoS One* 8: e65956. doi:10.1371/journal.pone.0065956.

249. El Kafsi H, Binesse J, Loux V, Buratti J, Boudebouze S, Dervyn R, Hammani A, Maguin E van de GM (n.d.) Genome sequence of *Lactobacillus delbrueckii* ssp. *lactis* CNRZ327, a dairy bacterium with anti-inflammatory properties. Submitted.
250. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685.
251. Leverrier P, Vissers JPC, Rouault A, Boyaval P, Jan G (2004) Mass spectrometry proteomic analysis of stress adaptation reveals both common and distinct response pathways in *Propionibacterium freudenreichii*. *Arch Microbiol* 181: 215–230. doi:10.1007/s00203-003-0646-0.
252. Le Maréchal C, Jardin J, Jan G, Even S, Pulido C, et al. (2011) *Staphylococcus aureus* seroproteomes discriminate ruminant isolates causing mild or severe mastitis. *Vet Res* 42: 35. doi:10.1186/1297-9716-42-35.
253. Dupierris V, Masselon C, Court M, Kieffer-Jaquinod S, Bruley C (2009) A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. *Bioinformatics* 25: 1980–1981. doi:10.1093/bioinformatics/btp301.
254. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19: 185–193.
255. Canosi U, Morelli G, Trautner TA (1978) The relationship between molecular structure and transformation efficiency of some *S. aureus* plasmids isolated from *B. subtilis*. *Mol Gen Genet* 166: 259–267.
256. Aseev L V, Boni I V (n.d.) [Extraribosomal functions of bacterial ribosomal proteins]. *Mol Biol (Mosk)* 45: 805–816.
257. Henderson B, Martin A (2011) Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect Immun* 79: 3476–3491. doi:10.1128/IAI.00179-11.
258. Maguire M, Coates ARM, Henderson B (2002) Chaperonin 60 unfolds its secrets of cellular communication. *Cell Stress Chaperones* 7: 317–329.
259. Efron B (2007) Size, power and false discovery rates. *Ann Stat* 35: 1351–1377. doi:10.1214/009053606000001460.
260. Turpin W, Humblot C, Noordine M-L, Thomas M, Guyot J-P (2012) Lactobacillaceae and cell adhesion: genomic and functional screening. *PLoS One* 7: e38034. doi:10.1371/journal.pone.0038034.
261. Vélez MP, De Keersmaecker SCJ, Vanderleyden J (2007) Adherence factors of *Lactobacillus* in the human gastrointestinal tract. *FEMS Microbiol Lett* 276: 140–148. doi:10.1111/j.1574-6968.2007.00908.x.

262. Roos S, Jonsson H (2002) A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. *Microbiology* 148: 433–442.
263. Buck BL, Altermann E, Svingerud T, Klaenhammer TR (2005) Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. *Appl Environ Microbiol* 71: 8344–8351. doi:10.1128/AEM.71.12.8344-8351.2005.
264. Du Y, He Y-X, Zhang Z-Y, Yang Y-H, Shi W-W, et al. (2011) Crystal structure of the mucin-binding domain of Spr1345 from *Streptococcus pneumoniae*. *J Struct Biol* 174: 252–257. doi:10.1016/j.jsb.2010.10.016.
265. Granato D, Bergonzelli GE, Pridmore RD, Marvin L, Rouvet M, et al. (2004) Cell surface-associated elongation factor Tu mediates the attachment of *Lactobacillus johnsonii* NCC533 (La1) to human intestinal cells and mucins. *Infect Immun* 72: 2160–2169.
266. Bergonzelli GE, Granato D, Pridmore RD, Marvin-Guy LF, Donnicola D, et al. (2006) GroEL of *Lactobacillus johnsonii* La1 (NCC 533) is cell surface associated: potential role in interactions with the host and the gastric pathogen *Helicobacter pylori*. *Infect Immun* 74: 425–434. doi:10.1128/IAI.74.1.425-434.2006.
267. Weiss N, Schillinger U, Kandler O (1983) *Lactobacillus lactis*, *Lactobacillus leichmannii* and *Lactobacillus bulgaricus*, Subjective Synonyms of *Lactobacillus delbrueckii*, and Description of *Lactobacillus delbrueckii* subsp. *lactis* comb. nov. and *Lactobacillus delbrueckii* subsp. *bulgaricus* comb. nov. *Syst Appl Microbiol* 4: 552–557. doi:10.1016/S0723-2020(83)80012-5.

## **Etudes comparatives de *Lactobacillus delbrueckii* sous-espèces *lactis* et *bulgaricus*.**

### **Identification des déterminants du phénotype anti-inflammatoire.**

Le travail décrit dans cette thèse a commencé avec la découverte d'effets anti-inflammatoires chez certaines souches de *L. delbrueckii*. Il avait été montré que l'effet anti-inflammatoire est souche-dépendant, et implique l'action de protéines exposées à la surface de la bactérie.

