



Analyse intégrative des déterminants de la spécificité organoleptique d'une souche de *Lactococcus. lactis* ssp. *lactis* dans sa fonction de ferment laitier

Amandine Dhaisne

► To cite this version:

Amandine Dhaisne. Analyse intégrative des déterminants de la spécificité organoleptique d'une souche de *Lactococcus. lactis* ssp. *lactis* dans sa fonction de ferment laitier. Microbiologie et Parasitologie. INSA de Toulouse, 2013. Français. NNT : 2013ISAT0052 . tel-04416322

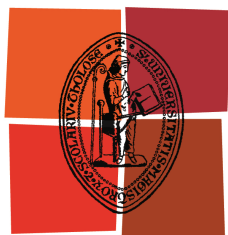
HAL Id: tel-04416322

<https://theses.hal.science/tel-04416322>

Submitted on 25 Jan 2024

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.



Université
de Toulouse

THESE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par :

Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présenté par :

Amandine DHAISNE

Titre :

Analyse intégrative des déterminants de la spécificité organoleptique d'une souche de *Lactococcus lactis* ssp. *lactis* dans sa fonction de levain laitier.

ED SEVAB : Ingénieries microbienne et enzymatique

Unité de recherche :

Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés UMR 5504 CNRS – UMR 792 INRA

Directeur de Thèse :

Pascal LOUBIERE

Rapporteurs :

Gisèle LAPOINTE

Christophe MONNET

Examineur :

Jean GUZZO

RESUME

Lactococcus lactis est une bactérie lactique couramment utilisée dans l'industrie laitière. Elle assure, en tant que ferment, des fonctions technologiques multiples qui impactent la flaveur et la texture finale des produits. Cependant, la diversité fonctionnelle constatée au sein des levains de cette espèce impose de mettre en place un processus de sélection des souches. Ces travaux ont pour objectif d'identifier les déterminants de la spécificité organoleptique dite « crème » de la souche industrielle *L. lactis* ssp. *lactis* (propriété de Yoplait).

Dans un premier temps, un diagnostic macro-cinétique de l'activité de ce ferment a été réalisé en lait pour évaluer l'impact sur la physiologie cellulaire de trois paramètres environnementaux (l'acidification, le stress oxydatif, et la thermisation différentielle du lait). Définir la singularité de notre souche d'intérêt nécessite d'évaluer la diversité fonctionnelle de levains laitiers de *L. lactis* ssp. *lactis*. Cette démarche s'est appuyée sur une approche de biologie intégrative du génotype au phénotype. Pour cela, un panel de neuf souches, représentatif de la diversité génétique de cette sous-population, a été constitué et évalué en lait sur 82 variables incluant des descripteurs physiologiques, la production de métabolites et de composés organiques volatiles (COVs). Cette étude a confirmé la spécificité des signatures phénotypiques de ces souches génétiquement proches. Pour réduire le temps d'expérimentation, une sélection des variables discriminantes a été conduite. Vingt COVs ont été identifiés comme les principaux déterminants de cette variabilité. Ils définissent quatre groupes de souches à l'orientation métabolique divergente : lipolyse, glycolyse, protéolyse et « non exprimée ». L'apport des variables génotypiques a permis d'ajuster la classification fonctionnelle en y intégrant les divergences génétiques non exprimées dans les conditions expérimentales. La présence de marqueurs protéiques de cette spécificité a été également déterminée par analyse comparative des différents protéomes en lait. Après sélection d'un sous-ensemble d'une trentaine de protéines, les trois niveaux (génotype, protéomique et phénotype) responsables de l'expression de cette diversité ont été intégrés pour proposer la première classification prédictive de la diversité fonctionnelle de ferments. En positionnant notre souche d'intérêt au sein de ce modèle, quatorze déterminants phénotypiques ainsi que quinze protéines ont pu être identifiés pour expliquer sa spécificité. L'un des composés clef de cette typicité aromatique a fait l'objet d'une étude approfondie afin de tester les différents paramètres pouvant influencer sa synthèse. La dernière partie, plus applicative, s'est articulée sur la modélisation de la signature en fonction de quatre facteurs industriels (matière grasse, sèche, oxygénation et température) par utilisation de la méthodologie des surfaces de réponse. Elle a abouti à la définition d'un procédé modifié susceptible de renforcer les caractéristiques organoleptiques « crèmeuses » des produits.

Mots clefs : *Lactococcus lactis*, diversité, biologie intégrative, phénotype, génotype, levain laitier.

ABSTRACT

The lactic acid bacterium *Lactococcus lactis* is widely used in the dairy industry. Used as a starter, *L. lactis* subsp. *lactis* is in charge of many functional characteristics which determine the final texture and flavour of fermented products. However, the phenotypic diversity observed within the species requires strain selection development. This work aims at identifying the determinants of the creamy sensory specificity of an industrial strain of *L. lactis* subsp. *lactis* (Yoplait property).

Firstly, a diagnosis of macro-kinetic activity of this starter was carried out to assess the impact on cellular physiology of three environmental parameters (acidification, oxidative stress, and milk heat treatment). In order to assess the uniqueness of our strain of interest, a systems biology approach from genotype to phenotype was implemented to study the functional diversity of *L. lactis* subsp. *lactis* starters. A group of nine strains representative of the genetic diversity of this subpopulation was made up. 82 variables selected as important dairy features; including physiological descriptors, production of metabolites and volatile organic compounds (VOCs); were monitored. This study demonstrated the phenotypic uniqueness of each of these genetically closely related strains, allowing strain discrimination in dairy products. To reduce the time-consuming experimentation, we developed a method of variable selection. Twenty VOCs were therefore identified as major phenotypic determinants of this diversity. They define four robust stain clusters depending on their metabolic orientations: lipolysis, proteolysis, glycolysis and “unexpressed”. Inclusion of genotypic diversity in addition to phenotypic characters led the functional classification to be adjusted by integrating strain unexpressed genotypic potentialities. Comparative proteomics analysis in milk identified protein markers of this specificity. After selecting a subset of forty-six proteins, the three levels (genotype, proteomics and phenotype) involved in this diversity expression were integrated to provide a predictive classification of starter diversity.

The integration of our strain to this model enabled the identification of fourteen phenotypic determinants and seven proteins to explain its specificity. A key aromatic compound of this typicity was subjected to a comprehensive analysis to point out the different factors that may influence its synthesis. On a more applied aspect, the last part was focused on the signature modeling based on four industrial factors (milk fat and total solid level, oxygen and temperature) using the response surface methodology. It enabled the “creamy” organoleptic characteristics of the fermented products to be enhanced.

Keywords: *Lactococcus lactis*, diversity, systems biology, phenotype, genotype, protein, dairy starter.

AVERTISSEMENT

**A la demande de Yoplait France SA, le contenu de ce
manuscrit est strictement confidentiel.**

SOMMAIRE

SOMMAIRE	I
INTRODUCTION	- 1 -
ETUDE BIBLIOGRAPHIQUE	- 7 -
I <i>Lactococcus lactis</i>, bactérie lactique importante pour l'industrie laitière	- 9 -
I.1 Le lait et ses produits de transformations.....	- 9 -
I.2 La fonction technologique des ferments lactiques	- 10 -
I.3 Caractéristiques générales des bactéries lactiques.....	- 12 -
I.4 Phylogénie et diversité de <i>Lactococcus lactis</i>	- 13 -
II <i>Métabolisme de Lactococcus lactis</i>	- 18 -
II.1 Métabolisme carboné	- 18 -
II.2 Nutrition azotée en matrice lait	- 31 -
II.3 Lipolyse et catabolisme des acides gras en lait	- 48 -
II.4 Métabolisme des bases puriques, pyrimidiques et des nucléotides	- 54 -
III <i>Diversité technologique des levains et identification de signatures spécifiques</i>	- 58 -
III.1 Diversité technologique des levains de <i>L. lactis</i>	- 58 -
III.2 Identification de la signature d'une souche	- 69 -
PROBLEMATIQUE	- 82 -
MATERIELS & METHODES.....	- 88 -
RESULTATS	- 118 -
CHAPITRE I : <u>ANALYSE FONCTIONNELLE DU COMPORTEMENT MACRO-CINÉTIQUE DE LA SOUCHE L. LACTIS</u>	
SSP. LACTIS ND5F EN MATRICE LAIT- 120 -	
INTRODUCTION	- 122 -
I <i>Diagnostic du comportement de ND5F dans la condition de référence</i>	- 123 -
I.1 Croissance diphasique et caractérisation de l'activité acidifiante.....	- 123 -
I.1 Cinétiques de consommation d'oxygène et activité réductrice.....	- 127 -
I.2 Analyse macro-cinétique du métabolisme.....	- 129 -
II <i>Impact des facteurs environnementaux</i>	- 136 -
II.1 Activité acidifiante et effet du stress acide sur la physiologie de ND5F.....	- 136 -
II.2 Effet de l'environnement gazeux sur la croissance et l'acidification.....	- 142 -
II.3 Thermisation différentielle de la matrice lait	- 145 -
III <i>Exigences nutritionnelles</i>	- 149 -
CONCLUSION	- 152 -
CHAPITRE II : <u>IDENTIFICATION ET CARACTÉRISATION DE LA SPÉCIFICITÉ DE L. LACTIS SSP. LACTISND5F EN LAIT</u>	
PAR UNE DÉMARCHE INTÉGRATIVE GÉNOTYPE-PHÉNOTYPE.	- 154 -
INTRODUCTION	- 156 -
I <i>Analyse du contenu génétique de ND5F</i>	- 157 -
I.1 Positionnement phylogénétique par MLST de ND5F	- 157 -
I.2 Evaluation génomique de ND5F.....	- 159 -
I.3 Caractérisation génomique de ND5F	- 161 -
II <i>Etude de la diversité de levains par une approche intégrative génotype-phénotype en lait</i>.....	- 168 -
II.1 Résumé article.....	- 168 -
II.2 Publication.....	- 170 -
III <i>Identification et caractérisation de la signature spécifique de ND5F</i>.....	- 171 -
III.1 Positionnement génotypique de ND5F	- 171 -
III.2 Singularité de la signature fonctionnelle de ND5F	- 172 -
III.3 Caractérisation métabolique de la typicité organoleptique de ND5F	- 177 -
III.4 Approche intégrative génotype-phénotype pour affiner la classification fonctionnelle de ND5F	- 184 -
CONCLUSION	- 186 -

CHAPITRE III : _APPORT DE LA DIVERSITÉ PROTÉIQUE DE LEVAINS DE <i>L. LACTIS</i> SSP. <i>LACTIS</i> A LA CLASSIFICATION INTÉGRATIVE GÉNOTYPE-PHÉNOTYPE	- 187 -
INTRODUCTION	- 190 -
I <i>Exploration de la diversité protéique de levains de <i>L. lactis</i> ssp. <i>lactis</i> en lait</i>	- 191 -
I.1 Exploration fonctionnelle des protéomes	- 191 -
I.2 Sélection des protéines discriminantes.....	- 199 -
I.3 Identification des signatures protéiques.....	- 204 -
I.4 Vers une classification intégrée de la diversité des levains.....	- 208 -
II <i>Positionnement protéique de la signature de ND5F.....</i>	- 216 -
II.1 Identification des protéines spécifiques à ND5F	- 216 -
II.2 Positionnement de la signature de ND5F.....	- 224 -
CONCLUSION	- 227 -
CHAPITRE IV : _LEVIERS ET VERROUS DE LA PRODUCTION AROMATIQUE DE <i>L. LACTIS</i> SSP. <i>LACTIS</i> ND5F..	- 229 -
INTRODUCTION	- 230 -
I <i>Approche mécanistique du contrôle de la production de pentane-2,3-dione.....</i>	- 231 -
I.1 Paramètres environnementaux sélectionnés.....	- 231 -
I.2 Analyse comparative du comportement macro-cinétique	- 239 -
I.3 Impact sur la production de la pentane-2,3-dione.....	- 248 -
I.4 Réponses protéomique comparatives	- 251 -
II <i>Modélisation de la singularité de la signature de <i>L. lactis</i>ssp. <i>lactis</i>ND5F en matrice lait</i>	- 258 -
II.1 Présentation de la démarche expérimentale	- 258 -
II.2 Analyse des réponses individuelles et modélisation des surfaces de réponses	- 263 -
II.3 Recherche d'un point de fonctionnement permettant l'optimisation de la réponse globale.....	- 274 -
II.4 Transfert et prédictibilité du point de fonctionnement du modèle à l'échelle pilote	- 282 -
CONCLUSION	- 285 -
CONCLUSION& PERSPECTIVES.....	- 287 -
REFERENCES BIBLIOGRAPHIQUES.....	- 301 -
TABLE DES ILLUSTRATIONS.....	- 334 -
INDEX DES ABREVIATIONS	- 346 -
LISTE DES PUBLICATIONS & COMMUNICATIONS	- 350 -

INTRODUCTION

INTRODUCTION

Employé dans le langage courant, le terme de « biodiversité » est la contraction des mots biologie et diversité, qui désigne la diversité du monde vivant. Elle est estimée à plus de 8,7 millions d'espèces et reflète principalement celle des microorganismes procaryotes, organismes vivants les plus répandus sur le globe terrestre [Whitman et al. 1998; Mora et al. 2011]. Pour la décrire, trois niveaux hiérarchisés, fondés autour de la notion fondamentale de l'espèce, peuvent être étudiés : la diversité intra-espèce (ou diversité intra spécifique), la diversité inter-espèces et la diversité des écosystèmes. L'étude de la biodiversité revêt une nécessité socio-économique car la survie de l'Homme dépend de très nombreuses espèces pour assurer son alimentation, sa santé, l'épuration de l'eau ou encore l'élimination des déchets.

Il est probable que l'une des premières exploitations du potentiel des microorganismes par l'Homme soit survenue il y a 10 000 ans, lors de la découverte des capacités de certaines bactéries à acidifier le lait, prolongeant ainsi sa conservation. Même si leur utilisation était à l'origine empirique, les études de caractérisation phénotypique du début du siècle dernier ont permis d'identifier les spécificités morphologiques, biochimiques et physiologiques des différentes espèces de bactéries lactiques. Ces connaissances ont contribué à rationaliser les étapes de fabrication des produits laitiers et à étendre l'utilisation de ce groupe bactérien d'intérêt à différentes matrices alimentaires. La capacité des bactéries lactiques à acidifier les matrices alimentaires par production d'acide lactique rend ces bactéries indispensables à l'industrie agroalimentaire [Leroy & De Vuyst 2004; Das & Goyal 2012]. Le marché de la production de ferments industriels représentait près de 250 millions de dollars en 2002 [Hansen 2002]. L'addition de ces bactéries au milieu, non seulement permet de prolonger la conservation du produit, mais de plus confère aux produits fermentés une saveur et une texture particulière [Marshall 1987; Leroy & De Vuyst 2004; Mills et al. 2010]. Cependant, l'utilisation de ferments pour la fabrication de produits laitiers a démontré le caractère souche-dépendant des propriétés technologiques des levains : acidification, texture, couleur, production d'arômes ou encore d'exopolysaccharides... Ces observations interrogent sur l'origine de cette importante diversité intra-espèce. D'un point de vue applicatif, l'existence de cette diversité intra-espèce garantit une diversification des produits (fromages de terroir notamment, création de nouveaux produits), mais pose la problématique de la caractérisation et de la sélection des souches.

L'avènement du séquençage, au début des années 2000, a révolutionné la microbiologie. Il a notamment ouvert la voie au décodage de l'ADN, support de l'information génétique, et à la recherche de marqueurs génétiques. L'analyse de la diversité génotypique de *Lactococcus lactis*, principal ferment mésophile de l'industrie laitière, a révélé que les souches adaptées à cette niche écologique partagent un large génome cœur et une faible variabilité génétique, à la différence des isolats végétaux [Kelly et al. 2010; Passerini et al. 2010; Tan-a-ram et al. 2011]. Cependant, les déterminants de la spécificité fonctionnelle de cette sous-population, dans l'environnement d'intérêt, restent à déterminer. Afin d'appréhender les mécanismes de régulation de l'expression génétique, et de comprendre les processus d'adaptation de la bactérie à son environnement, une approche dite de biologie intégrative a récemment émergé pour réconcilier les différentes données de l'ère « -omique » (génomique, transcriptomique, protéomique, métabolomique).

Ces travaux de thèse s'inscrivent dans cette démarche moderne de caractérisation et comparaison de souches microbiennes par une approche de biologie intégrative. Ils ont porté sur deux notions interdépendantes que sont l'évaluation de la diversité fonctionnelle de la sous-espèce *L. lactis* ssp. *lactis*, et l'identification du caractère spécifique de certaines souches. Pour ancrer cette problématique dans un environnement applicatif fort, ce projet de recherches s'est inscrit dans un partenariat entre l'équipe « Génie du métabolisme des procaryotes » du LISBP et l'industriel Yoplait autour d'un procédé modèle de production de lait fermenté. Cette étude est centrée sur une analyse approfondie du comportement d'une souche-modèle de *L. lactis* ssp. *lactis* d'origine industrielle (Yoplait) à la singularité organoleptique non expliquée. L'objectif de cette thèse est d'identifier les déterminants de cette spécificité afin d'orienter la signature aromatique de la souche d'intérêt en fonction du procédé. Ces travaux se distinguent par l'approche de biologie intégrative menée du génotype au phénotype pour déterminer ces marqueurs, et par l'analyse réalisée en matrice lait.

Ce manuscrit débute par une introduction bibliographique qui vise à restituer le contexte de cette étude. Après une présentation du rôle technologique des levains laitiers et de la diversité génétique inter-espèce de *L. lactis*, l'état de l'art sur les différentes régulations des quatre principaux métabolismes intervenant lors de la croissance en lait est exposé. En complément, la diversité fonctionnelle des levains et les principales méthodes d'identification de signatures spécifiques sont décrites. Enfin, les objectifs de cette thèse seront présentés en lien avec les différents points bibliographiques abordés. Les résultats de cette thèse sont organisés en trois axes principaux déclinés en quatre chapitres. Le premier axe porte sur un diagnostic macro-cinétique des premières heures de fermentation afin d'appréhender l'activité de la souche modèle et de décrire les principales interactions de la souche avec le procédé laitier (Chapitre 1). Le second axe concerne le

positionnement spécifique de cette souche au sein de la diversité génétique - phénotypique (Chapitre 2) ainsi que protéique (Chapitre 3) d'un panel de souches sélectionnées de *L. lactis*. Le dernier axe de cette thèse, plus applicatif, est consacré à l'identification des leviers et verrous de la production métabolique afin de moduler la signature aromatique de la souche d'intérêt (Chapitre 4). Pour clore ces travaux, l'ensemble des résultats seront repris pour proposer une image complète de la diversité fonctionnelle de la sous-espèce *L. lactis* et de la singularité de la souche d'intérêt. Cette partie s'ouvrira sur un ensemble de perspectives qui seront discutées.

ETUDE BIBLIOGRAPHIQUE

Lactococcus lactis, bactérie lactique importante pour l'industrie laitière

Ce chapitre a pour objectif d'introduire les caractéristiques de la matrice lait, de présenter les fonctions et des ferments lactiques et de positionner la phylogénie et la diversité génotypique inter et intra-espèce de *L. lactis* ssp. *lactis*. L'enjeu de la sélection des levains sera également explicité.

Le lait et ses produits de transformations

Le lait, aliment du début de la vie, est un fluide biologique, riche en éléments nutritifs pour répondre aux premiers besoins alimentaires des mammifères. De composition variable selon les espèces, à l'échelle industrielle, le lait utilisé est surtout celui de vache. C'est une solution aqueuse (87% d'eau) de nature colloïdale avec une émulsion huile dans eau, hautement complexe (plus de 100 000 molécules présentes). La composition moyenne des 13% de matière solide (MS) du lait de vache est répartie entre les trois grands groupes de substances alimentaires en proportions équilibrées [Singh & Bennett 2005] :

- Le sucre, le lactose, glucide diholoside (galactose-glucose) est à une concentration moyenne de 50 g.L⁻¹.

- La matière grasse (MG) est essentiellement sous forme de triglycérides (98%), entre 35 et 45 g.L⁻¹.

- Les protides représentent une matière azotée totale (MAT) de 35 g.L⁻¹, ils se répartissant en 78 à 80 % de caséines sous formes de micelles, 15 à 17 % de protéines du lactosérum, et 5 % partagé entre oligopeptides/acides aminés libres et azote non protéique tel que l'urée.

Un éventail intéressant de substances minoritaires est également présent avec i) des sels minéraux (dont 30 % de calcium et de phosphore) à 7 g.L⁻¹, ii) des vitamines liposolubles (A, D₃ et E autour de 0,46 mg.L⁻¹) et hydrosolubles (du groupe B, pour près de 1,69 mg.L⁻¹), et iii) des enzymes : phosphatase, peroxydase, amylase ...

Utilisé comme matière première pour de nombreux sous-produits (crème, beurre), le lait peut également subir une étape de transformation microbiologique par fermentation par la flore microbienne, comme dans le cas de la fabrication des laits fermentés ou fromages. En France, l'appellation lait fermenté désigne « un produit laitier composé exclusivement de matières premières d'origine laitière ayant subi [...] une fermentation par des micro-organismes spécifiques et caractérisé par une teneur en acide lactique minimale (0,6%) [...] ». Le terme yaourt est réservé aux produits fermentés exclusivement avec les deux bactéries lactiques thermophiles *Lactobacillus delbrueckii* ssp. *bulgaricus* et *Streptococcus thermophilus*.

La production de laits fermentés à l'échelle industrielle impose de rationaliser les procédés de production afin d'apporter régularité et qualité des produits aux consommateurs. Ainsi, la

composition du lait est-elle souvent « standardisée » en MS et MG pour limiter l'impact des variations dues aux conditions environnementales (période de lactation, alimentation du bétail) et s'adapter aux technologies laitières. Les réglementations sanitaires imposent une thermisation préliminaire, principalement la pasteurisation, dont le but est la destruction des germes pathogènes. Les germes les plus souvent évoqués [Brisabois et al. 1997] sont les mycobactéries, *Brucella*, *Listeria monocytogenes*, *Staphylococcus aureus*, et les entérobactéries, parmi lesquelles les *Escherichia coli* producteurs de toxines et *Salmonella sp.* Cependant, la pasteurisation réduit également la flore originelle du lait (bactéries lactiques ou psychotrophes, microcoques, levures ou encore moisissures) à l'origine de la diversité des produits au lait cru du terroir français [Wouters et al. 2002; Delgado et al. 2013]. Il est alors nécessaire pour les productions industrielles de substituer cette flore endogène par une sélection de microorganismes ajoutés au lait pour démarrer la fermentation, appelés levains laitiers ou « starters ».

La fonction technologique des ferments lactiques

Les ferments sont définis comme étant des cultures pures ou des mélanges de bactéries lactiques sélectionnées et utilisées pour la fabrication de produits fermentés [Leroy & De Vuyst 2004]. Parmi ces cultures, il est distingué les *ferments naturels*, souvent composés d'un mélange de nombreuses souches de bactéries dont la composition exacte est indéterminée, et les *ferments définis*, composés de plusieurs souches sélectionnées et cultivées séparément.

Dans l'industrie laitière, les bactéries lactiques sont à la base de la fabrication de fromages, yaourts et de nombreuses variétés de laits fermentés. D'autres bactéries, levures et moisissures, sont également retrouvées mais principalement dans les écosystèmes fromagers. Elles ne seront pas abordées dans la suite de cette revue puisqu'elles ne font pas partie des flores majoritaires d'adjonction des laits fermentés industriels. Deux types de ferments lactiques sont utilisés : i) les ferments mésophiles dont la température optimale de croissance est proche de 30°C et ii) les ferments thermophiles dont la température optimale est de 43°C. Les ferments mésophiles sont employés dans l'élaboration des produits laitiers frais (fromage frais, beurre, crème fraîche et lait fermenté) et de fromages affinés.

La sélection des levains est dépendante de l'ensemble des propriétés technologiques désirées dans le produit fini résultant de l'action des micro-organismes. Elle repose donc à la fois sur la connaissance des procédés mais également sur la compréhension de l'adaptation du comportement de la bactérie à cet environnement de mise en œuvre pour prédire la fonctionnalité de la souche. En effet, le ferment joue des rôles multiples au cours de la maturation du lait, ce qui contribue à créer un éventail de critères selon lesquels les souches doivent être examinées et éventuellement discriminées pour procéder à leur sélection.

- Production d'acide lactique à partir du lactose qui va aboutir à i) l'abaissement du pH (action conservatrice par la limitation du développement d'autres micro-organismes), ii) la déstabilisation des micelles de caséines qui vont créer un gel lactique par coagulation (Figure 1) et iii) la solubilisation des minéraux liés à la caséine ;
- Dégradation des caséines par son activité protéolytique qui i) active la croissance des bactéries [Juillard et al. 1996] et ii) développe les arômes par la dégradation des acides aminés qui s'ensuit ;
- Aromatisation du produit en fonction du métabolisme, avec l'implication des voies de lipolyse, protéolyse et catabolisme des hydrates de carbones (lactose, citrate) [Smit et al. 2005a; Cheng 2010];
- Augmentation de la viscosité du milieu par production d'exopolysaccharides (EPS) [Welman & Maddox 2003] ;
- Inhibition des germes indésirables par la production d'acides organiques et d'éventuelles substances inhibitrices telles les bactériocines [Klaenhammer 1988; De Vuyst & Leroy 2007], ou encore action en tant que probiotique, c'est-à-dire des « *micro-organismes vivants qui, lorsqu'ils sont ingérés en quantité suffisante, exercent des effets positifs sur la santé, au-delà des effets nutritionnels traditionnels* » (OMS, 2001).

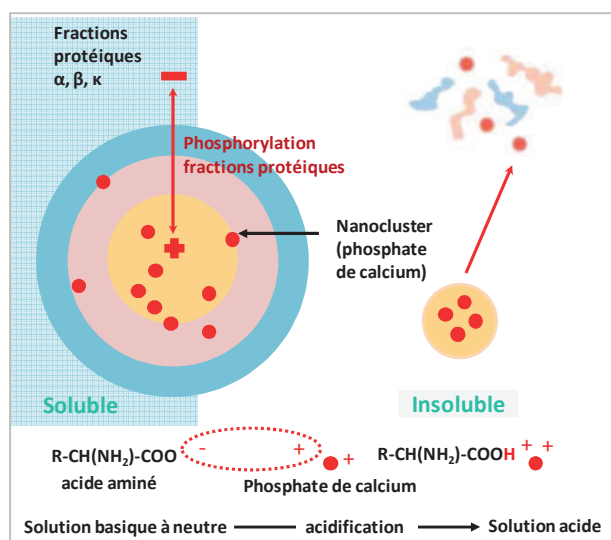


Figure 1 : Représentation schématisée de l'action déstabilisatrice de l'acidification sur le phosphate de calcium micellaire à l'origine de la coagulation du lait [Branger et al. 2007].

Les différentes fractions protéiques (α, β, κ) de la caséine sont liées par l'intermédiaire de phosphate de calcium organisé en nanocluster qui se fixe au groupement sérine présent sous forme dissociée à pH 6,8. La structure micellaire est donc stabilisée par les nombreuses interactions électrostatiques entre les fractions protéiques et les nanoclusters. Avec l'acidification du milieu, les groupements carboxyles de la sérine sont protonés et empêchent les interactions entre protéines/minéraux. Le calcium migre alors dans la phase hydrique du lait et l'édifice micellaire se désagrège pour coaguler.

Les études ont montré que les deux principaux critères d'acceptation et de préférence des consommateurs sont la qualité texturale (dépendante de la vitesse d'acidification et de la production d'EPS) et la flaveur (fonction des voies métaboliques) des produits. Ce sont donc principalement sur ces deux critères que les souches industrielles ont été souvent empiriquement sélectionnées comme de « bons ferments ».

Caractéristiques générales des bactéries lactiques

Les bactéries lactiques (Lactic Acid Bacteria, LAB) représentent une famille hétérogène de bactéries à GRAM positive, avec une composition en GC inférieure à 50%, aéro-tolérantes, acido-tolérantes et non sporulantes. Les LAB sont considérées généralement comme des espèces non pathogènes, dites GRAS (Generally Recognized As Safe), dont les propriétés fermentaires sont largement utilisées dans l'industrie agroalimentaire pour la fabrication de nombreux produits : boissons fermentées (vins, bières), salaisons (saucisses), levains de panifications et produits laitiers. Capables de coloniser différentes niches écologiques, elles sont retrouvées aussi bien à la surface des fruits et légumes, de muqueuses, ou dans des produits alimentaires. Les LAB fermentent de multiples sources carbonées soit par voie homofermentaire en produisant essentiellement de l'acide lactique, soit par voie mixte. Dans cette dernière, elles produisent en supplément, des acides organiques (acétate, formiate), de l'éthanol et du CO₂.

Ce groupe est composé des genres *Lactobacillus*, *Lactococcus*, *Streptococcus* (seulement *S. thermophilus*, les autres étant pathogènes), *Oenococcus*, *Leuconostoc*, *Pediococcus*, *Aerococcus*, *Carnobacterium*, *Enterococcus*, *Tetragenococcus*, *Vagococcus*, *Weissella*. La classification des principaux genres est présentée Figure 2.

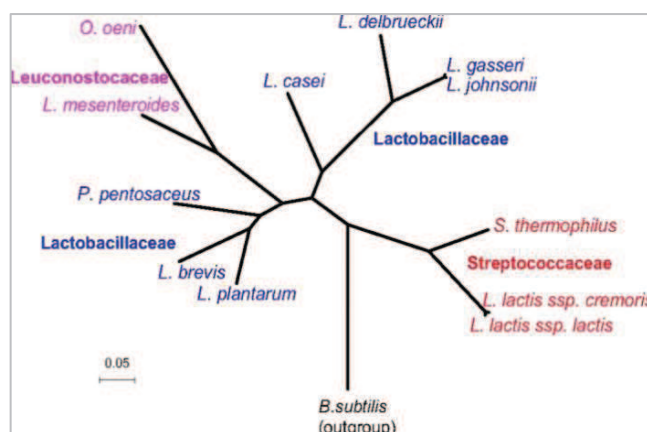


Figure 2 : Arbre phylogénétique des principaux genres du groupe des LAB basé sur la concaténation des séquences protéiques ribosomales (bootstrap >75 %) [Makarova et al. 2006]. Les différentes espèces sont regroupées par un code couleur en fonction de leur famille d'appartenance indiquée en gras.

Du fait de leur importance économique, les LAB font l'objet de recherches intensives qui sont améliorées par la disponibilité de séquences complètes des génomes de nombreuses bactéries. Depuis le premier génome séquencé appartenant à la souche *Lactococcus lactis* IL1403, en 2001 [Bolotin et al. 2001], 74 génomes complets sont aujourd'hui disponibles sur les bases de données publiques (NCBI, mise à jour juillet 2013). La majorité de ces génomes appartiennent au genre *Lactobacillus* (46 génomes).

Phylogénie et diversité de *Lactococcus lactis*

Le genre *Lactococcus*, apparenté au genre *Streptococcus*, rassemble des bactéries catalase négative en forme de coques, isolées, regroupées ou associées en chainettes de longueurs variables. Elles ne présentent aucune activité β -hémolytique, et sont mésophiles (température optimale de croissance proche de 30°C). Il inclut 7 espèces sur la base d'analyse de l'ADNr 16S: *L. lactis*, *L. raffinolactis* (lait cru et fromage), *L. garviae* (poissons, animaux et lait), *L. plantarum* (plantes), *L. piscium* (saumon), *L. chungangensis* (eaux usées) [Cho et al. 2008] et plus récemment *L. fujiensis* (choux) [Cai et al. 2011]. Parmi ces 7 espèces, *L. lactis* est considéré comme la plus importante espèce d'un point de vue de l'intérêt agroalimentaire qu'elle représente. *L. lactis* est subdivisé en 3 sous-espèces : *L. lactis* ssp. *hordniae*, ssp. *lactis* (incluant le biovar bv. *diacetylactis*) et ssp. *cremoris*. Seules les deux dernières sous-espèces sont utilisées en tant que levains. Bien qu'elles présentent une proximité nucléotidique au niveau de l'ADNr 16S de 99,3% [Salama et al. 1991], elles se différencient autant sur le plan génomique, notamment divergence sur la totalité des séquences codantes de 15% [Wegmann et al. 2007], que sur des critères phénotypiques. En effet, *L. lactis* ssp. *lactis*, contrairement à ssp. *cremoris*, est capable de croître sur des milieux à 40°C, en présence de NaCl 4% (pH 9,2), possède une glutamate décarboxylase ou encore est capable d'hydrolyser l'arginine [Nomura et al. 1999]. Le biovar *diacetylactis*, quant à lui, est constitué de souches capables de produire du diacétyle, arôme recherché par les industriels pour son goût de beurre retrouvé dans les produits laitiers.

Cependant, la taxonomie de l'espèce a souvent été remise en question puisque des souches identifiées comme ayant un génotype ssp. *cremoris* sur la base de l'ADNr 16S ont été caractérisées comme possédant un phénotype caractéristique de ssp. *lactis* [Tailliez et al. 1998; Kelly & Ward 2002; Nomura et al. 2002; Rademaker et al. 2007]. Pour étudier la diversité génétique de l'espèce *Lactococcus lactis*, 9 génomes complètement séquencés sont disponibles, et leurs caractéristiques sont résumées Tableau 1.

A l'émergence des technologies de génomique, s'ajoutent de multiples outils de typages moléculaires disponibles [Li et al. 2009] utilisés pour explorer la diversité génétique inter et intra espèce de *L. lactis* aussi bien à l'échelle génomique que génétique.

Tableau 1 : Récapitulatif des *Lactococcus lactis* complètement séquencés, disponible sur le site NCBI.
Mise à jour Juillet 2013.

Sous-espèce	Souche	Taille du génome (Mb)	Nombres de plasmides séquencés (taille en kb)	Origine	N°Accesion GenBank	Référence
<i>Lactococcus lactis</i> ssp. <i>cremoris</i>	MG1363	2,5	0	Fromage	AM40667	[Wegmann et al. 2007]
	NZ9000	2,5	0	Fromage	CP002094	[Linares et al. 2010]
	UC509.9	2,5	8(80-39-53-11-7-6-5-4)	Levain	CP003157	[Ainsworth et al. 2013]
	SK11	2,6	5 (74-47-14-14-9)	Fromage	CP000425	(Makarova et al. 2006)
	A76	2,6	4 (54-49-18-4)	fromage	CP003132	[Bolotin et al. 2012]
<i>Lactococcus lactis</i> ssp. <i>lactis</i>	CV56	2,5	5 (44-36-31-6-2)	Probiotique	CP002365	[Gao et al. 2011]
	IL1403	2,4	0	Souche laboratoire	AE005176	[Bolotin et al. 2001]
	IO-1	2,4	0	Eaux usées	AP012281	[Kato et al. 2012]
	KF147	2,6	1 (38)	Plante	CP001834	[Siezen et al. 2010]

Diversité génomique de *Lactococcus lactis*

Le développement des empreintes génomiques, notamment par électrophorèse en champs pulsés (Pulsed-Field Gel Electrophoresis, PFGE), ou encore par polymorphisme de longueur de fragments de restriction (Restriction Fragment Length Polymorphism, RFLP) [Yu et al. 2011] ou de fragments amplifiés (Amplified Fragment Length Polymorphism, AFLP) [Kütahya et al. 2011], a révélé une distinction dans l'organisation des génomes en fonction de la niche écologique. En particulier, les souches isolées de plantes possèdent un plus haut degré de divergence que celles présentes au sein des produits laitiers. L'hybridation comparative de génome (Comparative Genome Hybridation, CGH) a été utilisée pour établir des liens entre diversité génotypique, niche écologique et marqueurs phénotypiques au sein de l'espèce *lactis*. En particulier, les études de Makarova et coll en 2006 et Siezen et coll en 2011 ont permis d'identifier, à partir d'une analyse chromosomique des groupes de gènes orthologues (chrOGs), une liste de 1268 chrOGs communs à toutes les souches de *L. lactis*, décrivant les procédés cellulaires essentiels de l'espèce, et appelée « core chrOGs » [Makarova et al. 2006; Siezen et al. 2011]. De plus, uniquement 72 de ces chrOGs sont spécifiques de l'espèce *lactis* après comparaison aux clusters de gènes orthologues de la famille des *Lactobacillales* (LacGOGs). Il a pu également être mis en évidence respectivement 151 et 72 chrOGs présents systématiquement dans les souches de ssp. *cremoris* et ssp. *lactis* qui peuvent donc servir de marqueurs génétiques de la sous-espèce.

Diversité génétique de *Lactococcus lactis*, approche MLST

La technique MLST (MultiLocus Sequence Typing) repose sur le séquençage d'un fragment interne d'environ 450 - 500 pb, de 5 à 7 gènes de ménage [Maiden et al. 1998]. Chaque isolat est identifié par un profil allélique, ou ST (Sequence Type), dépendant de son génotype (pour plus de détails, voir Figure 3). D'une part, l'analyse des STs permet d'établir le lien de parenté, ou ressemblance génotypique entre les souches. D'autre part, elle permet de mettre en évidence les premières étapes de diversification, c'est-à-dire de connaître l'évolution récente d'une population et de caractériser les relations phylogénétiques. La MLST a démontré être une méthode robuste et fiable pour décrire la phylogénie et la structure de la population avec un nombre limité de gènes. Des analyses récentes sur la sous-espèce *lactis* ont confirmé l'adaptation à la niche écologique laitière des souches et ont conduit à proposer une nouvelle classification scindée en deux écotypes [Passerini et al. 2010]. Le premier correspond aux souches dites « domestiquées », souches essentiellement isolées de produits laitiers ou fermentés, et le second aux souches « environnementales » retrouvées dans des sources diverses (plantes, animaux, lait cru) et qui représente le réservoir de diversité de la sous espèce *lactis* (plus de détails Figure 4).

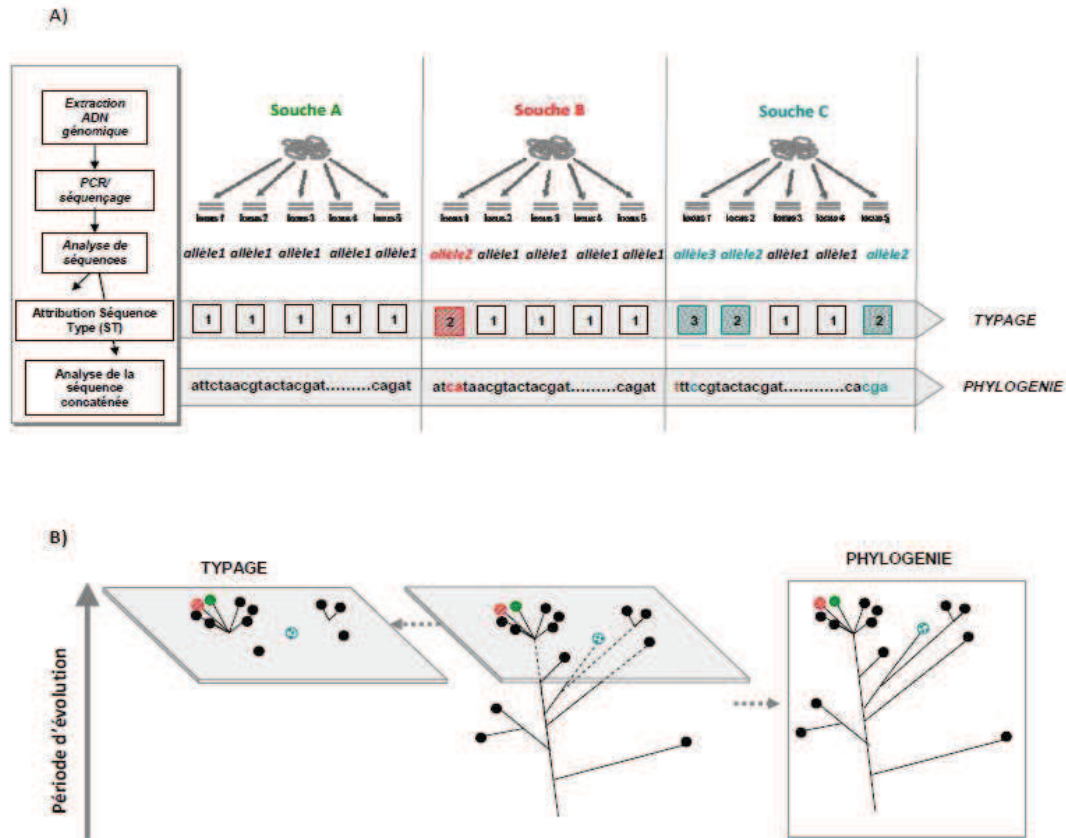


Figure 3 : Méthodologie et applications de l'approche MLST [Passerini 2011]. A) L'analyse des séquences de fragments internes de 5 gènes des souches A, B et C, permet de leur attribuer un profil allélique (ou Séquence Type, ST) sous forme d'un code à 5 chiffres. Le typage, regroupement des souches selon leur ST, permet de visualiser l'évolution récente d'une population. La concaténation des 5 STs permet d'analyser la phylogénie, repositionne dans le temps les évènements de séparation. B) Représentations d'un typage et de la phylogénie.

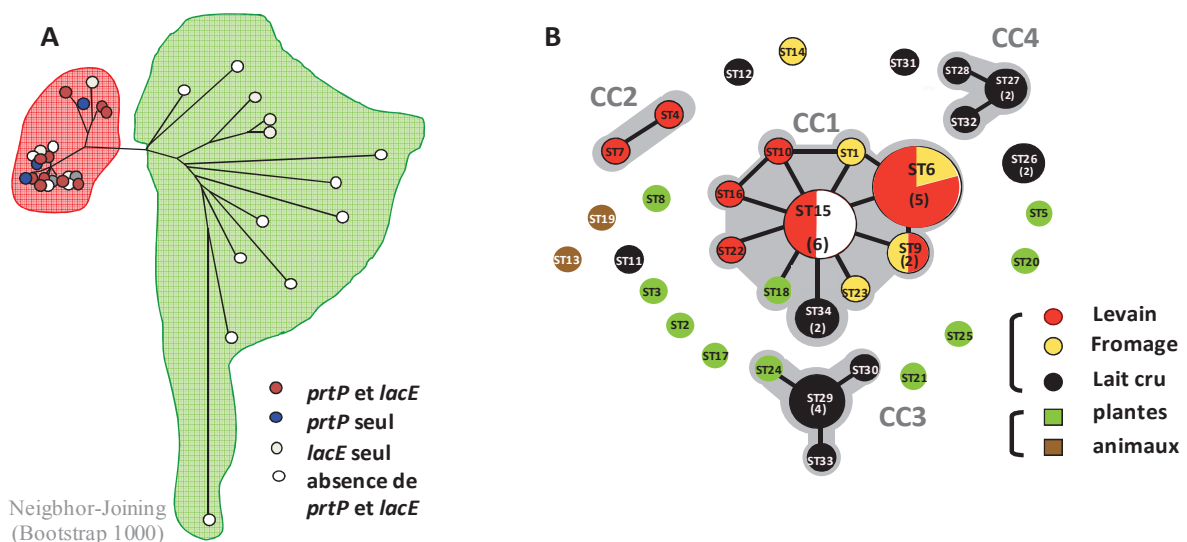


Figure 4 : Phylogénie de la sous-espèce *Lactococcus lactis* ssp. *lactis* par l'approche MLST [Passerini 2011]. A) Le positionnement phylogénétique des souches avec respectivement en rouge et en vert, la représentation

des deux écotypes « domestiqué » et « environnemental ». B) Analyse eBurst des 54 souches *ssp. lactis* organisées en 4 complexes clonaux (CC1 à CC4) avec, indiqué par un cercle coloré, l'origine des souches.

I Métabolisme de *Lactococcus lactis*

Ce chapitre décrit les quatre grands types de métabolisme de *Lactococcus lactis* : carboné, azoté, lipidique et nucléotidique. Chaque partie sera abordée par une approche transversale de biologie des systèmes afin de décrire tous les éléments de régulation connus, du gène au métabolite, pour expliciter le comportement de la souche face à son environnement. Les données spécifiques à la matrice lait seront présentées en détail pour éclairer au mieux le cadre de cette thèse.

Métabolisme carboné

L'utilisation de *Lactococcus lactis* dans l'industrie alimentaire ainsi que sa présence dans différentes niches écologiques la confronte à de multiples sources de carbone. Après avoir brièvement abordé les transporteurs spécifiques des sucres fermentescibles par *L. lactis*, nous nous intéresserons plus particulièrement dans cette section aux voies cataboliques impliquées dans l'utilisation du lactose, du glucose, du galactose ainsi que du citrate en raison de leur présence en milieu laitier et de leur contribution à la production d'arômes.

Transports des sucres

La membrane cytoplasmique, de structure hydrophobe, représente une barrière physique imperméable, séparant le cytoplasme (siège du métabolisme) du milieu extérieur (source de substrat). Pour pénétrer dans la cellule, les sucres vont nécessiter la présence de système de transport actif (Figure 5). Chez *L. lactis*, comme chez les bactéries GRAM positives, deux systèmes membranaires sont prépondérants : le système Phosphoenolpyruvate-phosphotransférase (PTS) qui couple translocation et phosphorylation du sucre [Poolman 1993] et le système perméase qui internalise le sucre sous forme libre [Thompson 1987].

a. Système perméase

Dans le système perméase énergie-dépendent, les sucres (cas du glucose, lactose et galactose, mais aussi du citrate) sont internalisés sous leur forme native et éventuellement contre leur gradient de concentration aux dépens de l'énergie générée sous forme de gradient de protons de part et d'autre de la membrane appelée force protomotrice, ou FPM. Les hydrates de carbone internalisés sont alors phosphorylés par une kinase ATP-dépendante avant d'être dégradés. Le lactose internalisé par la lactose perméase, va être tout d'abord hydrolysé par la β -galactosidase pour donner du glucose (phosphorylé en Glucose-6P, G6P) et du galactose. Ce dernier, également intégré par une perméase, va suivre la voie de Leloir pour être transformé en G6P et intégrer la glycolyse (Figure 5).