Dans le but d'identifier l'effecteur bactérien à l'origine de l'effet immuno-modulateur, 8 souches de *L. delbrueckii* ont été sélectionnées. Deux de ces souches sont à fort effet anti-inflammatoires, et les 6 restantes sont à effet faible ou intermédiaire. Pour l'identification des protéines potentiellement responsables pour l'effet anti-inflammatoire, des études de génomique et transcriptomique comparatives des 8 souches de *L. delbrueckii* ont été entreprises, ainsi qu'une étude comparative du protéome de surface bactérienne.

La première partie de cette thèse décrit les résultats de finition du génome d'une des deux souches hautement anti-inflammatoires. Cette étape a révélé que la partie manquante de la séquence génomique était principalement composée de séquences répétées de type séquences d'insertions (IS), dont le nombre s'avère particulièrement élevé.

La deuxième partie de la thèse décrit une étape de valorisation des données génomiques à travers une étude comparative entre souches de la ssp. *lactis* et souches de la ssp. *bulgaricus*. Cette étude révèle que les deux ssp. de *L. delbrueckii* évoluent en adaptation au milieu lait. Toutefois, la ssp. *bulgaricus* semble avoir atteint un stade d'adaptation plus avancé que celui de la ssp. *lactis*. L'adaptation des deux ssp. à leur environnement se fait principalement par un phénomène de perte spontanée de gènes devenus superflus.

Une étude plus avancée de la structure génomique des deux ssp. révèle deux nouveaux aspects des différences de structure génomique.

Tout d'abord, au sein du core génome des deux sous-espèces, les événements d'échange génétique et recombinaison ont contribué plus à la diversité au sein de la ssp. *lactis* qu'à la diversité chez la ssp. *bulgaricus*.

Ensuite, une structure inversée répétée de grande taille, rarement observée dans les génomes bactériens, s'avère caractéristique de la ssp. *bulgaricus*.

La troisième partie de thèse décrit les trois approches comparatives menées dans le but d'identifier les protéines bactériennes à l'origine de l'effet anti-inflammatoire. Au bout de ces études, nous avons sélectionné 56 gènes candidats, dont 41 ont été clonés dans un système d'expression hétérologue. Pour l'instant, 17 clones d'expression ont été testés *in vitro* pour leur potentiel immuno-modulateur. Les résultats préliminaires ont permis l'identification d'une protéine à effet anti-inflammatoire.

### **Comparative studies of *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus*.**

#### **Identification of anti-inflammatory bacterial effectors.**

The work described in this thesis began with the discovery of anti-inflammatory effects in certain strains of *L. delbrueckii*. The anti-inflammatory effect had been shown to be strain-dependent, and implicate the action of bacterial surface exposed proteins.

In order to identify the bacterial effector responsible for the immunomodulatory effect, eight strains of *L. delbrueckii* were selected. Two of these strains have a strong anti-inflammatory effect, whereas the remaining strains have a weak to intermediary effect. To identify proteins that may be responsible for the anti-inflammatory effect, comparative genomic and transcriptomic studies of the 8 *L. delbrueckii* strains were conducted, as well as a comparative study of the bacterial surface proteome.

The first part of this thesis describes the genome finishing of one of the highly anti-inflammatory strains. This step revealed that the missing part of the genome sequence was mainly composed of repeated sequences such as insertions sequences (IS), that were present in a particularly high number.

The second part of the thesis describes the valorisation of genomic data through a comparative study between ssp. *lactis* strains and ssp. *bulgaricus* strains. This study reveals that both *L. delbrueckii* ssp. are evolving in adaptation to the milk environment. However, the ssp. *bulgaricus* appears to have reached a more advanced stage of adaptation than the ssp. *lactis*. The adaptation of both ssp. to their environment is primarily a phenomenon of spontaneous loss of genes that have become superfluous.

A more detailed study of the genome structure of the two ssp. reveals two important differences. Firstly, within the *L. delbrueckii* core genome, genetic exchange and recombination contributed much more to the ssp. *lactis* diversity than to the ssp. *bulgaricus* diversity. Furthermore, a large inverted repeat structure, rarely observed in bacterial genomes, appears to be characteristic of the ssp. *bulgaricus*.

The third part of the thesis describes the three comparative approaches used to identify bacterial proteins responsible for the anti-inflammatory effect. We selected 56 candidate genes, of which 41 were cloned in a heterologous expression system. So far, 17 expression clones were tested *in vitro* for their immunomodulatory potential. The preliminary results allowed the identification of one protein with anti-inflammatory effects.