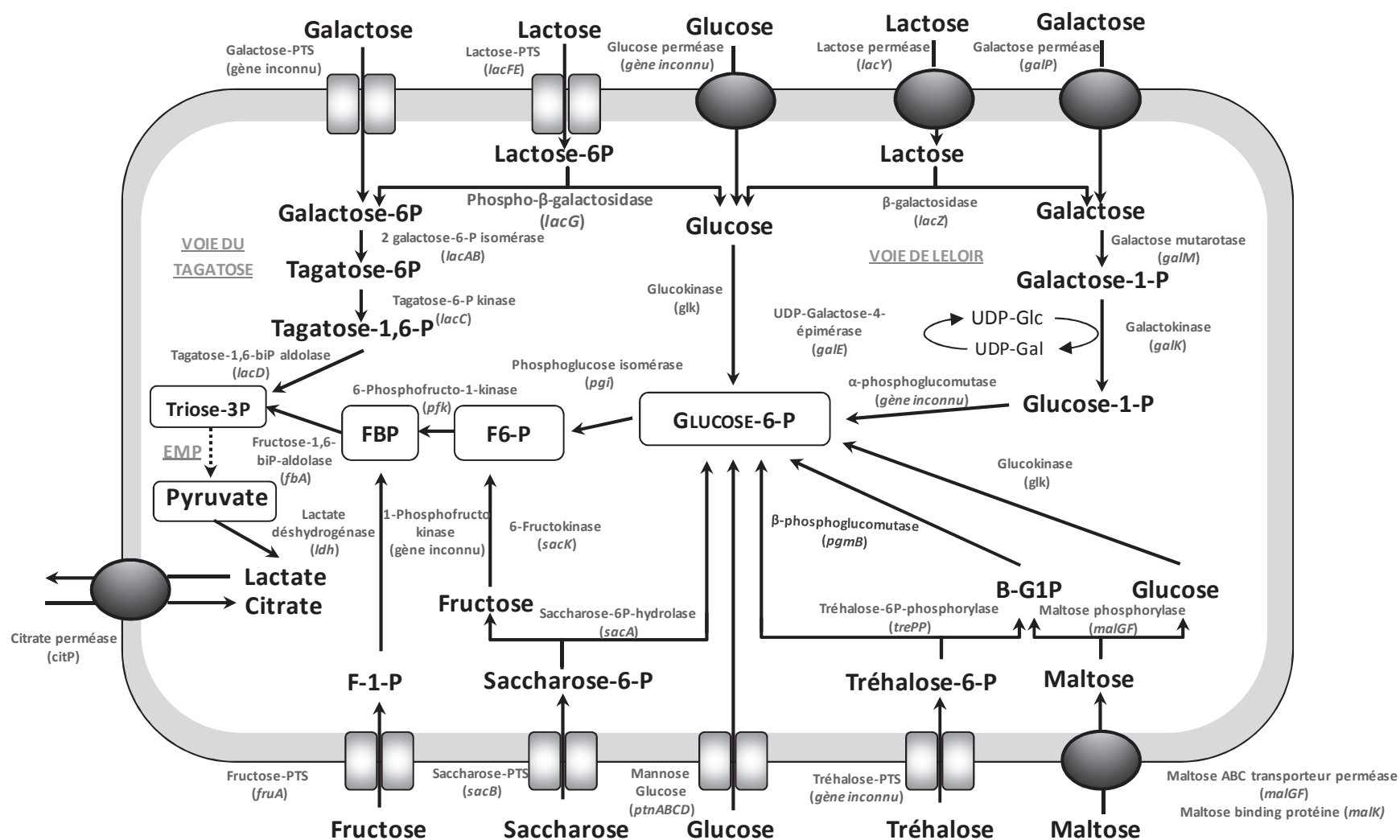


Figure 5 : Vue globale des voies métaboliques de transport et de catabolisme des sucres et de l'acide citrique chez *Lactococcus lactis ssp. lactis*. Il est indiqué pour chaque réaction les enzymes qui catalysent ces voies ainsi que les gènes impliqués entre parenthèses. Les symboles ronds grisés représentent des perméases, et les canaux grisés, les enzymes membranaires du PTS. Les métabolites encadrés appartiennent à la glycolyse. EMP : Embden-Meyerhof-Parnas. [Neves et al. 2005; Coccagn-Bousquet et al. 2002].

b. Système PTS

D'un point de vue bioénergétique, le système PTS est probablement le transport le plus efficace puisque le sucre est simultanément internalisé et phosphorylé grâce à l'utilisation d'une molécule de PEP (Phosphoénol pyruvate ; équivalent à une molécule d'ATP consommé par carbohydrate internalisé). En effet, un sucre internalisé par un transport-non PTS, implique l'utilisation d'ATP à la fois pour générer la FPM nécessaire au transport du sucre ainsi que pour la phosphorylation puisque les deux phénomènes sont physiquement distincts.

Le PTS est un système multi composants qui se compose de deux protéines aspécifiques impliquées dans le couplage énergétique, l'enzyme I (EI) et HPr, et de protéines spécifiques au sucre appartenant au complexe « enzyme II » en charge de l'internalisation et de la phosphorylation (voir Figure 6)[Poolman 1993]. Le processus débute avec le transfert du groupement phosphorylé du PEP vers EI puis vers la protéine HPr qui, une fois phosphorylée sur le résidu His-15, catalyse la phosphorylation des protéines EII. Les protéines EII B et C transfèrent alors le groupement jusqu'au sucre.

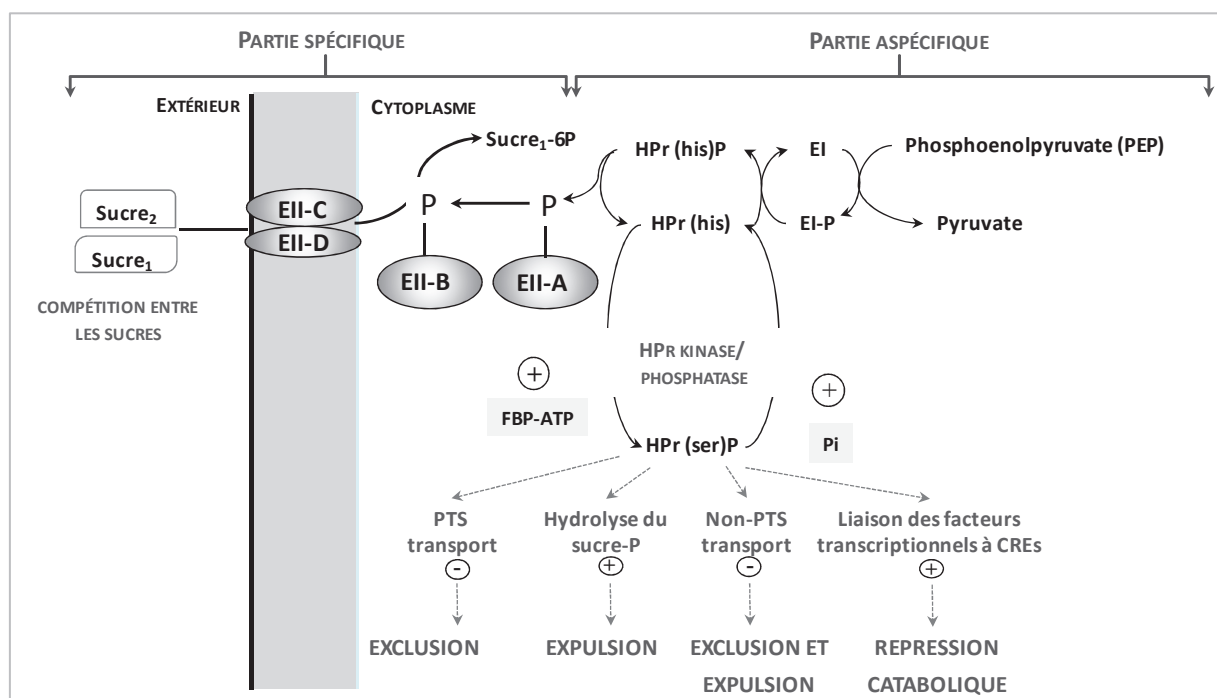


Figure 6 : Système Phosphoenolpyruvate-phosphotransférase (PTS) de transport des sucres et positionnement des mécanismes de régulation de type biochimique et génétique. Cas d'un système à 4 protéines au sein de la perméase II (A, B, C, D) avec les domaines EII-B et C cytoplasmiques. CRE : Catabolite Response Element.

A la différence du glucose, directement intégré sous forme de G6P par le PTSmannose/glucose, le galactose internalisé par le PTS-galactose ou issu du clivage du lactose transporté par le PTS-lactose va suivre la voie du tagatose. Le galactose-6-P va alors être transformé en trioses phosphate qui vont rejoindre la partie basse de la glycolyse (Figure 5).

c. Contrôle du métabolisme des sucres

Pour s'adapter à des changements de source de carbone, *L. lactis* doit pouvoir moduler les activités de ses enzymes et réguler les flux de produits en réponse aux variations de substrats. Les mécanismes de régulations impliqués dans le contrôle du métabolisme des sucres sont recensés dans la revue de Kowalczyk et Bardowski [Kowalczyk & Bardowski 2007]. Ils se divisent en deux groupes selon qu'ils sont de nature biochimique (modulation des activités enzymatiques) pour une action rapide, ou génétique (induction ou répression des gènes cataboliques) dans le cadre d'une réponse à plus long terme (résumé Figure 6). Dans ces deux cas, la réponse implique les acteurs du transport PTS. L'objectif de ces régulations est de pouvoir hiérarchiser la consommation des sucres si plusieurs sont présents, et de s'adapter à une carence en sucre afin de favoriser l'assimilation du composé au rendement de croissance maximal.

Deux phénomènes liés aux régulations biochimiques vont être abordés, la compétition entre les sucres et les mécanismes d'expulsion, d'exclusion et de répression de l'inducteur. Le premier de ces mécanismes intervient lors du transport des sucres au niveau des protéines du complexe EII, par un phénomène de compétition entre les sucres. Ces régulations expliquent le différentiel observé chez *L. lactis* entre les vitesses de consommation du glucose (rapidement métabolisé) et du galactose (plus faiblement internalisé en présence de glucose) à l'origine du phénomène de diauxie [Thompson et al. 1978]. Cette hiérarchisation dans la consommation du substrat dépend de la préférence de la bactérie pour les différents sucres. Ces critères de préférences sont fonction de i) la constante d'affinité du sucre pour une même EII membranaire (cas du PTS mannose qui peut internaliser du mannose ou du glucose [Thompson 1987]) et de ii) la compétition existante entre différentes protéines EII non membranaires pour être phosphorylées par une même HPr-P.

Le second type de mécanisme implique la protéine HPr-P dans trois phénomènes : i) l'exclusion et ii) l'expulsion de l'inducteur, ainsi que iii) la répression catabolique (voir Figure 6).

Le mécanisme d'exclusion a pour but d'empêcher l'entrée d'un sucre ou la formation d'inducteurs de gènes cataboliques par inhibition allostérique d'un transporteur de type PTS ou perméase [Saier 1989]. Intervient dans cette régulation une phosphorylation alternative de la protéine HPr sur le résidu sérine 46 en lieu et place de l'histidine 15 par l'HPr kinase/phosphatase. Il a été démontré chez *Lactobacillus casei* que la HPr-(Ser)P ne peut pas entrer dans la cascade de phosphorylation utilisable pour phosphoryler le sucre [Viana et al. 2000]. Elle peut cependant interagir avec certaines perméases PTS et non PTS pour bloquer l'entrée d'inducteurs. La phosphorylation de HPr sur la sérine 46 est stimulée par la présence d'intermédiaires glycolytiques tels que le fructose-1,6-biphosphate (FBP), l'ATP, et inhibée par la présence de phosphate inorganique [Galinier et al. 1999]. Ainsi la formation de HPr-(ser)P est dépendante du flux glycolytique et donc de l'entrée d'un sucre

métabolisable. Lorsque la bactérie entre en état de carence d'un sucre rapidement métabolisable, les faibles niveaux de FBP, d'ATP et les hautes concentrations en Pi favorisent la phosphorylation sur le résidu histidine de HPr afin de permettre l'entrée d'un autre sucre.

Le second mécanisme de régulation, appelé expulsion de l'inducteur, est un procédé en deux étapes successives qui comprend, en présence d'un sucre rapidement métabolisable (le glucose), la déphosphorylation du sucre phosphate non préférentiel (le galactose) internalisé puis son expulsion vers le milieu extérieur [Reizer et al. 1983; Reizer & Saier 1983]. L'implication de la protéine HPr-(Ser)P dans l'activation d'une phosphorylase intracellulaire du sucre-P est suggérée par les travaux de Ye en 1994 [Ye et al. 1994].

Le dernier mécanisme concerne la régulation génétique des gènes du catabolisme des sucres lors du changement de source carbonée. Toutes les bactéries sont capables de percevoir l'état nutritionnel de leur environnement grâce à des réponses régulatrices, dites CCR (Carbon Catabolite Repression). Chez *L. lactis*, la « CCR » est sous le contrôle des protéines HPr, HPr kinase/phosphatase, et du régulateur transcriptionnel CcpA (Catabolite control protein-A) qui détectent et transmettent l'information sur l'état nutritionnel. La détermination du régulon CcpA par une étude transcriptomique comparative entre la souche *L. lactis* ssp. *cremoris* MG1363 et le mutant $\Delta ccpA$ a d'ailleurs révélé l'étendue du contrôle catabolique par CcpA [Zomer et al. 2007]. Brièvement, lors de l'internalisation d'un sucre rapidement métabolisable par le système PTS, les concentrations intracellulaires de FBP augmentent et vont donc activer l'activité kinase de HPr kinase/phosphatase qui va phosphoryler HPr sur la sérine. HPr-(ser)P va alors agir comme co-répresseur ou co-activateur avec CcpA en fonction de la localisation des sites *cre* (catabolite response element) sur l'ADN. La liaison aux sites *cre*, situés à divers endroits sur les gènes du catabolisme des sucres, va réguler la transcription de ces gènes. La fixation de CcpA sur un site *cre* en i) amont des promoteurs, sera considérée comme activatrice, à la différence des sites à ii) l'intérieur de ces derniers ou iii) en aval, qui impliquent respectivement la répression et l'avortement de la transcription [Ye et al. 1994; Coccagn-Bousquet et al. 2002].

La glycolyse

La voie centrale du catabolisme des sucres est la glycolyse. En effet, quelque soit le sucre assimilé par la cellule, une fois transporté à l'intérieur de la cellule, il va rejoindre la glycolyse, soit au niveau du glucose-6-phosphate (G6P), du fructose-6-phosphate (F6P), ou bien au niveau du fructose-1,6-phosphate (FBP) pour le galactose/lactose internalisé par le système PTS. La glycolyse est une succession de 7 réactions qui conduit à la formation de 2 moles de pyruvate par mole de sucre en C6 (voir Figure 7). Dans la littérature, les études sur les régulations biochimiques identifient trois

enzymes potentiellement clés de la glycolyse, la phosphofructokinase (PFK), la glyceraldéhyde-3-phosphate déshydrogénase (GAPDH) et la pyruvate kinase (PK) [Llanos et al. 1993; Luesink et al. 1998; Even et al. 1999; Oh et al. 2011]. La phosphofructokinase catalyse la conversion du F6P en FBP en utilisant un ATP. La PFK de *Lb. bulgaricus* a été purifiée, et l'étude de sa structure révèle un tétramère composé de quatre sous unités [Le Bras & Garel 1991]. Le gène *pfk* codant cette enzyme est transcrit en opéron bicistronique avec le gène, *pyk*, de la pyruvate kinase [Branny et al. 1993; Branny et al. 1996]. La régulation de ces deux activités enzymatiques (PFK et PK) est impliquée dans une boucle de rétro-inhibition de la glycolyse, dans laquelle le substrat d'une enzyme est l'effecteur allostérique de l'autre. Ainsi, la PFK est partiellement inhibée par le phosphoénolpyruvate (PEP) à bas pH [Le Bras & Garel 1993], qui est le substrat de la pyruvate kinase (PK). En effet, cette dernière catalyse la conversion du PEP en pyruvate en synthétisant un ATP. Une large gamme d'activateurs de la PK est connue : G6P, ribose-5-P, et le F6P, substrat de la PFK. Cette action activatrice est supprimée en présence de FBP [Le Bras & Garel 1993]. En revanche, la PK est inhibée principalement par le phosphate inorganique (Pi) chez *L. lactis* [Llanos et al. 1993; Collins & Thomas 1974]. Cette combinaison d'une ante-activation par le F6P et d'une rétro-inhibition par le PEP assure l'amortissement des variations du flux glycolytique. La GAPDH catalyse la conversion du glyceraldéhyde-3-phosphate (GAP) en 1,3-diphoglycérate en réduisant un NAD^+ . Deux gènes, *gapA* et *gapB*, homologues à 80%, sont présents chez *L. lactis* [Bolotin et al. 2001]. La GAPDH exerce un contrôle majeur chez *L. lactis* lors du métabolisme homolactique. Ainsi, l'enzyme est inhibée par le NADH selon un mécanisme de type inhibition compétitive modélisable par le ratio NADH/NAD^+ [Even et al. 1999; Garrigues et al. 1997].

Métabolisme du citrate

Voie alternative à la glycolyse pour la production de pyruvate, le catabolisme du citrate au sein de l'espèce *L. lactis* ssp. *lactis* est lié à la présence d'un plasmide de 7,9 Kb contenant les gènes de l'opéron *citQRP* nécessaires à l'internalisation du citrate [Felipe et al. 1995]. En particulier, *citP* code une citrate perméase, seul transporteur connu pour l'internalisation du citrate chez *L. lactis*. La protéine CitR est impliquée dans la régulation du gène *citP*. Après l'internalisation du citrate et la production d'acétate par l'action de la citrate lyase (*citDEF*), l'oxaloacétate formé va être décarboxylé en pyruvate et CO_2 par l'oxaloacétate décarboxylase (*citM*) (voir détail Figure 7).

Cette réaction contribue à la formation des ouvertures dans certains fromages (Roquefort et Gouda). Ces deux dernières enzymes sont codées par des gènes chromosomiques. Les LAB internalisant et métabolisant le citrate sont identifiées comme appartenant au biovar *diacetylactis* [Hugenholtz et al. 1993].

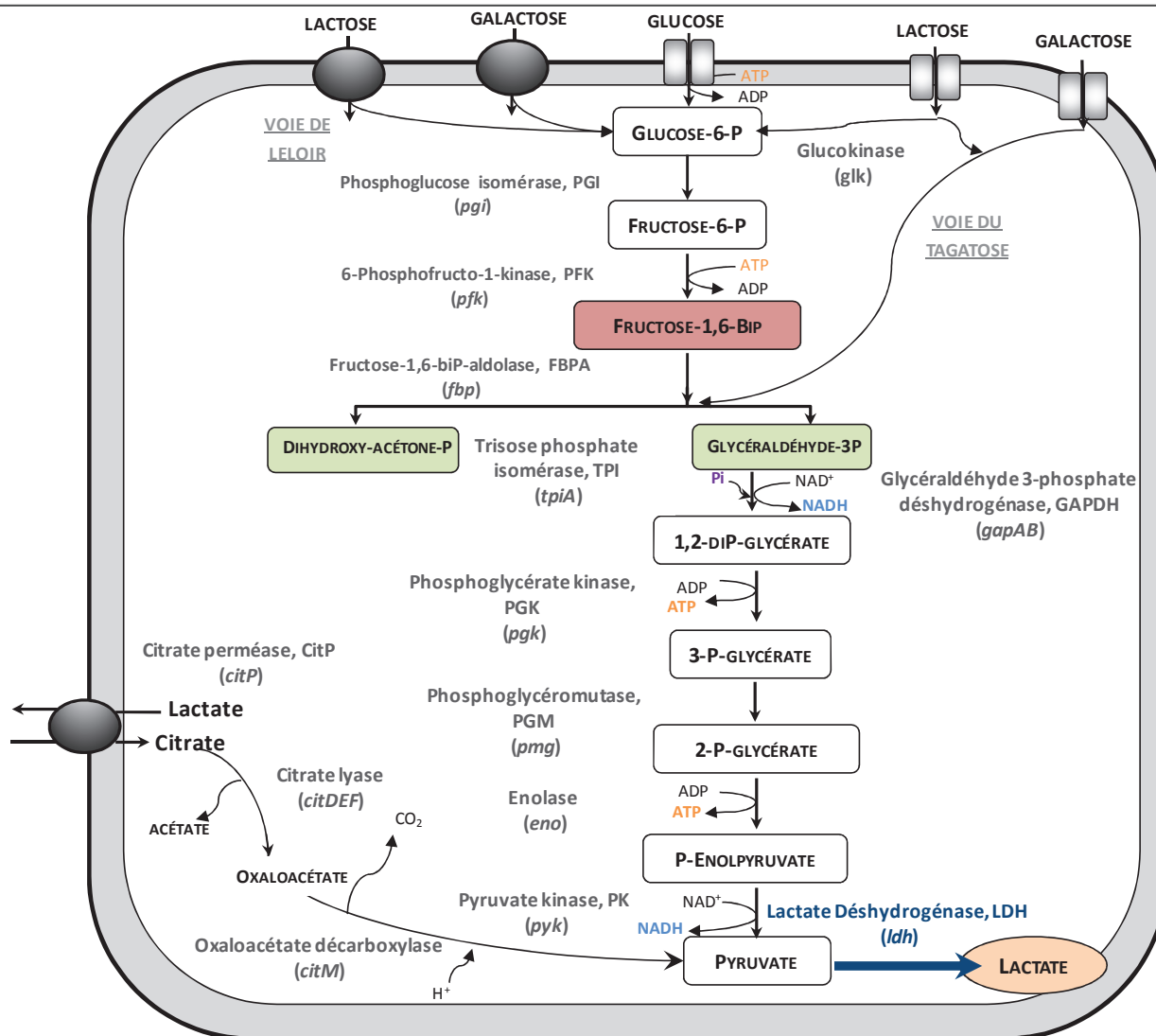


Figure 7 : Nutrition carbonée en lait : voie d'entrée des carbohydrates (lactose, galactose, glucose et citrate), glycolyse, et métabolisme du citrate. Il est indiqué pour chaque réaction le nom des enzymes qui catalysent ces voies, leur abréviation ainsi que les gènes impliqués. Les intermédiaires glycolytiques apparaissent encadrés et le principal produit fermentaire est sous forme d'ellipse. Les intermédiaires glycolytiques en rouge et vert jouent un rôle dans la régulation du métabolisme.

Cette dénomination provient du co-métabolisme citrate-lactose qui entraîne la production de coproduits d'intérêts industriels, diacétyle et CO₂ (voir Figure 8), qui participent activement aux qualités aromatiques et texturales des produits laitiers. Cependant, des travaux récents ont démontré que cette capacité de production de diacétyle/acétoïne n'est pas exclusive des souches du biovar *diacetylactis* et que d'autres souches ne possédant pas le gène *citP* sont capables de niveaux de production comparables [Passerini et al. 2013]. L'internalisation du citrate par la perméase CitP est également impliquée dans la résistance de *L. lactis* ssp. *lactis* biovar. *diacetylactis* aux conditions acides [Magni et al. 1999].

Le lactate produit est éliminé dans le milieu exocellulaire via la citrate perméase qui effectue également le passage du citrate dans le cytoplasme de la cellule. Cet échange crée un potentiel trans-membranaire. La conversion du citrate en lactate s'accompagne de la consommation de protons au niveau de la décarboxylation de l'oxaloacétate en pyruvate, et par conséquent d'une alcalinisation du milieu intracellulaire.

Destin du pyruvate

Véritable nœud métabolique du métabolisme central, le pyruvate est l'intermédiaire métabolique à l'origine des différents produits fermentaires. En fonction des ratios des différents produits, le métabolisme peut-être dit homolactique (l'acide lactique est le produit très majoritaire) ou mixte (d'autres co-produits sont formés).

a. Fermentation homolactique

Les lactocoques sont considérés comme des microorganismes homolactiques puisque généralement, ils convertissent 90% du pyruvate en acide lactique. Ces microorganismes étant dépourvus de chaîne respiratoire fonctionnelle, la réoxydation du NADH produit lors des étapes de la glycolyse en condition anaérobie est assurée par la lactate déshydrogénase (LDH) qui transforme le pyruvate en acide lactique.

b. Fermentation mixte et production d'arômes en aval du pyruvate

▪ Métabolisme mixte :

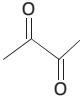
Dans le cas d'un métabolisme mixte, le flux à travers la LDH diminue et le lactate n'est plus l'unique produit de fermentation. En effet, le pyruvate peut alors être métabolisé en acétyl-CoA soit par la pyruvate formiate lyase (PFL), soit par la pyruvate déshydrogénase (PDH), en fonction des conditions d'oxygénation (voir Figure 8). La PFL n'est fonctionnelle qu'en anaérobiose, et conduit également à la formation de formiate [Takahashi et al. 1982]. Cette enzyme est inhibée par les trioses-phosphate (GAP et DHAP) [Abbe et al. 1982]. La décarboxylation du pyruvate par la PDH, permet de générer outre du CO₂ (intérêt en fromagerie pour la formation de cavités) de l'acétyl-CoA et un autre

cofacteur réduit (NADH). Faiblement exprimée en aérobiose, elle est inhibée en anaérobiose et sensible aussi bien au potentiel rédox qu'au ratio NADH/NAD [Snoep et al. 1993]. La métabolisation de l'acétyl-CoA en acétate est réalisée par actions successives de la phosphotransacétylase (PTA) et de l'acétate kinase (ACK), alors que la formation d'acétaldéhyde et d'éthanol est catalysée par l'alcool déshydrogénase (ADHE). Si la voie de l'acétate permet la synthèse d'une molécule d'ATP supplémentaire, celle de l'éthanol régénère 2 NADs⁺. Ainsi, un équilibre réducteur équivalent au métabolisme homolactique peut alors être obtenu par répartition équimolaire de la production acétate/éthanol avec un gain net de 50% en ATP.

▪ Production d'arômes à partir du pyruvate :

Parmi les plus de 90 composés aromatiques identifiés dans les laits fermentés [Ott et al. 1997], l'acétaldéhyde produit à partir de l'acétyl-CoA ainsi que les produits issus de la conversion de l' α -acétolactate (diacétyl, acétaldéhyde, butanediol) sont considérés comme les principales molécules responsables de la flaveur des laits fermentés [Cheng 2010; Tamime & Robinson 1999; Friedrich & Acree 1998]. En effet, ces composés ont la particularité d'être à la fois retrouvés en concentrations importantes dans les différentes matrices étudiées, et de présenter des seuils olfactifs de reconnaissance particulièrement bas [Grosch 1993; Imhof & Bosset 1994; Ott et al. 2000; Czerny et al. 2008]. Les propriétés chimiques et olfactives de ces composés sont résumés Tableau 2.

Tableau 2 : Propriétés chimiques et organoleptiques des quatre composés carbonés clefs issus du pyruvate dans les laits fermentés.¹ : [Corrieu 2008]. ND : Non déterminé.

Nom usuel (nomenclature chimique)	Formule	Descripteurs de l'arôme	Gamme de concentration retrouvée dans les laits fermentés (mM)	Seuil olfactif en ppm (eau/huile) ¹
Acétaldéhyde		Ether, frais, vert yaourt et acre	1 à 13,5	1,3/0,1
Diacétyl (butane-2,3-dione)		Beurre, crémeux, noisette	0,1 à 13,0	0,0054/0,014
Acétoïne (3-hydroxy-2-butanone)		Beurre, lait acide	Trace à 8,2	1
Butane-2,3-diol		fruité	Trace à 5	ND

L' α -acétolactate est synthétisé à partir de 2 molécules de pyruvate par l'acétolactate synthase catabolique (ALS) (Figure 8). De nature chimique instable, il peut-être soit spontanément décomposé en diacétyl en présence d'oxygène, soit converti en acétoïne par l' α -acétolactate décarboxylase

(ALD). La diacétyle-acétoïne réductase va alors catalyser les transformations successives du diacétyle en acétoïne puis en 2,3-butandiol avec régénération de NADH à chaque étape. Il existe une seconde acétolactate synthase, dite acétolactate synthase anabolique (ILV) codée par les gènes *ilvBN* impliquée dans la synthèse des acides aminés branchés [Godon et al. 1992].

c. Régulation du shift homolactique/mixte

La régulation du shift homolactique/mixte est principalement gouvernée par un contrôle du ratio NADH/NAD⁺. Le mécanisme a été mis en évidence lors des transitions nutritionnelles glucose-galactose chez la souche *NCDO 2118* [Neves, Wietske A. Pool, et al. 2005; Garrigues et al. 1997]. En effet, lorsque la croissance s'effectue sur une source de carbone rapidement métabolisable, le flux glycolytique est élevé et le métabolisme est homolactique. Lorsque la croissance est limitée (carence en carbone ou sucre faiblement métabolisable) un shift vers un métabolisme mixte s'opère. Un flux glycolytique élevé impose une production importante de cofacteur réduit par la glycolyse et donc un ratio NADH/NAD⁺ fort. En conséquence, les concentrations intracellulaires des trioses-P et du FBP augmentent (limitation de la glyceraldéhyde-3-P déshydrogénase (GAPDH) par le fort ratio NADH/NAD⁺). Ils vont respectivement inhiber la PFL et activer la LDH assurant ainsi une production de lactate, régénératrice de NAD⁺. Au contraire, lors d'un ralentissement du flux glycolytique à vitesse anabolique identique, le ratio NADH/NAD⁺ est plus faible et la limitation de l'activité de la GAPDH est levée. Les concentrations des trioses-P et FBP alors diminuent, la LDH est moins activée et par conséquent, le flux du pyruvate va alors être réorienté vers les voies fermentaires du métabolisme mixte pour garantir l'équilibre énergétique.

La présence d'oxygène dans le milieu est un autre facteur environnemental pouvant déclencher un shift vers le métabolisme mixte. En effet, en aérobiose, l'activité de la NADH oxydase rentre en compétition avec la LDH comme voie alternative de régénération de NADH en utilisant l'oxygène comme accepteur d'hydrogène (Figure 8). Le ratio NADH/NAD⁺ va diminuer et le flux de pyruvate au travers de la LDH d'autant. Le pool de pyruvate disponible va être réorienté vers l'ALS et la pyruvate déshydrogénase, augmentant ainsi la production des composés d'intérêt aromatique, acétoïne/diacétyle, acétaldéhyde. De plus, la présence d'oxygène inhibe la PFL, ce qui bloque toute production de formiate bien que les concentrations en trioses-P ne soient pas élevées.

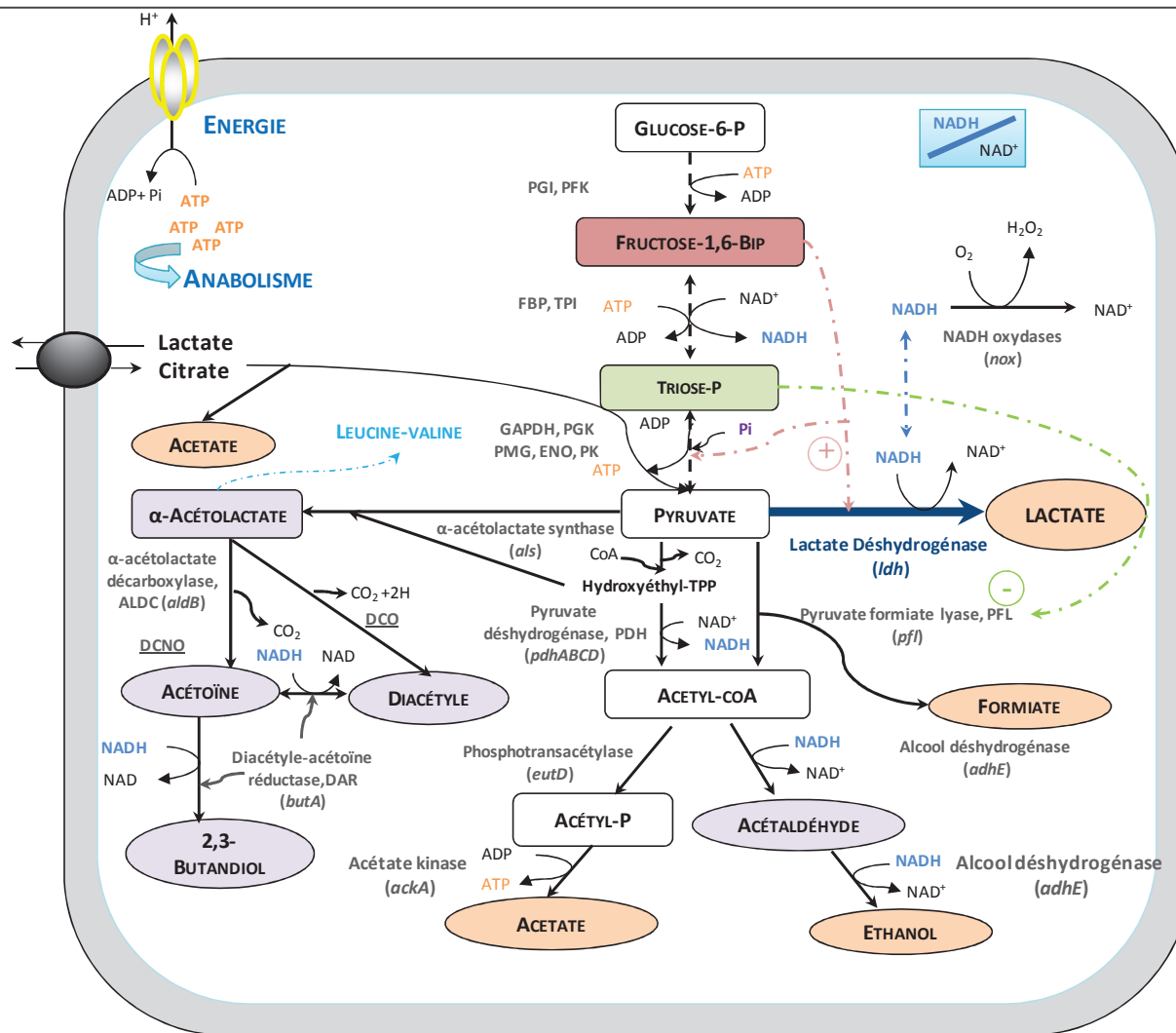


Figure 8 : Métabolisme du pyruvate et rôle majeur du ratio NADH/NAD⁺ dans le contrôle du shift homolactique vers un métabolisme mixte pour la production d'arômes. Il est indiqué pour chaque réaction les enzymes qui catalysent ces voies ainsi que les gènes impliqués. Les carrés représentent les intermédiaires réactionnels, et les ellipses les produits fermentaires. Sur fond violet, les composés responsables d'arômes dans les produits laitiers et en orange les non aromatiques. Les flèches en pointillé matérialisent les actions régulatrices d'intermédiaires glycolytiques (en vert les actions inhibitrices et en rouge les actions activatrices). DCO : décarboxylation chimique oxydative et DCNO : décarboxylation chimique non oxydative. Les abréviations des enzymes de la glycolyse sont détaillées Figure 7.

La production d'énergie

Il est vital pour une cellule vivante de produire de l'énergie pour satisfaire d'une part, les processus cellulaires de synthèse de biomasse (approvisionnement en monomères et assemblage en macromolécules), et d'autre part, la maintenance cellulaire liée à l'homéostasie cellulaire et au renouvellement des macromolécules. Cette énergie métaboliquement utilisable par la cellule peut-être produite par deux processus i) le couplage entre le catabolisme et la production de liaisons phosphodiester riches en énergie, stockée notamment sous forme d'ATP (énergie chimique) et ii) le transfert d'électrons selon les gradients ioniques des solutés (énergie chimiosmotique).

La majeure partie de l'énergie chimique des bactéries fermentaires est fournie par le catabolisme des substrats et les réactions de couplage énergétique. *L. lactis* est une bactérie homofermentaire qui obtient l'essentiel de son ATP par la glycolyse. Le bilan énergétique de cette voie conduit en condition optimale de croissance à 2 moles d'ATP et 2 moles de NADH produites par mole de glucose consommée. Ce rendement est de quatre moles d'ATP et de NADH par mole de lactose métabolisé. Puisque le sucre peut également participer à la production de biomasse, la production d'ATP est plus sûrement évaluée à partir de la quantité des produits de fermentation. Ainsi une mole d'ATP est produite par mole de lactate et d'éthanol à partir de la transformation des trioses phosphates dans la partie basse de la glycolyse, plus une mole d'ATP par mole d'acétate synthétisée.

L'énergie chimiosmotique libre est convertie en énergie électrochimique par le couplage chimique des réactions de translocation d'ions à travers la membrane [Mitchell 1974]. La force d'expulsion des ions dépend de la différence de potentiel redox au niveau d'un site de translocation et de la variation d'énergie libre d'hydrolyse de l'ATP. Dans la majorité des bactéries, tout comme chez *L. lactis*, ces ions sont des protons. Le gradient électrochimique des H^+ exerce une force appelée force protomotrice (FPM). Les systèmes capables de générer ces gradients électrochimiques sont appelés transporteurs « primaires ». Les F_0F_1 -ATPases membranaires sont les systèmes les plus étudiés, impliquées notamment dans la pompe à protons ATPasique et les ATP Binding Cassette (ABC) transporteurs. Concrètement, chez les bactéries anaérobies comme les LAB, la FPM est très majoritairement générée par l'hydrolyse d'ATP au niveau de l'ATPase membranaire qui permet l'expulsion de protons.

La force FPM peut-être décomposée en deux termes additifs, le potentiel électrique ($\Delta\psi$) et le potentiel chimique des protons ($-Z\Delta pH$, avec $Z=2,3 \frac{RT}{F}$) à travers la membrane. De nombreuses fonctions sont conduites ou régulées par la FPM : synthèse d'ATP, transport de métabolites, translocation de protéines transmembranaires, la mobilité, activité d'enzymes autolytiques...

Chez les LAB, la FPM contribue à la majorité des transports secondaires (uniport, symport et antiport), convertissant ainsi l'énergie électrochimique du proton en énergie électrochimique pour permettre l'accumulation ou l'excrétion d'un soluté [Poolman & Konings 1993]. Une FPM peut-être également générée par l'excrétion d'un produit terminal lors d'un symport avec un proton, comme dans le cas de l'efflux de lactate chez *L. lactis* [Otto et al. 1980; Otto et al. 1982; Ten Brink et al. 1985]. Cependant, seule l'excrétion du lactate à pH basique nécessite deux protons ou plus et conduit, par la génération de FPM, à un gain énergétique durant le processus de fermentation du sucre. Ce phénomène est appelé recyclage énergétique [Michels et al. 1979].

Nutrition azotée en matrice lait

Les lactocoques ont des besoins nutritionnels complexes [Jeanson et al. 2009; Jensen 1993; Law et al. 1976] et souches-dépendants [Ayad et al. 1999]. Auxotrophes pour de nombreux acides aminés, les acides aminés branchés (BCAAs : isoleucine-leucine-valine), l'histidine, la méthionine et plus rarement l'arginine, l'acide glutamique et la proline, elles réclament une source d'azote exogène qu'elles assimilent via un processus de nutrition azotée [Monnet & Gripon 1994; Coccagn-Bousquet et al. 1995]. Cette auxotrophie peut provenir soit de l'absence de voie de biosynthèse fonctionnelle pour les acides aminés, soit de mécanismes de régulation spécifiques [Chopin 1993; Godon et al. 1993]. En particulier, il a été démontré que les besoins nutritionnels des souches laitières sont particulièrement complexes par rapport aux isolats de végétaux (Ayad et al. 1999; Garault et al. 2000; Bachmann et al. 2012). Cette perte de fonctionnalité est souvent interprétée comme une conséquence de l'adaptation des bactéries à la matrice laitière particulièrement riche en matière azotée [Ayad et al. 1999; Siezen et al. 2011]. Cependant, dans le lait, l'accès aux acides aminés est conditionné par une étape préalable d'hydrolyse des caséines puisque les concentrations d'acides aminés libres sont insuffisantes pour couvrir les besoins en acides aminés (notamment les chaînes latérales ramifiées, l'histidine et la méthionine). Cette section abordera donc la description de la cascade de réactions de la protéolyse, étape clef de la croissance en lait (pour revue Kok 1990; Doeven et al. 2005; Guédon et al. 2005; Savijoki et al. 2006), dont une représentation schématique est présentée Figure 9. Elle se terminera par la régulation des voies de synthèse des BCAAs.

La protéase de paroi, pivot de la protéolyse des caséines

La croissance est rendue possible chez les LAB se développant dans le lait par la présence d'une protéase communément appelée protéase de paroi, PrtP. En effet, chez *L. lactis*, la croissance en lait est diauxique et scindée en deux phases exponentielles successives séparées par une phase au cours de laquelle la synthèse des protéases de paroi est réalisée [Juillard et al. 1995; Juillard et al. 1998]. Au cours de la première phase, les bactéries consomment les acides aminés et peptides du milieu, tandis qu'au cours de la seconde phase, elles utilisent la caséine comme source d'azote. C'est donc la capacité à utiliser les caséines qui limite la croissance des souches. Liée de manière covalente à la paroi cellulaire la protéase PrtP hydrolyse les caséines en oligopeptides et di/tripeptides qui seront ensuite internalisés pour couvrir les besoins nutritionnels. En raison de son rôle pivot dans la protéolyse des levains industriels, des études ont visé à caractériser la protéase de *Lactococcus lactis* (pour revue Kok 1990; Doeven et al. 2005; Guédon et al. 2005; Savijoki et al. 2006).

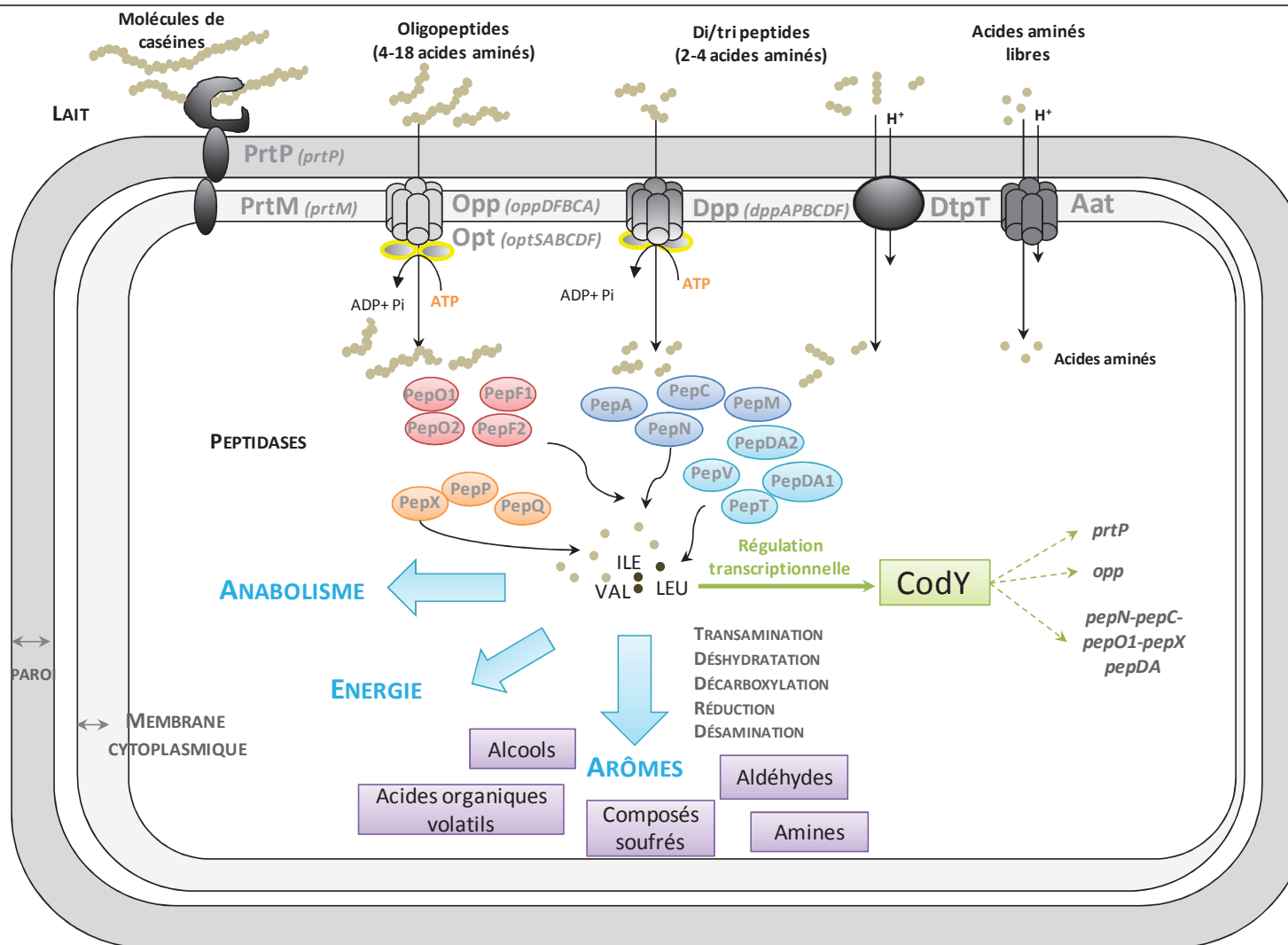


Figure 9: Représentation schématique du système protéolytique en lait et de nutrition azotée chez *Lactococcus lactis*, et des voies de régulation de l'expression génique, ainsi que du devenir des acides aminés. Adapté de [Guédon et al. 2001; Savijoki et al. 2006]. Les peptidases sont classées en quatre familles sur la base de leur spécificité de clivage : en rouge les endopeptidases, en bleu les aminopeptidases, en orange les proline peptidases, et en cyan les tri/di-peptidases. En bleu, les trois voies principales de devenir des acides aminés : anabolisme (synthèse protéique), formation d'énergie et production de molécules aromatiques. En vert, la régulation de l'expression génique du système protéolytique par le régulateur transcriptionnel CodY.

La protéase de paroi est une protéine de grande taille (200 kDa) appartenant à la famille des subtilisines (enzymes à sérine). Huit domaines conservés sont retrouvés chez les lactocoques [Siezen 1999], qui se partagent des fonctions de localisation, d'ancrage et de catalyse. La spécificité d'hydrolyse des protéases de paroi a été étudiée chez différentes souches de *L. lactis* et a révélé qu'un nombre restreint de modifications de la séquence conduisait à des différences de spécificité entre les types de caséines (α , β , κ) ou de longueur des fragments générés [Broadbent et al. 2006]. Ainsi, les protéases de type PI hydrolysent principalement les caséines β et dans une moindre mesure les caséines κ , alors que les protéases du second type, PIII, hydrolysent tous les types de caséines [Visser et al. 1986]. La spécificité des protéinases de paroi joue également un rôle sur le développement de l'amertume. En effet, ce défaut résulte de l'accumulation dans le milieu extracellulaire de peptides de petite taille, ayant une hydrophobicité élevée et dont la partie C-terminale est constituée soit de leucine, de phénylalanine ou de tyrosine [Broadbent et al. 2002]. Le choix des souches est particulièrement important pour éviter ce défaut. Le gène *prtP* code un précurseur de la protéase, alors que le gène *prtM* code une lipoprotéine impliquée dans la maturation de la protéase [de Vos et al. 1989]. Tous deux sont le plus souvent d'origine plasmidique et associés aux gènes nécessaires à l'utilisation du lactose (portés par le plasmide « lactose »).

Comme pour la majorité des composants nécessaires à la nutrition azotée, l'expression des gènes codant la protéase de paroi est négativement contrôlée en présence d'une source azotée riche. Plus précisément, au niveau transcriptionnel le régulateur transcriptionnel *CodY* va percevoir le pool d'acides aminés branchés dans la cellule, et, lorsque ce dernier est élevé, réprimer l'expression de *prtP* en se fixant sur une séquence consensus au niveau de sa région promotrice [Guédon et al. 2001].

Systèmes de transport des peptides et des acides aminés

L'utilisation des peptides, initialement présents dans le lait ou dérivés de l'hydrolyse des caséines, est indispensable à la croissance des lactocoques et requiert un système de transport spécifique [Tynkkynen et al. 1993]. Trois grandes classes de transporteurs qui se partagent l'internalisation des produits de dégradation de *PrtP* sont retrouvées sur la base de leur dépendance énergétique : i) les perméases qui fonctionnent grâce à la force protomotrice (PFM), ii) les ABC-transporteurs qui utilisent l'énergie issue de l'hydrolyse de l'ATP et iii) les antiports qui exploitent les gradients de concentration.

Les oligopeptides de 4 à 18 résidus sont la principale source d'acides aminés internalisés [Juillard et al. 1995]. Ils sont transportés par des systèmes de type ABC, dit Opp.

Les gènes *oppDFBCA* du transporteur Opp sont organisés en opéron avec le gène de l'aminopeptidase *PepO* et codent les 5 composants du transporteur :

- OppA est une protéine exocellulaire en charge de la reconnaissance de l'oligopeptide, et donc de la spécificité de transport, avant la liaison de l'oligopeptide au transporteur,
- OppB et OppC sont des protéines intra membranaires responsables de la translocation du peptide,
- OppD et OppC sont des protéines intracellulaires impliquées dans l'hydrolyse de l'ATP.

Les di/tripeptides peuvent être internalisés soit par le transporteur perméase DtpT (gène *dtpT*), si ces derniers sont de nature hydrophile et chargés, soit par le transporteur ABC Dpp lorsque les di- et tri-peptides sont hydrophobes et riches en acides aminés branchés [Hagting et al. 1994; Foucaud et al. 1995]. L'organisation génétique (opéron *dppAPBCDF*) et fonctionnelle de Dpp est similaire au transporteur oligopeptidique Opp à l'exception de la protéine de liaison non active DppP.

Les acides aminés libres peuvent être internalisés dans la cellule par une douzaine de systèmes spécifiques transportant les acides aminés de structures proches, selon les trois classes de mécanismes énergétiques (voir Tableau 3).

Tableau 3 : Systèmes de transport d'acides aminés identifiés chez les bactéries lactiques [Corrieu 2008].

Transporteur (moteur énergétique)	Acides aminés transportés
Perméase (force protomotrice)	Alanine, Glycine
	Thréonine, Sérine
	Isoleucine, Leucine, Valine
	Tyrosine, Phénylalanine
	Méthionine
	Histidine
	Lysine
ABC-transporteur (hydrolyse de l'ATP)	Glutamine, Glutamate
	Asparagine
	Proline, Glycine-bétaïne
Antiport (gradient de concentration)	Arginine/Ornithine
	Glutamate/ γ -aminobutyrate

Ce transport d'acides aminés représente un coût énergétique important pour la cellule par comparaison aux transporteurs d'oligopeptides. Un premier rôle évoqué de ces transporteurs serait d'assurer l'appoint en acides aminés spécifiques (ce qui justifierait leur forte affinité) de manière à compléter une déficience en acide aminé particulier [Flambard et al. 1998]. En complément de ce rôle nutritionnel, ces transporteurs d'acides aminés peuvent assurer différentes fonctions liées à la survie cellulaire. Par exemple, la proline, internalisé par BusA et également un transporteur hautement spécifique de la glycine bétaïne, intervient dans la résistance aux chocs osmotiques [Obis et al. 1999; Romeo et al. 2003]. De même, l'antiport glutamate/ γ -aminobutyrate participe à la

formation d'une force proton motrice par la décarboxylation du glutamate [Higuchi et al. 1997]. Une autre fonction est l'expulsion d'un acide aminé en trop forte concentration à l'intérieur de la cellule lors de l'internalisation d'un peptide, qui pourrait atteindre des concentrations toxiques s'il n'était pas éliminé.

Dégradation des peptides en acides aminés par les peptidases intracellulaires

Une fois internalisés, les peptides vont être dégradés par les peptidases intracellulaires qui diffèrent selon leurs spécificités de substrat et leurs mécanismes de clivage (voir Figure 9). Les peptidases sont appelées Pep, suivi d'une lettre indiquant leur spécificité. Les aminopeptidases hydrolysent les acides aminés aux extrémités des chaînes peptidiques (PepN et PepC sont des aminopeptidases générales alors que PepA est spécifique des acides aminés acides). Ces dernières se distinguent des endopeptidases qui hydrolysent les liaisons à l'intérieur des chaînes (PepF1-PepF2 et PepO1-PepO2). Un certain nombre de peptidases sont spécifiques des résidus prolines, qui de par leur structure particulière, sont plus difficiles à hydrolyser (PepP et PepX). Il faut noter par ailleurs que les caséines sont des protéines qui contiennent un pourcentage élevé de proline par rapport à la moyenne des autres protéines. Dernière famille de peptidases, les di/tri-peptidases sont composées de la tri-peptidase PepT et des di-peptidases PepDA et PepV.

L'étude systématique de la transcription des gènes codant les éléments du système protéolytique a permis de mettre en évidence l'existence d'un régulon *CodY* répondant au pool d'acides aminés branchés (BCAAs) (voir détails0). Les peptidases PepO, PepN, PepC, PepX et PepDA font partie de ce régulon, ce qui confirme leur importance dans le processus de nutrition azotée [Doeven et al. 2005; Gitton et al. 2005].

Catabolisme des acides aminés, production d'énergie et de molécules aromatiques

Les acides aminés présents dans le cytoplasme peuvent-être utilisés directement pour la synthèse protéique, ou être catabolisés. Le catabolisme des acides aminés est fortement impliqué dans la qualité et la sécurité des produits fermentés. En effet, les acides aminés peuvent générer des composés toxiques tels que les amines biogènes, mais représentent également une voie majeure de production de molécules aromatiques. De plus, ils peuvent également fournir de l'énergie à la bactérie.

a. Voies de production d'énergie

Deux voies principales génèrent de l'énergie lorsque la bactérie est en condition de stress, la décarboxylation des acides aminés, et le catabolisme de l'arginine. La décarboxylation des acides aminés, en particulier de l'aspartate, du glutamate, de l'histidine, de la tyrosine et du tryptophane, conduit à la formation d'une amine et à la consommation intracellulaire d'un proton, correspondant

à une alcalinisation du cytoplasme. De plus, le transport de l'amine hors de la bactérie génère un gradient électrochimique qui permet une économie d'ATP, et peut également permettre de réguler le pH intracellulaire [Abbe et al. 1982; Konings et al. 1989; Higuchi et al. 1997]. Cependant, la décarboxylation de l'histidine, de la tyrosine et du tryptophane, peut provoquer la production d'amines biogènes, composés toxiques et néfastes pour les qualités organoleptiques [Molenaar et al. 1993]. La principale voie du catabolisme de l'arginine est la voie de l'arginine déiminase (ADI). Cette voie produit de l'énergie via la production d'ATP. Elle engendre également de l'ammoniac qui contribue à la survie des bactéries à bas pH, probablement en augmentant le pH de la culture. Cette voie ADI est couplée au système antiport arginine/ornithine (ArcD1/2) qui est utilisé pour internaliser l'arginine. La voie fait intervenir 3 enzymes clefs : l'arginine déiminase (*arcA*), la carbamate kinase (*arcC*) et l'ornithine carbamoyl-transférase (*arcB*). Elle aboutit à la production de 2 moles d'ammoniac, 1 mole d'ornithine, 1 mole de CO₂ et 1 mole d'ATP par arginine métabolisée. L'arginine ainsi que le glutamate peuvent-être internalisés par le transporteur Glts (*gltS*) et suivre une voie parallèle de métabolisation pour rejoindre au niveau de la citrulline la voie ADI (Figure 10).

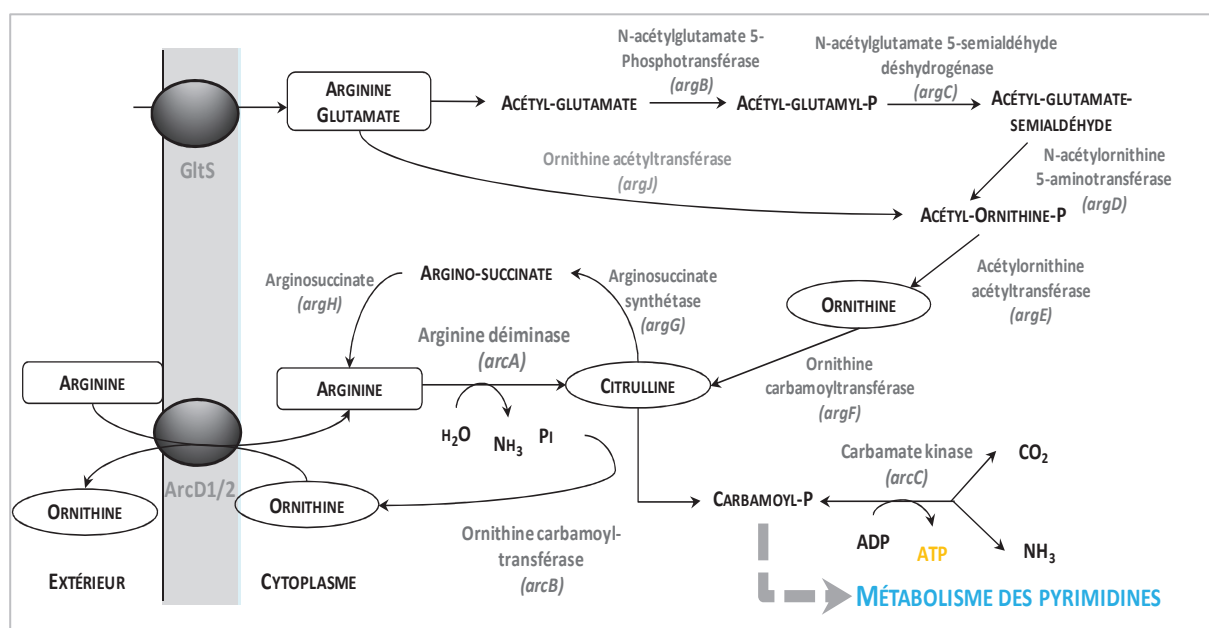


Figure 10 : Représentation schématisée du métabolisme de l'arginine chez *Lactococcus lactis*. Adapté de [Larsen et al. 2004]. Les enzymes ainsi que les noms de gènes (entre parenthèse) sont indiqués pour chaque réaction catalysée.

b. Voies de production de molécules aromatiques

Le catabolisme des acides aminés génère des molécules aromatiques que l'on peut classer en 5 familles : les aldéhydes, les alcools, les acides, les composés soufrés (notamment les thiols) et les amines. Bien que chaque molécule ait une note caractéristique, les molécules dérivant pour une grande partie des acides aminés aromatiques (AraAs) présentent une note plutôt florale, celles dérivant des acides aminés branchés (BCAAs) ont une note plutôt animale, et celles dérivant de la

méthionine possèdent une note soufrée, aillée [Smit et al. 2005]. Généralement, ces composés ont un effet bénéfique sur l'arôme du produit, mais plusieurs d'entre eux peuvent-être associés à des défauts d'arômes lorsqu'ils sont produits en trop grandes quantités [Reineccius 1998]. Le seuil de détection de ces molécules est de l'ordre du ppm, voire inférieur, ce qui explique leur fort impact même en très faible quantité. Un récapitulatif des caractéristiques de ces principaux arômes est présenté Tableau 4. Deux types de mécanismes régissent le catabolisme des acides aminés : la réaction de transamination et la réaction d'élimination de la chaîne latérale.

- La voie de transamination est initiée par les transaminases qui catalysent le transfert d'un groupement amine d'un acide aminé à un céto-acide accepteur pour produire le céto-acide correspondant à l'acide aminé de départ. Trois aminotransférases ont été caractérisées chez *L. lactis* : ArAT ayant une affinité particulière pour les acides aminés aromatiques et la méthionine, BcAT active sur les acides aminés ramifiés et la méthionine, ainsi que Asp-AT spécifique de l'aspartate [Alting et al. 1995; Yvon et al. 1997; Gao & Steele 1998; Engels et al. 2000]. Les gènes codant BcAT et ArAT sont monocistroniques et leur expression est régulée par des facteurs nutritionnels via le régulateur transcriptionnel pléiotrope CodY [Chambellon & Yvon 2003; Petranovic et al. 2004] (voir Figure 9). La transamination de l'aspartate en oxaloacétate (OAA) permet la production de pyruvate, interconnexion des voies entre métabolisme azoté et carboné, pour synthétiser notamment du diacétyle. Cette transamination nécessite la présence d'un α -céto-acide accepteur qui est généralement l' α -cétooglutarate (α -kg). L'étude de Yvon en 1998 a montré que l' α -cétooglutarate était le premier facteur limitant l'activité aminotransférase de *L. lactis* dans le fromage, puisque son addition intensifie fortement la flaveur du fromage [Yvon et al. 1998]. Deux voies fonctionnelles ont été mises en évidence dans la production d' α -kg. La première voie, mineure, est spécifique des souches du biovar *diacetylactis* qui peuvent, à partir de l'internalisation du citrate, le transformer en OAA grâce à l'action successive de la citrate perméase et de la citrate lyase [Tanous et al. 2005b]. L'Asp-AT va alors intervenir pour transférer le groupement amine du glutamate sur l'OAA, donnant de l'aspartate et de l' α -kg [Pudlik & Lolkema 2013]. Cette voie est cependant limitée par les faibles concentrations intracellulaires de citrate. La seconde voie fait intervenir la glutamate déshydrogénase (GDH) qui catalyse la désamination oxydative du glutamate en α -kg [Rijnen et al. 1999; Rijnen et al. 2000; Tanous et al. 2002]. Le gène *gdh* est porté chez *L. lactis* par le plasmide *pGdh442* incompatible avec les plasmides de type *SK08* qui portent les gènes de transport et d'utilisation du lactose et le gène *prtp* codant la protéase de paroi nécessaire à l'utilisation de la caséine, du fait de l'identité des systèmes de réplifications et partitions des 2 plasmides. Cette incompatibilité explique pourquoi la majorité des souches GDH-positives n'est pas d'origine laitière [Tanous et al. 2006].

Tableau 4 : Notes aromatiques et seuils de perception des molécules aromatiques dérivées des acides

Acide aminé précurseur	Famille chimique	Nomenclature chimique (nom usuel)	Descripteurs de l'arôme	Seuil de perception eau/huile/fromage

aminés [Corrieu 2008]. (A. : Acide)

				(ppm)
Thréonine	Aldéhyde	Acétaldéhyde	Vert, éther, âcre	1,3/0,1/-
	Acide	A. propionique	Vinaigre, aigre	40/-/-
Valine	Aldéhyde	2-méthylpropanal	Vert, malté, animal	0,1/-
	Alcool	2-méthylpropanol	Alcool, malté	-
	Acide	A. 2-méthylpropanoïque (isobutyrique)	Pomme pourrie, beurre, sueur	5,3/-/-
	Autre	Ethylbutyrate	Ananas, banane	-
Isoleucine	Aldéhyde	2-méthylbutanal	Vert, malté, animal	0,13/-/-
	Alcool	2-méthylbutanol	Malté	-
	Acide	A. 2-méthylbutanoïque	Fruité, sueur, aigre	1,6/3,2/-
Leucine	Aldéhyde	3-méthylbutanal (isovaléraldéhyde)	Vert, malté, animal	0,06/0,013/-/-
	Alcool	3-méthylbutanol	Fruité, malté, alcool	4,8/-/-
	Acide	A. 3-méthylbutanoïque (isovalérique)	Fruit pourri, sueur, pied	0,07/1/-
Phénylalanine	Aldéhyde	Phénylacétaldéhyde	Floral, rose	0,004/-/0,02
		Benzaldéhyde	Amande amère, douce	0,35/-/0,02
	Alcool	Phényléthanol	Rose, floral	1/0,2/0,07-10
	Acide	A. phénylacétique	Floral, vert	-
	Autre	Phénylacétate	Floral, rose	-
		Acétophénone	Floral, fleur d'oranger, doux	-
Tyrosine	Acide	A. hydroxyphénylacétique	Floral, vert	-
	Autre	p-crésol	Médical	-/-/0,3
		Phénol	Floral, médical	1-10/0,15/0,01
Tryptophane	Autre composé	Stackol	-	-
		Indole	Putride	0,14/0,02
Méthionine	Aldéhyde	3-méthylthiopropional (méthional)	Pomme de terre bouillie	-/0,2/60
		Méthylthioacétaldéhyde	Vert, pomme	-
	Alcool	3-méthylthiopropanol	Pomme de terre crue	-
				-
	Acide Thiols et autres composés	A. méthylthiopropionique	-	-
		Méthanethiol	Choux cuit, ail	0,002/0,0006/-
		Diméthyldisulfide (DMDS)	Choux-fleur, fromage affiné	0,012-0,12/-/-
		Diméthyltrisulfide (DMTS)	Ail, fromage trop affiné, putride	-/0,025/0,0001
		Diméthylsulfide (DMS)	Choux, souffré, ail	-/0,01-0,17/0,019
		2-4-dithiapentane	Ail	-/-/0,06
		Méthylthioacétate	Choix-fleur cuit	-/-/0,05
Asparagine, Glycine, Alanine et Sérine	Acide	Acide Acétique	Vinaigre, aigre, âcre	22-54/3-7/-

[THESE CONFIDENTIELLE]

Les céto-acides ainsi produits sont le point de départ de 4 réactions enzymatiques ou chimiques :

-**Réduction en hydroxyacides**, molécules non aromatiques, par des hydroxyisocaproique acides désydrégénases (HicDH) qui peuvent régénérer le NAD^+ .

-**Décarboxylation non-oxydative** par des céto-acides décarboxylases qui donnent des aldéhydes. Ces dernières peuvent-être oxydées en acides par les aldéhydes déshydrogénases ou réduits en alcools par les alcools déshydrogénases. Ces deux voies utilisent le cofacteur NADH. Aldéhydes, acides et alcools constituent un pool de molécules fortement aromatiques (Tableau 4).

-**Décarboxylation oxydative** par des α -céto acides déshydrogénases NAD dépendantes qui produisent des acides carboxyliques via des acyl-CoAs. De plus, cette voie peut générer de l'ATP, tout comme le catabolisme du pyruvate en acétate. Les lactocoques sont également capables de produire des esters à partir des acides (ou acyl-CoA) et d'alcools grâce à l'estérase *EstA* notamment. Cependant, cette activité est souche-dépendante [Liu et al. 1998; Holland et al. 2005a]. Par ailleurs, les acides indolacétique et hydroxyphenylacétique, dérivés du tryptophane et de la tyrosine, peuvent-être spontanément dégradés en skatole et p-crésol, composés impliqués dans les défauts organoleptiques de certains fromages.

-**Transformation chimique** des α -céto acides, et notamment l'oxydation de ceux dérivés de la phénylalanine, leucine et méthionine, qui produit des molécules fortement aromatiques (Tableau 4). Il est à noter que toutes ces voies décrites impliquent de nombreuses réactions d'oxydoréduction et de régénération de co-facteurs. Ainsi ces réactions métaboliques sont-elles fortement dépendantes des conditions d'oxydoréduction du milieu. Il a été démontré que des conditions oxydantes favorisent la production d'aldéhydes et des composés soufrés, alors que les conditions réductrices tendent à stimuler la production d'acides carboxyliques et d'hydroxyacides [Kieronczyk et al. 2006].

- La production d'arômes par les réactions d'élimination se fait en une seule étape chez les LAB. Ces réactions sont catalysées par des lyases, spécifiques d'un acide aminé (Figure 11). Elles touchent essentiellement la méthionine, la thréonine et plus faiblement le tryptophane et la tyrosine. Deux enzymes, capables de produire du méthane-thiol directement à partir de la L-méthionine, ont été isolées et caractérisées chez *L. lactis* ssp. *cremoris*. Il s'agit de la cystathionine γ -lyase (CGL) isolée à partir de la souche *SK11* [Bruinenberg et al. 1997], et de la cystathionine β -lyase (CBL) isolée à partir de la souche *B78* [Alting et al. 1995]. Ces enzymes catalysent respectivement l'élimination en α - β et α - γ de la cystathionine, en produisant de la cystéine ou de l'homocystéine en plus de l' α -céto-butyrat et de l'ammoniac. Le méthane-thiol pourra être par la suite oxydé en diméthyl-disulfure (DMS) et/ou en diméthyl-trisulfure (DMTS). Le catabolisme de la thréonine est initié par une thréonine aldolase (TA) qui convertit cette dernière en glycine puis acétaldéhyde. L'acétaldéhyde, composé majeur des yaourts et également des fromages, peut également être

converti en éthanol via une alcool déshydrogénase [Marshall & Cole 1983]. Le tryptophane et la tyrosine sont également les substrats respectifs de la tryptophane indole lyase (TIL) et de la tyrosine phénol lyase (TPL) qui produisent indole et phénol.

Le catabolisme des acides aminés est une voie majeure pour la formation d'arômes chez L. lactis et par conséquent, son contrôle est une piste pour intensifier et diversifier l'arôme des produits fermentés. Certains aspects étant souches-dépendants, la sélection des levains doit tenir compte de cette diversité pour sélectionner et associer au mieux les souches.

Synthèse des acides aminés, le cas des acides aminés branchés

Les LAB ont besoin du système protéolytique décrit précédemment pour croître en lait. Cependant, les études des différents génomes de *Lactococcus lactis* ont permis de caractériser trois clusters de gènes impliqués dans la biosynthèse du tryptophane [Bardowski et al. 1992], de l'histidine [Delorme et al. 1992; Delorme et al. 1993], et des acides aminés branchés, BCAAs [Godon et al. 1992].

En raison du rôle prépondérant des BCAAs dans la régulation de la nutrition azotée, la génération d'arômes et dans l'interconnexion avec le nœud du pyruvate, la régulation de la synthèse de ces acides aminés va être détaillée.

Les principales étapes de biosynthèse des acides aminés branchés, à partir du pyruvate et de la thréonine, nécessitent 4 enzymes communes suffisantes pour la synthèse de la valine, alors que respectivement une et trois enzymes supplémentaires sont nécessaires pour l'isoleucine et la leucine (voir Figure 12) pour revue, [Chopin 1993]. Les gènes codant ces enzymes sont organisés en opéron *leuABCD-ilvDBNCA* avec les gènes *aldB* codant l'acétolactate décarboxylase et *aldR*, codant le régulateur transcriptionnel AldR. Ils sont transcrits sous le contrôle de trois promoteurs, P₁, P₂ et P₃, régulés de manière différentielle. P₁ et P₂ sont réprimés par CodY en présence d'isoleucine, et P₁ est, de plus, soumis à un deuxième contrôle de l'élongation de la transcription par un mécanisme de terminaison-anti terminaison modulé en fonction de la traduction d'un peptide riche en leucine [Godon et al. 1992; Godon et al. 1993; Goupil et al. 1996]. Le promoteur P₃ initiant la transcription de l'ALDC est réprimé par la réponse stringente induite par un arrêt de traduction [Goupil-Feuillerat et al. 2000] (pour détails voir 0). Un autre point de contrôle de cette synthèse est une régulation allostérique de l'ALDC codée par *aldB* et dégradant l'acétolactate, précurseur de la valine et de la leucine.

En présence de leucine, ALDC est activée et le flux d'acétolactate est dirigé vers la synthèse d'acétoïne/butanediol [Goupil-Feuillerat et al. 1997] puisque ce précurseur n'est pas nécessaire à la synthèse de leucine et de valine déjà présente dans le milieu (Goupil-Feuillerat et al. 1997). En plus d'un contrôle de son activité par la leucine, l'ALDC est contrôlée, au niveau de son expression, par un

mécanisme d'atténuation traductionnelle sur les transcrits initiés aux promoteurs P1 et P2 qui bloque la fixation des ribosomes susceptibles de traduire *aldB* par la formation de structures tiges-boucles.

Le résultat de cet ensemble de mécanismes est que la croissance de *L. lactis* est inhibée dans un milieu où l'apport des acides aminés branchés est déséquilibré. Ainsi en présence d'isoleucine, la faible transcription de l'opéron *leu-ilv* entraîne une synthèse limitante de leucine et valine. En présence d'excès de leucine, la synthèse de valine est arrêtée provoquant une inhibition de la croissance dans un milieu carencé en cet acide aminé.

Régulation du métabolisme azoté

Le métabolisme azoté et l'approvisionnement en acides aminés sont essentiels pour une croissance optimale des bactéries. Cet apport, comme nous venons de le voir, peut se faire soit par l'import d'une source exogène d'acides aminés, soit en synthétisant les acides aminés à partir d'une source azotée plus simple. Cependant, l'assimilation de nutriments azotés est la voie énergétique la plus favorable pour assurer la croissance des bactéries lactiques. Ainsi, les gènes impliqués dans ces deux voies métaboliques sont sujets à des régulations transcriptionnelles afin de moduler leur expression selon les conditions environnementales. Chez *L. lactis*, cette régulation peut-être effectuée par l'intermédiaire i) de régulateurs spécifiques, ii) de la réponse stringente et iii) du régulateur pléiotropique CodY.

- Les régulateurs spécifiques sont des protéines de liaisons à l'ADN qui interviennent dans la régulation de la transcription. *GlnR* est le régulateur transcriptionnel de la glutamine synthase (opéron *glnRA*), mais également de l'opéron *glnPQ* (glutamine ABC transporteur) [Larsen et al. 2006]. Un motif conservé est reconnu sur l'ADN, 5'-TGTNA-7N-TNACAT-3', en réponse aux variations de concentrations exocellulaires de glutamine et d'ammonium. *ArgR* et *AhrC* sont des régulateurs impliqués dans la biosynthèse et le catabolisme de l'arginine [Larsen et al. 2004]. En effet, les trois opérons de biosynthèse de l'arginine *argCJDBF*, *argGH*, et *gltS-argE* sont réprimés par les produits d'*argR* et d'*ahrC*. De plus, l'opéron catabolique de l'arginine *arcABD1C1C2TD2* est activé par la présence d'*AhrC* mais non par *ArgR*. Il a été démontré également que les deux régulateurs sont interdépendants et incapables de se compléter l'un l'autre [Larsen et al. 2004]. *AldR* est un activateur de la transcription de l'opéron *ilvBNCD*A qui code la synthèse des acides aminés branchés (voir 0) [Goupil-Feuillerat et al. 1997]. *FhuR* (=CmbR) est un activateur transcriptionnel de la famille Lys-R impliqué dans la régulation de 18 gènes du métabolisme de la cystéine et de la méthionine [Sperandio et al. 2005]. Un site de liaison putatif a pu être identifié sur l'ADN des gènes du régulon, 5'-TGATA-N(9)-TATCA-N(2-4)-TGATA-3' [Kovaleva & Gelfand 2007].

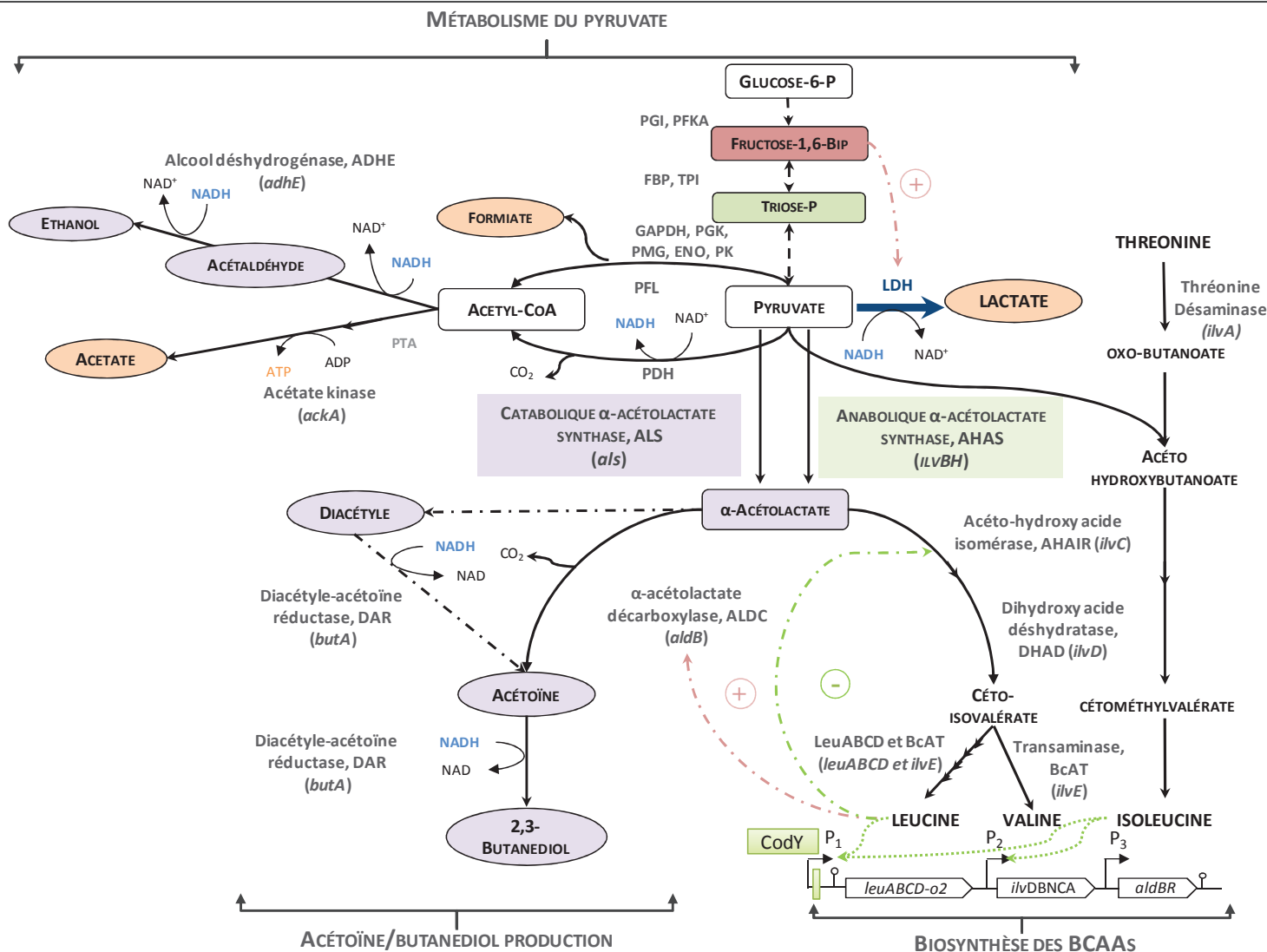


Figure 12 : Interconnexion des voies du métabolisme du pyruvate, de production d'acétoïne/butanediol et de la régulation de la biosynthèse des acides aminés branchés. Adapté de [Goupil-Feuillerat et al. 1997]. Sur fond violet, les composés responsables d'arômes dans les produits laitiers, et en orange les non aromatiques. L'opéron *leu-ilv* est représenté dans la partie basse de la figure. Les flèches en pointillé matérialisent les actions régulatrices transcriptionnelles du régulateur CodY, et les flèches en tirets-points les régulations allostériques (en vert inhibition et en rouge activation).

▪ La réponse stringente a été initialement caractérisée lors de l'état de carence en acides aminés de la cellule comme un échec de l'estérification des acides aminés sur l'extrémité 3' des ARNt, qui conduit à l'insatisfaction des besoins de synthèse protéique [Cashel 1969]. Cependant, les études ultérieures ont mis en évidence le rôle pléiotropique de cette réponse (plus de 100 gènes identifiés) lors de l'adaptation à différents stress : carence carbone et stress oxydatif chez *B. subtilis* et *E. coli* [Wendrich & Marahiel 1997; Eymann et al. 2002; Chang et al. 2002] ou encore acide chez *Lactococcus* [Rallu et al. 1996; Budin-Verneuil et al. 2007]. Elle a été essentiellement caractérisée suite à l'étude de la carence en acides aminés ou d'ajout de norvaline, inhibiteur analogue à la valine (Eymann et al. 2002), par des approches combinées de biochimie, de transcriptomique et de protéomique [Eymann et al. 2002; Chang et al. 2002; Nyström 2004; Magnusson et al. 2005; Durfee et al. 2008; Traxler et al. 2008]. En particulier, deux études à l'échelle *omique* ont étudié la réponse stringente chez *Lactococcus lactis* [Dressaire et al. 2008; Dressaire et al. 2011].

La réponse stringente implique une reprogrammation globale des processus cellulaires et en particulier de l'expression génétique dans la cellule. D'une part, il est observé un contrôle négatif des gènes associés à la croissance (ARN ribosomiques et ARN de transfert, synthèse des acides gras et lipides membranaires, réplication, métabolisme des purines et pyrimidines, transcription par l'ARN polymérase ainsi que la synthèse des facteurs d'initiation et d'élongation de la traduction) [Cashel et al. 1997; Magnusson et al. 2005; Baker et al. 2012]. D'autre part, les gènes de biosynthèse des acides aminés (*His*, *Arg*, *Cys*, *Ilv*) et de réponses aux stress vont être surexprimés (pour revue [Nyström 2004; Srivatsan & Wang 2008; Wolz et al. 2010; Kanjee et al. 2012]). Cependant, il est communément admis que le principal effet sur la physiologie cellulaire intervient au niveau du contrôle négatif de la transcription des ARN dits stables (ribosomaux et de transfert) [Cashel et al. 1997; Paul et al. 2004; Dennis et al. 2004; Jin et al. 2012; Srivatsan & Wang 2008; Kanjee et al. 2012].

La réponse à la carence en acides aminés est médiée par l'accumulation des nucléotides guanosine 3', 5'-diphosphate (ppGpp) et guanosine 3'-diphosphate, 5'-triphosphate (pppGpp) (abréviation commune (p)ppGpp) d'un facteur 100 au-dessus du niveau basal [Cashel 1969; Baracchini & Bremer 1988; Cashel et al. 1997]. Alors que chez *E. coli*, l'activité de synthèse de (p)ppGpp à partir de GTP, de GDP et d'ATP est assurée par deux enzymes, la (p)ppGpp synthétase I ou RelA, et la (p)ppGpp synthétase II (SpoT) [Cashel et al. 1997; Heinemeyer & Richter 1978; Hernandez & Bremer 1991], seule l'enzyme RelA, associée à la sous-unité 70S du ribosome, a été identifiée chez *L. lactis* [Bolotin et al. 2001]. Le signal déclencheur de synthèse de (p)ppGpp par la protéine RelA, est l'accumulation dans la cellule d'ARNt non chargés qui vont se positionner sur le site A du ribosome en réponse à la pause traductionnelle de ce dernier [Wendrich et al. 2002]. La protéine ribosomique L11 régule

l'activité de RelA [Yang & Ishiguro 2001]. Le mécanisme de l'effet inhibiteur de (p)ppGpp sur la synthèse d'ARN a été principalement investigué chez *E. coli* mais reste néanmoins partiellement inconnu. Cependant, à la différence des principaux régulateurs transcriptionnels qui se fixent à l'ADN, le (p)ppGpp interagit en se fixant aux sous-unités de l'ARN polymérase (*rpoLB*) pour moduler son activité et sa spécificité [Chatterji et al. 1998]. Les récentes études biochimiques, génétiques et de cristallographie ont situé le site de liaison du (p)ppGpp à l'interface des sous-unités β' et ω de l'ARN polymérase et à distance du centre catalytique et de la liaison à l'acide nucléique [Zuo et al. 2013; Ross et al. 2013]. Il a été également démontré que la fixation de la protéine DksA dans le canal de l'ARN polymérase à plus de 30 Å du site ppGpp [Mechold et al. 2013; Ross et al. 2013; Zuo et al. 2013] augmentait le contrôle de la transcription par le (p)ppGpp [Cashel 1969; Paul et al. 2004]. Il est avancé que la liaison des deux sous-unités de l'ARN polymérase par (p)ppGpp/DksA pourrait restreindre les mouvements pivotants de l'ARN polymérase et bloquer la fermeture du centre actif. Par conséquent, les cycles d'addition de nucléotides seraient ralentis et les complexes ouverts d'initiation de la transcription du promoteur P1 des ARNr seraient déstabilisés [Barker et al. 2001; Dennis et al. 2004]. De plus, les séquences en amont du point +1 de la transcription sur les promoteurs P1 et P2 des *rrn* à -35 et à -10 contribuent à réduire la demi-vie des complexes ARN polymérase-ARNr [Travers 1980; Travers 1984; Travers 1988].

Deux mécanismes vraisemblablement concomitant ont été postulés pour expliquer la régulation positive assurée par le (p)ppGpp. Le premier s'appuie sur une approche cinétique indirecte de relocalisation de l'ARN polymérase active et libre [Dennis et al. 2004; Ehrenberg et al. 2010; Baker et al. 2012; Ehrenberg et al. 2013]. En effet, la disponibilité accrue de l'ARN polymérase active libérée des promoteurs stringents (promoteurs P1 et P2 des *rrn*) expliquerait la fixation renforcée sur les promoteurs faibles (comme les promoteurs d'opérons des acides aminés) et donc une initiation préférentielle de la transcription de ces derniers [Barker et al. 2001]. Le second modèle est basé sur la modulation de la spécificité de l'activité de l'ARN polymérase envers certains promoteurs en fonction du recrutement du facteur σ de la transcription [Ishihama 2000]. En présence de (p)ppGpp, le recrutement par l'ARN polymérase de facteurs σ dits alternatifs, notamment $\sigma^{54/N}$ (impliqué dans la croissance à faible concentration en azote), σ^{38} (en réponse aux stress, à l'adaptation à la phase stationnaire et à la survie) et σ^{32} (gènes exprimés suite à un choc thermique), est augmenté (affinité accrue et synthèse augmentée de ces facteurs) [Jishage et al. 2002]. Ainsi, les gènes associés à la croissance qui nécessitent une quantité importante de σ^{70} seraient négativement régulés à la différence de l'expression de gènes sous contrôle de facteurs σ alternatifs tels que les gènes de biosynthèse d'acides aminés ou de réponses au stress qui pourraient-être transcrits [Magnusson et

al. 2005]. En l'absence de facteurs σ chez *L. lactis*, seul le modèle de relocalisation cinétique de l'ARN polymérase pourrait expliquer la réorientation transcriptionnelle observée.

- Le dernier acteur de la régulation du métabolisme azoté est la protéine répresseur CodY [Sonenshein 2005; Guédon et al. 2005a; Stenz et al. 2011]. Plus de 100 gènes cibles ont été identifiés pour ce régulateur pléiotropique de la transcription des bactéries Gram-positives. Le régulon CodY comprend l'opéron *ilv-leu* de la synthèse des BCAAs, les gènes du catabolisme des acides aminés, de la protéolyse (protéase *prtP*), du transporteur des peptides (opérons *dpp* et *opp*) et de dégradation des peptides par les peptidases. Un motif de fixation de ce régulateur, 5'-AATTTTCNGAAAATT-3' appelé CodY-box, a été déterminé en comparant les zones promotrices du régulon de CodY [den Hengst et al. 2005; Guédon et al. 2005].

L'activité de CodY est contrôlée par la concentration intracellulaire du pool d'acides aminés branchés [Guédon et al. 2005a; Dressaire et al. 2011]. En effet, en carence en BCAAs, il y a une dérepression transcriptionnelle des gènes cibles de CodY par détachement de CodY de l'ADN. D'un point de vue mécanistique, il a été découvert chez *B. subtilis* que la liaison du BCAA sur l'extrémité N-Terminale du domaine GAF de CodY induit un changement de conformation qui augmente l'affinité de CodY pour la CodY-box sur l'ADN [den Hengst et al. 2005; Levnikov et al. 2006; Levnikov et al. 2009]. Ainsi, CodY agit comme un senseur nutritionnel pour permettre l'adaptation de la bactérie à son environnement nutritionnel.

Lipolyse et catabolisme des acides gras en lait

Dernière voie métabolique impliquée dans l'aromatisation des produits laitiers, la dégradation de la matière grasse, ou lipolyse, joue un rôle essentiel dans l'aromatisation. En effet, la fabrication de produits laitiers maigres, bien que soutenue par une demande forte du marché, ne satisfait pas toujours les exigences organoleptiques [Foda et al. 1974; Milo & Reineccius 1997; Wijesundera & Drury 1999]. Phénomène particulièrement important dans le procédé de fabrication des fromages affinés, les études se sont concentrées sur la compréhension de ce mécanisme afin de le maîtriser. En effet, un excès de lipolyse conduit à un goût rance, et son absence à une perte drastique d'aromatisation. Souvent sous-estimée au profit de l'activité de protéolyse, la lipolyse a été mise en évidence dans les laits fermentés par les LAB, comme le prouvent les différents inventaires de molécules détectées [Ott et al. 1997; Cheng 2010]. Les LAB réduisant fortement le milieu lors de la croissance dans le lait, les réactions d'oxydation chimique des lipides du lait sont considérées comme mineures dans la dégradation de la matière grasse. A l'inverse, l'hydrolyse enzymatique de la matière grasse libère des acides gras qui participent soit par eux-mêmes, soit par les composés issus de leur catabolisme, aux caractéristiques sensorielles des produits [McSweeney & Sousa 2000; Smit et al. 2005; Ott et al. 1997; Collins et al. 2003]. Cette section introduira les deux types d'enzymes à l'origine du métabolisme lipidique, les estérases et les lipases, en présentant leurs origines, leurs fonctionnalités ainsi que l'intérêt technologique qu'elles représentent au niveau du développement d'arômes.

Les agents lipolytiques dans le lait et les produits laitiers

La lipolyse est due à la présence d'enzymes lipolytiques définies comme des carboxylestérases qui hydrolysent les liaisons esters entre les acides gras et le glycérol au sein des triglycérides. Elles produisent des acides gras libres, du glycérol et, selon les cas, des mono, ou diglycérides [Deeth & Touch 2000]. Les enzymes lipolytiques sont classées en fonction de leur spécificité envers le substrat en tenant compte des éléments suivants : i) la position sur le corps glycérol de la chaîne à hydrolyser, ii) la longueur de l'acide gras, ou iii) la stéréospécificité [Jensen et al. 1990; Jensen 1983; Jaeger et al. 1994]. Ainsi, selon la longueur de la chaîne des acides gras, est-il distingué les estérases ou carboxylases (EC 3.1.1.1) qui hydrolysent des acylglycérols aux chaînes comprises entre 2 et 8 carbones, et les lipases ou triacylglycérol acylhydrolases (EC 3.1.1.3) impliquées dans la libération d'acide gras de plus de 10 carbones. Autre différence, les estérases sont actives en solution aqueuse alors que les lipases hydrolysent préférentiellement à l'interface des globules lipide-eau. Les enzymes lipolytiques ont été particulièrement étudiées dans la cadre de l'affinage des fromages à pâte

pressée comme l'emmental, le beaufort, le salers ou le cheddar, et trois sources possibles de lipases ont été identifiées: i) la présence endogène dans le lait, ii) l'ajout de lipase exogène ou la contamination de la présure par des lipases gastriques et iii) leur production par les différents microorganismes (levains lactiques ou la flore microbienne d'affinage) [McSweeney & Sousa 2000].

a. Lipoprotéine lipase du lait

Bien qu'il existe des activités estérases dans le lait [Chen 2000], les données de la littérature se sont essentiellement focalisées sur la lipase lipoprotéine du lait ou LPL. C'est un dimère de chaînes glycoprotéiques qui hydrolyse spécifiquement les positions 1,3 sur les triglycérides. Synthétisée par les glandes mammaires, le niveau de LPL dans le lait dépend de la période de lactation, de l'alimentation et de la saison, mais est compris en moyenne entre 10 et 20 nM [Mc Sweeney 2004]. Associée à près de 90% avec la caséine, la LPL est physiquement dissociée de son substrat présent dans les globules lipidiques, ce qui réduit considérablement son activité. Cependant, si la membrane lipoprotéique est endommagée (brassage excessif du lait, moussage), les triglycérides sont libérés et il peut résulter de l'action de la LPL le développement d'un goût « rance » du lait. La pasteurisation du lait inactive à près de 95% l'activité de la LPL endogène, ce qui permet de considérer son action comme nulle au sein des produits pasteurisés, à la différence des produits au lait cru.

L'utilisation de présure étant caractéristique des procédés fromagers, nous ne détaillerons pas l'action des lipases gastriques.

b. Les activités lipolytiques des estérases des bactéries lactiques

Les LAB, et spécifiquement les genres *Lactococcus* et *Lactobacillus*, sont considérées comme des bactéries faiblement lipolytiques par comparaison à d'autres genres bactériens (*Pseudomonas*, *Bacillus*, *Flavobacterium* et *Achromobacter*) [Andersson et al. 1979; Kalogridou-Vassiliadou 1984; Deeth & Touch 2000; Chen et al. 2003]. Cependant, des activités estérases/lipases d'hydrolyse d'esters ont pu être mesurées à de nombreuses reprises chez les LAB [Harper et al. 1980; Kamaly & Marth 1989; Kamaly et al. 1990; Tsakalidou & Kalantzopoulos 1992; Katz et al. 2002]. De plus, du fait de leur présence à haute concentration en tant que levains ainsi que de leur persistance durant les phases d'affinage des fromages, les LAB sont considérées comme responsables de la libération d'un niveau significatif d'acides gras.

Une seule et unique estérase, EstA, a été identifiée et purifiée chez *Lactococcus lactis* [Tsakalidou & Kalantzopoulos 1992; Fernández et al. 2000; Nardi et al. 2002]. Ces résultats sont corroborés aussi bien par les données génétiques, un seul gène homologue à des estérases ayant été détecté [Bolotin et al. 2001], que par la construction de mutants déficients en *estA* [Fernández et al. 2000; Nardi et al.

2002]. De plus, aucune lipase n'a pu être isolée, en accord avec les données des génomes des lactocoques.

Trois points communs se dégagent pour les estérases des LAB :

- Elles hydrolysent préférentiellement des esters aux chaînes d'acides gras en C₄ et C₆.
- Elles sont inhibées par des inhibiteurs d'enzymes à sérine, ce qui permet de supposer la présence de sérine dans leur site catalytique et leur appartenance à la famille des sérines hydrolases.
- Il a été également démontré que les estérases des lactocoques, sont sélectives des mono- et diglycérides et ne peuvent hydrolyser la matière grasse du lait lorsque cette dernière est intacte [Liu et al. 2001; Holland et al. 2005].

Elles sont décrites comme majoritairement intracellulaires, ce qui implique une étape préliminaire de lyse des cellules pour être relarguées dans la matrice laitière et rentrer en contact avec la matière grasse du lait. Les travaux menés sur la souche *L. lactis* ssp. *cremoris* AM2 ont démontré que les niveaux en acides gras libres (octanoïque et hexanoïque) étaient plus importants sur ce type de souche autolytique par rapport à un levain classique, mettant ainsi en évidence un lien entre autolyse et lipolyse

D'un point de vue physiologique, le rôle des estérases n'est pas encore élucidé, tout comme l'expression et la régulation des gènes concernés. Il a pu, cependant, être démontré par l'analyse du comportement des mutants déficients en *estA* chez *L. lactis* ssp. *lactis*, que l'estérase ne participe à la croissance ni en lait, ni sur milieu synthétique. L'une des hypothèses avancées serait que les estérases jouent un rôle structurel dans la maintenance de la membrane du fait de l'association de la tributyrine estérase à la membrane de *L. lactis* ssp. *cremoris* E8 [Holland & Coolbear 1996]. Un rôle de détoxification vis-à-vis du formaldéhyde a également été avancé sur la base d'homologie des estérases aux enzymes de la famille des S-formylglutathione hydrolases [Harms et al. 1996].

En supplément de leur activité d'hydrolyse, les estérases sont également capables de catalyser la réaction réverse d'estérification. Ces deux activités antagonistes sont régulées par la cellule en fonction des paramètres environnementaux (concentration en substrat, activité de l'eau, température et pH). Là encore, cette double activité n'a pu être explicitée d'un point de vue physiologique mais joue un rôle technologique important avec la formation d'esters aux seuils de détection olfactif bas et aux notes essentiellement florales.

Le catabolisme des acides gras

La composition des acides gras libérés dépend à la fois de la nature des triglycérides (essentiellement les acides hexanoïque et octanoïque qui représentent respectivement 25 et 27 % du total des lipides [Gunstone et al. 2010]) et de la spécificité d'hydrolyse des estérases. Les acides gras libres

participent, soit par eux-mêmes, soit comme précurseurs, aux caractéristiques sensorielles des produits [Collins et al. 2003].

Les proportions des acides gras libres ($C_{6:0}$ - $C_{18:3}$) sont similaires à celles de la matière grasse du lait à l'exception de l'acide butyrique (C_4) retrouvé en concentrations plus importantes. Tous ces acides gras libres à chaînes courtes ou intermédiaires ont une note olfactive rance, acre, grasseuse, fromagère ou encore dite lustrée, avec seuil olfactif de détection relativement haut (de l'ordre du ppm). Les travaux de Güler ont notamment montré une corrélation dans les yaourts salés entre la quantité d'acides gras libres, d'acides octanoïque et 9-décenoïque, et le descripteur « savonneux » [Güler 2007]. Comparativement, les produits du catabolisme des acides gras obtenus par une série de réactions (voir Figure 13) ont des seuils de détection plus faibles (de l'ordre du ppb pour les esters), ce qui impacte d'autant plus les conséquences aromatiques.

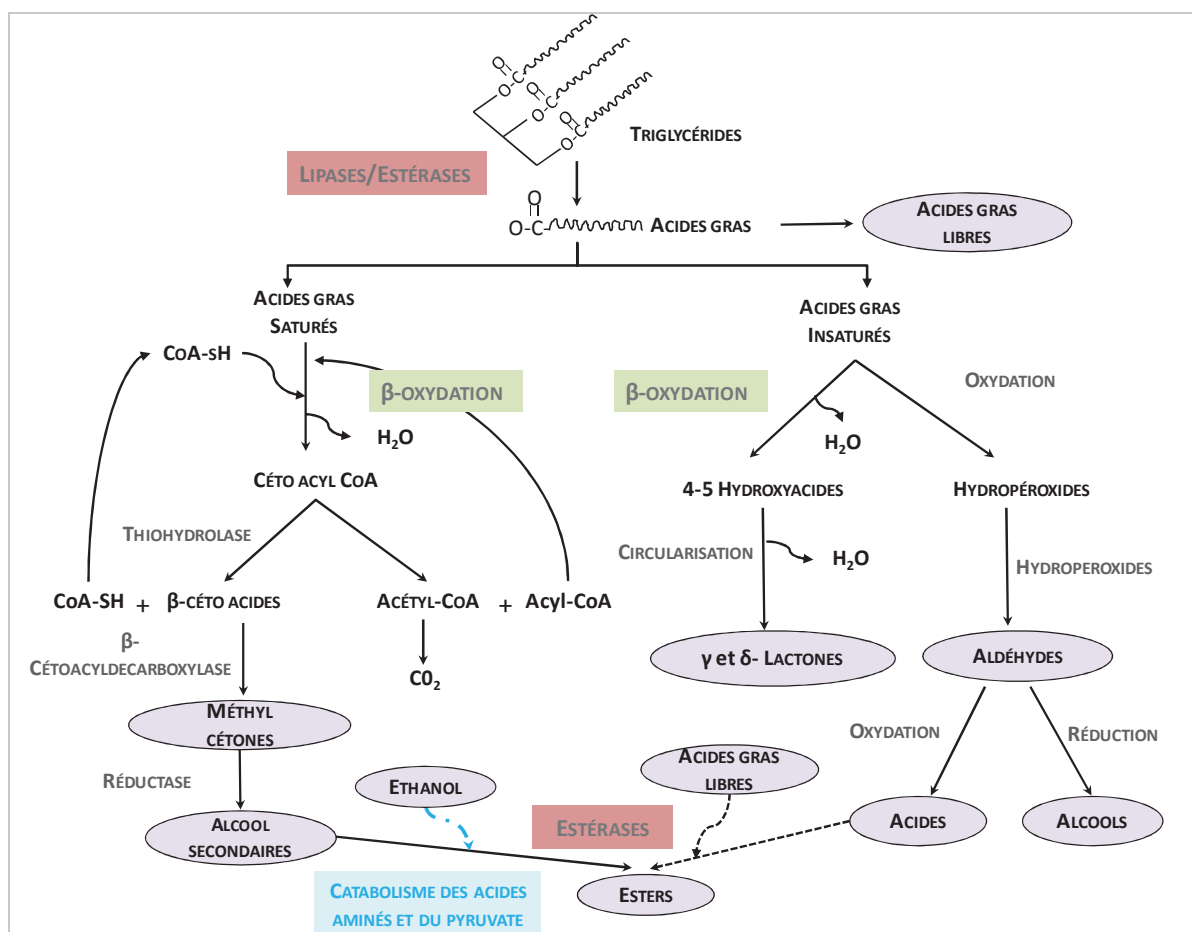


Figure 13 : Schéma général de la dégradation de la matière grasse du lait et du catabolisme des acides gras libres. Adapté et modifié d'après [McSweeney & Sousa 2000; Cheng 2010].

Les acides gras saturés peuvent subir une étape de β -oxydation pour conduire à la production de 2-méthyl cétones après décarboxylation. Cette voie est prépondérante chez les moisissures (*P. roqueforti*, *P. camemberti* ou *G. candidum*). Le taux de conversion en méthyl-cétones est fonction de nombreux facteurs (température, pH : optimal entre 5 et 7, et la concentration en acides gras libres).

Bien que 11 types de méthyl-cétones aient été identifiés, les plus retrouvées sont la pentane-2-one, l'heptane-2-one, et la nonan-2-one. Elles peuvent par la suite être réduites en alcools secondaires correspondants (pentan-2-ol, heptan-2-ol et nonan-2-ol) aux notes vertes et herbeuses.

Les lactones sont des composés cycliques formés à partir de l'estérification intramoléculaire des hydroxyacides. A la fois les γ - et δ -lactones (respectivement cycle à cinq et six carbones) ont été détectées, elles contribuent aux aspects fruités et floraux des produits. La γ -et la δ -dodecalactone ont été régulièrement identifiées dans les laits fermentés [Ott et al. 1997]. Acides et alcools peuvent être également synthétisés par oxydation et réduction des aldéhydes formées à partir des hydroperoxydes. L'estérification des acides gras libres produits par les lipases et les enzymes du catabolisme des lipides peut, avec les alcools, permettre la production de multiples esters. Alors que des méthyles, propyles et butyles esters peuvent-être produits, la majorité des esters sont des éthyles esters de raison de l'alcool majoritaire, l'éthanol. Ce dernier peut provenir du catabolisme du pyruvate ou des acides aminés. Les éthyles esters apportent des caractéristiques plaisantes aux produits, et sont identifiés avec les descripteurs sucrés et fruités [McSweeney & Sousa 2000; Mariaca et al. 2001; Fernández-García et al. 2002]. Cheng, dans sa revue sur les laits fermentés, répertorie la présence d'éthyle acétate, butanoate, hexanoate, octanoate et butyle acétate dans les laits fermentés par les LAB [Cheng 2010].

Tableau 5 : Notes aromatiques et seuils de perception des principales molécules aromatiques dérivées de la dégradation de la matière grasse retrouvées dans les laits fermentés. [Ott et al. 1997; McSweeney & Sousa

Nature du	Nomenclature	Descripteurs de l'arôme	Ordre de grandeur

2000; Cheng 2010]. (A. : Acide)

composé	(nom usuel)		du seuil de perception (ppm)
Acides	A. butanoïque (butyrique)	Aigre, rance, putride, fromage	1-100
	A. hexanoïque	Rance, floral	
	A. octanoïque	Cireux, savonneux, musqué, rance	
	A. decanoïque	Rance	
Cétones et Lactones	2-pentanone	Fruité, acétone	0,1-10
	2-hexanone	Floral, fruité	
	2-nonanone	Vert, herbeux, fruité	
	2-undacanone	Floral, rose, herbeux	
	δ-dodecalactone	Pêche, noix de coco, beurré, musqué	10 ⁻² -1
	γ-dodecalactone	Pêche, prune, noix de coco, beurré	
Alcools	2-pentanol	Alcool	1-50
	2-heptanol	Terreux, huileux	
	2-octanol	Vert	
	Oct-1-en-3-ol	Champignon	
Esters	Ethyle acétate	Solvant, fruité, ananas	10 ⁻³
	Ethyle butanoate	Fruité, sucre, banane	
	Ethyle hexanoate	Fruité, pomme, banane	
	Ethyle octanoate	Fruité, banane, pomme	
	Butyle acétate	Ananas	

Métabolisme des bases puriques, pyrimidiques et des nucléotides

Le métabolisme des nucléotides joue un rôle fondamental dans la vie cellulaire. Les nucléotides sont des sources d'énergie nécessaires à de nombreuses réactions. L'ATP est la source d'énergie la plus couramment rencontrée, mais le GTP est aussi employé dans la synthèse de protéines et dans d'autres réactions. L'UTP permet l'activation du glucose et du galactose, et l'AMP fait partie de la structure des coenzymes NAD et Coenzyme A. Par-dessus tout, les nucléotides sont les molécules essentielles servant à construire les acides nucléiques (ADN et ARN). Les besoins en nucléotides de la cellule peuvent être assouvis par deux moyens : la synthèse *de novo* des nucléotides à partir des sources de carbone et d'azote disponibles, ou l'exploitation des nucléotides, nucléosides et nucléobases de l'environnement grâce à un système de sauvetage et d'interconversion. Les bases nucléotidiques (adénine, guanine, uracile et xanthine) stimulent la croissance des lactocoques mais ne sont pas indispensables [Cocaign-Bousquet et al. 1995; Smith et al. 1975]. Dans le lait, la présence d'une source de pyrimidine, l'orotate, a été mise en évidence, mais toutes les souches ne sont pas capables de l'assimiler. Les besoins en purine pour des souches auxotrophes ne peuvent, quant à eux, pas être satisfaits dans le lait en absence de précurseurs [Garault et al. 2000]. *Lactococcus lactis* met en œuvre et régule ces deux mécanismes en fonction des conditions environnementales, comme la majorité des procaryotes ; toutefois, elle présente un certain nombre de caractères qui lui sont spécifiques (pour revue voir [Kilstrup et al. 2005]).

La synthèse de 5-phosphoribosyl-pyrophosphate, ou PRPP

Le 5-phosphoribosyl-pyrophosphate, ou PRPP, est un intermédiaire commun aux voies de biosynthèse de purines, de pyrimidines, et de sauvetage – interconversion. Il est également requis pour la synthèse des cofacteurs NAD⁺ et NADP⁺ et impliqué dans la synthèse de l'histidine et du tryptophane. Il est synthétisé par la PRPP synthétase codée par les gènes *prsA* et *prsB*, à partir du ribose de la voie des pentoses phosphate, et d'ATP. Le PRPP est également un inducteur de l'expression des gènes impliqués dans la synthèse *de novo* des purines chez *L. lactis*. En effet, l'expression de ces gènes est dépendante des concentrations de l'activateur PurR qui stimule la transcription à forte concentration de PRPP [Kilstrup & Martinussen 1998].

La synthèse *de novo* des nucléotides

La synthèse *de novo* consiste à utiliser la machinerie enzymatique afin de transformer les sources de carbone et d'azote en molécules servant à la production des nucléobases. Ainsi deux voies existent : la synthèse *de novo* des purines (adénine et guanine) et la synthèse *de novo* des pyrimidines (uracile, thymine et cytosine).

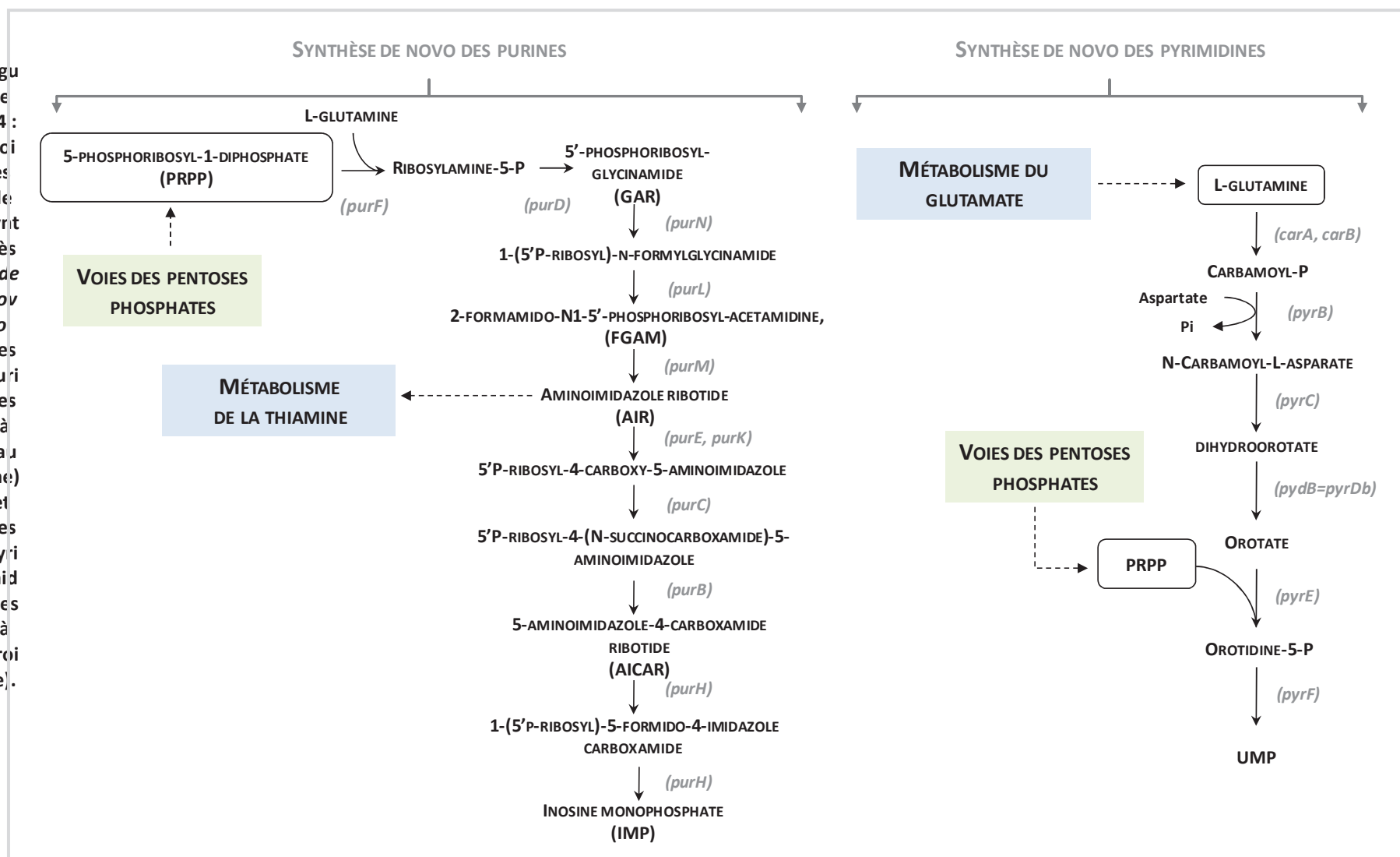
a. La synthèse de novo des purines

La synthèse *de novo* des purines résulte de la succession de 10 réactions enzymatiques conduisant, à partir du PRPP (5-phosphoribosyl-pyrophosphate) et de la L-glutamine, à la formation de l'inosine monophosphate, IMP (Figure 14). Dix gènes, organisés en opérons ou monocistroniques, codent les enzymes catalysant ces dix réactions : *purB*, *purC*, *purD*, *purE*, *purF*, *purH*, *purK*, *purL*, *purM*, *purN* [Kilstrup & Martinussen 1998; Jendresen et al. 2012]. La première réaction de la biosynthèse *de novo* des purines, catalysée par PurF, consiste en l'assemblage du pyrophosphate provenant du PRPP avec le groupement amine de la glutamine, formant le ribosylamine-5-P (PRA). Ensuite les enzymes PurD et PurN convertissent le ribosylamine-5-P successivement en glycinamide ribonucléotide (GAR) et formylglycinamide ribonucléotide (FGAR). La transformation du FGAR en formylglycinamide ribonucléotide (FGAM) requiert PurL, mais aussi, probablement chez *L. lactis*, les deux sous-unités PurQ et PurS. PurQ porterait un domaine glutamine amidotransférase alors que PurS servirait de lien entre PurQ et PurL. Ensuite, l'enzyme codée par *purM* convertit le FGAM en aminoimidazole ribonucléotide (AIR). La voie métabolique peut continuer vers l'IMP par la transformation de l'AIR en 5'-phospho-ribosyl-4-carboxy-5-aminoimidazole (CAIR) par l'activité de PurE et PurK, ou être déviée vers le métabolisme de la thiamine. PurC et PurB catalysent ensuite la production d'aminimidazole carboximide ribonucléotide (AICAR). Enfin, les deux dernières réactions qui conduisent à la synthèse de l'IMP sont catalysées par l'enzyme PurH [Zalkin & Dixon 1992].

b. La synthèse de novo des pyrimidines

La synthèse *de novo* des pyrimidines permet de produire de l'UMP à partir du PRPP et de la L-glutamine grâce à 6 réactions enzymatiques codées par les 7 gènes suivants : *carA*, *carB*, *pyrB*, *pyrC*, *pyrDb*, *pyrE* et *pyrF* (Figure 14). L'UMP sert ensuite de précurseur pour la formation de l'UTP, CTP, ou dTTP [Martinussen et al. 2001]. La première étape de la synthèse *de novo* des pyrimidines est la formation de carbamoyl-P. Les gènes *carA* et *carB* codent respectivement la petite et la grande sous-unité de la carbamoyl-P-synthase (CPSase) qui catalyse cette réaction. La deuxième réaction est la transformation du carbamoyl-P en carbamoyl-aspartate par l'aspartate-transcarboxylase (PyrB) [Martinussen & Hammer 1998]. Ensuite, le carbamoyl-aspartate est transformé en dihydroorotate par l'enzyme produit du gène *pyrC* [Bolotin et al. 1999] et PydB catalyse la production d'orotate. Une particularité de l'espèce *L. lactis* est qu'elle possède deux dihydroorotate-synthase, codées par les gènes *pyrDa* (ou *pydA*) et *pyrDb* (ou *pydB*) à la différence de tous les autres procaryotes étudiés à ce jour. Il a été montré que *L. lactis* peut n'avoir qu'une seule de ces deux enzymes sans perdre sa prototrophie pour les pyrimidines. L'enzyme PyrE synthétise l'orotidine-5-P à partir du PRPP et de l'orotate [Bolotin et al. 1999]. Enfin, *pyrF* code une OMP-décarboxylase qui convertit l'orotidine en UMP, produit final de la synthèse *de novo* des pyrimidines [Andersen et al. 1989].

Figure 14 : Voies de synthèse de novo des purines (à gauche) et des pyrimidines (à droite).



Transport, voies de sauvetage et d'interconversion

Afin d'éviter le détournement de matériel cellulaire et le coût énergétique imposé par la synthèse *de novo* de nucléotides, les bactéries lactiques sont capables d'utiliser les molécules préexistantes dans le milieu pour satisfaire leurs besoins en nucléotides. En effet, ces voies réactionnelles baptisées voies de sauvetage et d'interconversion permettent notamment la croissance de mutants auxotrophes pour les purines lors de l'ajout d'adénine, guanine, ou hypoxanthine [Nilsson & Lauridsen 1992]. Ces résultats démontrent l'existence de système de transport spécifique pour internaliser et transformer ces sources de purines exogènes. Chez *L. lactis* ssp. *lactis* IL1403, il existe une uracile-perméase, codée par le gène *pyrP* [Martinussen et al. 2001], et un transporteur de purine codé par le gène *pbuX* [Bolotin et al. 2001]. Comme précédemment évoqué, on notera également que certaines souches de *L. lactis* ont un transporteur d'orotate, codé par le gène *oroP* [Kilstrup et al. 2005]. Après leur internalisation, l'adénosine, l'inosine et la guanine peuvent-être déphosphorylées par la nucléoside-phosphorylase codée par le gène *deoD* pour former respectivement de l'adénine, de l'hypoxanthine ou de la guanine [Kilstrup et al. 2005]. Une autre voie possible pour l'adénosine et l'inosine est leur conversion respective en inosine et adénosine par l'adénosine-désaminase (*add*). Le sauvetage des purines s'effectue ensuite grâce à la conversion des ribonucléosides monophosphates par les phosphoribosyltransferases. Quatre gènes de purine-phosphoribosyltransferase ont été identifiés sur le génome de *L. lactis* IL1403. Les gènes *hpt* et *hprT* codent deux hypoxanthine-guanine-phosphoribosyltransferase, *xpt* code une xanthine-phosphoribosyltransferase, et *apt* une adenine-phosphoribosyltransferase [Nilsson & Lauridsen 1992].

En ce qui concerne les pyrimidines, toutes sont métabolisables par *Lactococcus lactis* à l'exception de la cytosine. La cellule importe des précurseurs pour ensuite les convertir en cytosine intracellulaire ou activer la synthèse *de novo*. La cytidine importée dans la cellule peut-être désaminée par une cytidine-désaminase (gène *cdd*) pour former l'uridine, ou être convertie en CMP par l'uridine kinase codée par *udk* [Kilstrup et al. 2005]. L'uridine peut également être prise en charge par l'uridine kinase (*udk*) pour former l'UMP, ou être convertie en uracyle par Pdp [Kilstrup et al. 2005]. L'enzyme de la voie de sauvetage de l'uracile (pyrimidine) est l'uracile-phosphoribosyltransferase (UPRTase), codée par le gène *upp*, qui conduit à la synthèse d'UMP à partir de l'uracyle (Martinussen and Hammer, 1994).

II Diversité technologique des levains et identification de signatures spécifiques

Caractériser la diversité fonctionnelle de *L. lactis* s'inscrit dans la continuité de l'étude sur le comportement des souches et représente une nécessité pour sélectionner des souches adaptées à l'application et aux propriétés désirées. En effet, les progrès effectués dans la connaissance des lactocoques ont mis en évidence un très large éventail de propriétés allant bien au-delà du potentiel acidifiant, et cela avec un caractère souche-dépendant très marqué au sein de l'espèce [Ayad et al. 1999; Yvon et al. 1997; Smit et al. 2004]. Les recherches se sont alors orientées vers la découverte ou l'ingénierie de nouvelles souches présentant des propriétés intéressantes telles que la production d'arômes, d'exopolysaccharides ou encore de bactériocines permettant de conférer aux produits typicité et singularité. La problématique repose donc sur l'évaluation, la caractérisation et la compréhension de cette diversité, afin d'identifier les souches à la signature fonctionnelle en adéquation avec le produit final désiré. Ce chapitre présente d'une part une synthèse non exhaustive des informations recensées concernant la diversité technologique des ferments de *L. lactis* (naturelle ou créée), et d'autre part les méthodes d'exploration et de sélection des souches.

Diversité technologique des levains de *L. lactis*

L'analyse de la diversité des souches de levains a été menée sur des critères divers mais spécifiques des caractéristiques technologiques qu'ils assument : acidification, production d'arômes, texture, autolyse, et synthèse de bactériocines comme la nisine, afin d'exploiter les différences des souches. Cependant, cette revue présente ces différentes fonctionnalités comme des critères indépendants alors que la spécificité d'une souche industrielle réside bien souvent dans la multiplicité de ces propriétés.

L'acidification

L'activité acidifiante des LAB représente la principale activité métabolique des cellules, avec la conversion de plus de 90 % du sucre consommé en acide lactique dans le cadre d'un métabolisme homo-fermentaire [Novák & Loubiere 2000]. Ce processus d'acidification du milieu, associé à la croissance, est à l'origine de la formation de caillés lactiques (voir 0) et de l'action de synérèse (égouttage des caillés de fromagerie en cuve et en moules par expulsion de lactosérum). L'accumulation d'acide lactique acidifie ; au pKa égal à 3,9 ; est à l'origine d'un état de stress acide pouvant aller jusqu'à l'auto inhibition des bactéries. Ce stress est attribué à la perméabilité de la membrane à la forme non dissociée de l'acide lactique qui diminue le pH intracellulaire. Cette acidification intracellulaire se traduit par une chute du taux de croissance, des activités des enzymes

glycolytiques, de la production d'ATP [Raynaud et al. 2005], et d'une dépurination et dépyrimidination de l'ADN [Lindahl & Nyberg 1972].

Du fait de son interaction avec de nombreuses autres activités d'intérêt technologique telles que la conservation (activité anti-microbienne démontrée des acides faibles [Presser et al. 1997]), la texture (formation du gel lactique [Beal et al. 1999; Kristo et al. 2003]) ainsi que la production aromatique [Visser 1993], l'acidification est considérée comme un paramètre clef de la fermentation des produits laitiers. Ainsi, pour chaque type de fromage ou de lait fermenté, un profil d'acidification bien défini est recherché. Par exemple, la fabrication des fromages frais et yaourts impose une acidification lente et régulière avec une phase de stabilisation du pH en fin de maturation afin de limiter le phénomène de « post-acidification » lors du stockage au froid.

Etant donné l'importance de la fonction acidifiante, la détermination des potentialités acidifiantes des souches a fait l'objet de nombreuses méthodes d'évaluation : acidité titrable, conductivité [Belfares et al. 1993], ou encore la fluorescence [John et al. 2003]. Cependant la pratique la plus usitée reste la mesure dynamique par CINAC (CINétique d'Acidification) mise au point par Corrieu, qui permet d'obtenir des descripteurs pertinents pour caractériser l'activité acidifiante de plusieurs fermentations en simultané [Corrieu et al. 1989]. Les indicateurs suivants sont les plus fréquemment utilisés :

- $V_{\max}^a$: la vitesse maximale d'acidification en unité pH.min⁻¹ ;
- $T_{\max}^a$: le temps nécessaire en heure pour atteindre $V_{\max}^a$;
- $\text{pH}_{\max}^a$: le pH en unité pH correspondant à $V_{\max}^a$;
- T^a : la durée de la phase de latence correspondant à une diminution de 0,08 unité pH. h⁻¹ ;
- $T_{\text{pH}4,6}$: le temps nécessaire pour atteindre un pH en heure, fréquemment le pH 4,6 de fin de maturation dans les laits fermentés ;

Les études ont montré que l'activité acidifiante dépendait d'une combinaison de facteurs : nombre de cellules, état physiologique et caractéristiques de la souche étudiée [Cachon et al. 2002; Picque et al. 1992; Casalta et al. 1995]. Il a ainsi été prouvé que les trois descripteurs $V_{\max}^a$, $T_{\max}^a$ et $\text{pH}_{\max}^a$ étaient robustes et constants (pour une gamme de concentration bactérienne) afin de décrire l'activité acidifiante d'une souche [Picque et al. 1992]. Il est donc possible d'établir une classification inter et intra espèces sur la base de ces critères, à la différence des autres indicateurs qui sont dépendants de la concentration, de l'état cellulaire, du pH initial ou encore de la température [Picque et al. 1992]. Ainsi, alors que la $V_{\max}^a$ traduit le potentiel d'acidifiant d'une souche, le descripteur $\text{pH}_{\max}^a$ éclaire sur l'aptitude de la souche à être active en présence d'acide lactique : plus la valeur de pH est faible, plus la souche est apte à croître et à acidifier en présence de lactate. Dans

des conditions de cultures favorables, les LAB thermophiles conduisent à des valeurs de $V_{\max}^a$ toujours supérieures à $10 \cdot 10^{-3}$ et jusqu'à $20 \cdot 10^{-3}$ upH.min⁻¹ (cas des associations *Streptococcus thermophilus* et *Lactobacillus delbrueckii* ssp. *bulgaricus*). En comparaison, les lactocoques, en tant que bactéries mésophiles, acidifient plus lentement, à des valeurs en-dessous de $10 \cdot 10^{-3}$ upH.min⁻¹ [Corrieu 2008]. La principale source de diversité entre les activités acidifiantes est le phénotype nutritionnel, avec la présence de la protéase de paroi, PrtP. En effet, en moyenne les souches de *Lactococcus lactis* PrtP⁻ présente une $V_{\max}^a$ deux fois plus faible que les souches PrtP⁺ en lait. Un autre facteur de variabilité intra-espèce de *L. lactis* ssp. *lactis* est l'appartenance des souches au biovar *diacetylactis* qui concourt à diminuer leur action acidifiante d'un facteur deux également (synthèse à partir du citrate d'acétate, d'acétoïne et de CO₂ en lieu et place de l'acide lactique). L'étude de Cachon et coll (2002) réalisée sur 7 souches de *Lactococcus lactis* PrtP⁺ révèle une variabilité de 11% entre les différentes bactéries testées.

L'étude comparative de la régulation de l'acidification sur différentes souches levains de *Lactococcus lactis* a démontré que les différences de $V_{\max}^a$ résultent principalement de divergences de taux de croissance de la souche (anabolisme) plus que de différences de capacités glycolytiques intrinsèques sur milieu lait [Raynaud et al. 2005]. Ces différences sont notamment liées à l'inhibition de la croissance par l'acidification provoquée par le produit métabolique final, le lactate. En effet, la membrane est perméable à la forme non dissociée de l'acide lactique qui peut alors diffuser librement pour s'équilibrer entre le milieu et le cytoplasme. L'acide lactique, dans l'environnement basique (par rapport au milieu extérieur) du pH cytoplasmique [Cook & Russell 1994] va se dissocier pour produire la forme anionique, le lactate, et des ions H⁺ qui vont acidifier le cytoplasme [Kashket 1987]. Les mécanismes de maintien de l'homéostasie du pH intracellulaire vont devoir se mettre en place (pompe ATPases, voie ADI, utilisation du citrate et décarboxylation des acides aminés) pour maintenir le pH intracellulaire, avec un coût énergétique d'autant plus élevé que le pH baisse. La sensibilité des enzymes de la glycolyse (PGK, PFK) et de la LDH au pH intracellulaire est un facteur relativement homogène, et situé à une valeur limite de 4,84 [Raynaud et al. 2005] responsable de l'arrêt du catabolisme des souches [Rallu et al. 1996] (schéma bilan Figure 15). Par ailleurs, il a pu être déterminé que le pH atteint à l'arrêt de croissance (fonction du phénotype nutritionnel) fixe la vitesse de post-acidification. Ainsi, un pH d'arrêt de culture peu élevé est peu inhibiteur vis-à-vis des enzymes du métabolisme central (en particulier PFK et PGK) et l'activité glycolytique est donc maintenue, d'où une post-acidification importante (Raynaud 2006).

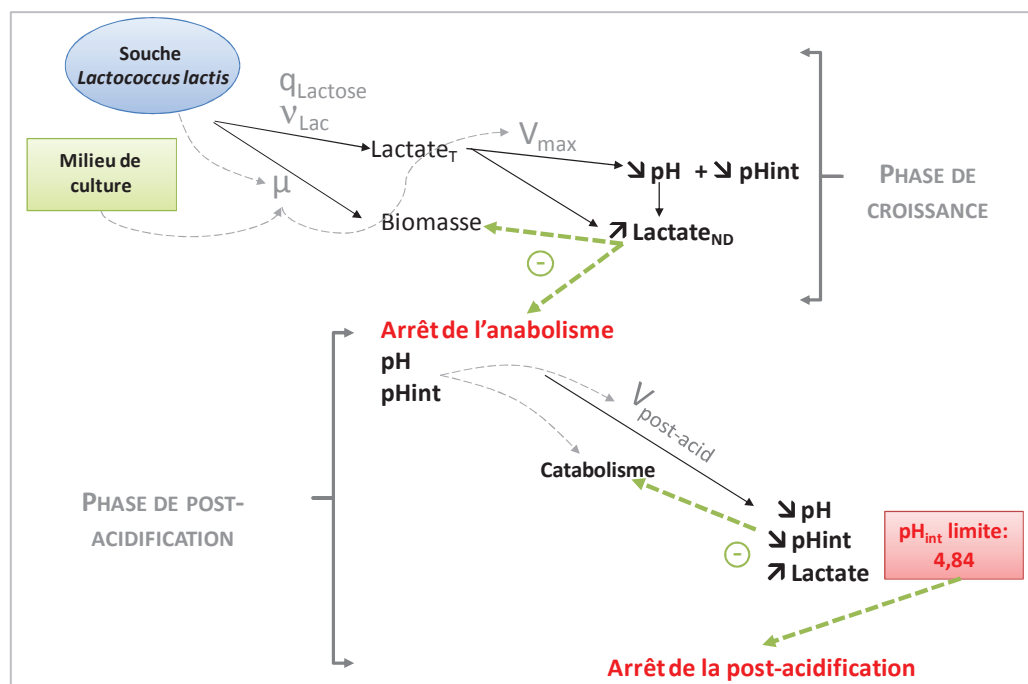


Figure 15 : Régulation de l'activité de l'acidification chez *Lactococcus lactis* dans des conditions de culture nutritionnellement non limitantes. [Raynaud 2006]

La production d'arômes

Le développement aromatique au sein d'un fromage ou d'un lait fermenté est un processus biochimique dynamique qui est impacté par la composition du lait, les conditions du procédé et les microorganismes et enzymes. Les LAB produisent sur matrice laitière, en plus de l'acide lactique, divers métabolites secondaires issus des trois grandes familles de substrats précédemment présentés : lactose et citrate (glycolyse et réorientation du flux du pyruvate vers un métabolisme mixte) (voir Figure 8), caséine (protéolyse et le catabolisme des acides aminés) (Figure 11 et Tableau 4) et triglycérides (lipolyse et catabolisme des acides gras) (Figure 13 et Tableau 5). A côté du criblage haut débit, la démarche pour isoler des souches (ou une combinaison de souches) présentant une production aromatique optimale en milieu lait peut-être accélérée par la connaissance des régulations des voies métaboliques concernées, par le travail sur la composition « idéale » en arômes pour obtenir l'effet sensoriel désiré, et par la compréhension des interactions entre les différentes souches et leurs environnement. Des exemples de ces différents points seront présentés ci-dessous dans le processus de sélection de souches.

La recherche de souches dites « aromatiques » s'est orientée traditionnellement vers les souches de *Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* en raison de leur capacité à produire diacétyle (D), molécule aux notes crème et beurrée. Utilisé en fromagerie (Gouda-Edam) ou dans la fabrication de laits fermentés en culture mixte, *L. lactis* ssp. *lactis* biovar *diacetylactis* peut représenter jusqu'à 20%

des ferments totaux [Urbach 1997; Fox & McSweeney 2004; Smit et al. 2004]. Ce biovar est usuellement identifié par croissance sur gélose Kempler&McKay [Kempler & McKay 1980] ou par la présence du gène plasmidique *citP* codant la citrate perméase. Cependant, une étude récente de criblage d'une collection de 36 souches productrices de Diacétyle/Acétolactate (A) a révélé que des souches isolées dans un environnement non laitier et ne possédant aucun des gènes du catabolisme du citrate sont capables de produire des quantités équivalentes de D/A. Ce résultat remet en question les procédures classiques d'identification de ces souches et démontre le réservoir de diversité encore inexploitée que représentent les isolats naturels [Passerini et al. 2013]. Par ailleurs, améliorer la production de diacétyle des souches par ingénierie métabolique a été une seconde approche développée. Plusieurs équipes ont étudié des biovariants de *L. lactis* ssp. *lactis* biovar. *diacetylactis* déficients en α -acétolactate décarboxylase (ALDC) qui accumulent de l' α -acétolactate dans le milieu (2 mM) [Jordan & Cogan 1988; Hugenholtz 1993; Monnet et al. 1994]. Cette accumulation, associée à une décarboxylation oxydative en présence d'O₂ dans le milieu, permet la surproduction de diacétyle. Sur la base de ce modèle de dérégulation, des mutants ont été générés de façon dirigée [Goupil-Feuillerat et al. 1997; Monnet et al. 2003], par mutation spontanée [Goupil et al. 1996] ou encore par mutagenèse aléatoire et isolés par criblage [Monnet et al. 1997; Monnet et al. 1999]. De nombreuses combinaisons de mutations de gènes en aval du pyruvate ont été testées (LDH-ALS-ILVBN). Parmi les différents mutants, l'atténuation de l'activité de la LDH en combinaison avec la déficience en ALDC est l'approche qui a permis d'obtenir les meilleurs résultats [Corrieu 2008].

Le second axe d'investigation dans la recherche de souches aux propriétés aromatiques a porté sur les capacités de protéolyse et de synthèses de composés volatils en provenance des acides aminés. L'exploration de cette diversité a été réalisée en s'appuyant d'une part sur les résultats de criblage de produits d'intérêts, et d'autre part sur les dosages d'activités des différents enzymes impliquées dans ces voies. La détection des niveaux de production de ces composés par GC-MS ainsi que les multiples analyses sensorielles ont révélé une grande hétérogénéité dans les capacités aromatiques de la sous-espèce *lactis*, renforçant le caractère extrêmement souche-dépendant de ce critère de sélection [Ayad et al. 1999; Fernández et al. 2011].

En particulier, la recherche de souches productrices de 2,3-méthylbutanal ou 2-méthylpropanal (arômes maltés/chocolatés typiques des fromages affinés), issus des BCAAs, est venue confirmer les disparités présentes au sein des collections, avec des différences d'un facteur 40 en concentration [Ayad et al. 2001; Smit et al. 2004]. Afin d'élucider l'origine de cette diversité phénotypique, les enzymes protéolytiques impliquées dans la production d'arômes (peptidases, aminotransférases, décarboxylases, déshydrogénases et estérases) ont été étudiées par comparaison génétique et de

régulation [Liu et al. 2008]. L'analyse de la diversité naturelle des enzymes a révélé que le niveau total de protéases varie d'un facteur de 8 à 15 au sein des sous-espèces *cremoris* et *lactis* respectivement [Crow et al. 1994]. Plus précisément, sur un jeu de 6 peptidases analysées, Crow a mesuré des variations d'activité allant de 2 à 22. Des facteurs du même ordre de grandeur, voire supérieurs (entre 1,5 et 249) pour la diversité intra-espèce chez *L. lactis*, ont été retrouvés pour les variations d'activités des aminotransférases BcaT (substrat acides aminés branchés) et AraT (substrat phénylalanine) et de l'hydroxyacides déshydrogénase (HicDH) [Fernández de Palencia et al. 2006; Brandsma et al. 2008; Bachmann et al. 2009b] (voir Figure 11 pour plus de détails sur les voies métaboliques). De plus, au sein des souches à forte activité en BcaT, des facteurs de variations entre 5 et 6 pour les activités de l'AraT et de l'HicDH sont conservés, suggérant que la sélection rationnelle de souches possédant des combinaisons uniques de niveaux d'enzymes du métabolisme azoté est possible.

Bien qu'une partie de ces résultats ait été confirmée lors d'expérimentations en matrice laitière [Ayad et al. 2001], l'absence de corrélations pour une même enzyme entre deux environnements de criblage (milieu synthétique ou riche) limite les capacités prédictives de ces résultats sur milieu laitier [Bachmann et al. 2009b] et accentue la nécessité de conditions de criblage au plus près des matrices. Cependant, une séparation des souches en fonction de leur écotype d'isolation (laitiers/non laitiers) a été démontrée sur milieu synthétique à partir des profils d'activités de BcaT et AraT, significativement inférieurs sur les isolats non laitiers. Le même résultat est obtenu pour la peptidase PepX, qui corrobore l'hypothèse d'une adaptation physiologique à la niche écologique [Siezen et al. 2011].

La découverte que la réaction de transamination, à la base du processus de formation des molécules aromatiques, était limitée faute de céto-acide accepteur en quantité suffisante, a initié des stratégies de sélection et de renforcement des voies de production d' α -cétoglutarate [Yvon et al. 1998]. L'activité de la GDH étant corrélée à la capacité des souches à cataboliser les acides aminés branchés et aromatiques en présence de glutamate [Tanous et al. 2002; Kieronczyk et al. 2004; Helinck et al. 2004], le criblage des souches pour une activité de GDH apparaît comme pertinent. Les travaux d'inactivation, surexpression ou encore d'expression de GDH hétérologue menés en parallèle confirment le rôle clef de la GDH dans la régulation de ce procédé [Rijnen et al. 2000; Tanous et al. 2005b]. L'analyse des génomes de LAB disponibles confirme la présence largement répandue de gènes homologues à la GDH au sein des séquences de *Streptococcus*, *Lactobacillus*, ou *Leuconostoc* à la différence des *Lactococcus* [Liu et al. 2008]. En effet, peu de souches de *L. lactis* disposent d'une activité GDH [Tanous et al. 2002], et ce sont presque exclusivement des isolats d'origine

environnementale. Ce constat s'explique par l'incompatibilité du plasmide *pGdh442* porteur du gène *gdh* avec les plasmides de type *SK08* qui portent les gènes de transport et d'utilisation du lactose et de la caséine [Tanous et al. 2006]. Ce résultat prouve une fois de plus le réservoir de diversité fonctionnelle que représentent les isolats végétaux au sein de la sous espèce *lactis* et la pertinence des cultures mixtes pour répartir les fonctions du ferment entre plusieurs souches [Ayad et al. 2001].

Cependant, ces démarches d'augmentation ciblée d'un composé volatil, ou même d'une voie, sont à relativiser en considérant que la perception d'un arôme dépend essentiellement de l'équilibre de ce dernier avec les autres composés [Ott et al. 1997; Wouters et al. 2002]. Le potentiel aromatique d'une souche, ou signature aromatique, est un phénomène relatif qui doit être considéré dans sa globalité et tenir compte également des interactions avec l'environnement. Ainsi, les effets de la présence d'oxygène sur la production aromatique (réactions de décarboxylation et implication dans le shift vers le métabolisme mixte) ont-ils été longuement étudiés, pour revue [Condon 1987; Felipe et al. 1998]. Autre facteur d'importance, le potentiel d'oxydoréduction extracellulaire, ou Eh, caractérise la capacité du milieu à donner (oxydation) ou à accepter (réduction) des électrons. Il dépend autant du milieu que de l'activité réductrice des bactéries. Ainsi, tout comme pour l'acidification, le potentiel de réduction du milieu peut-être utilisé pour effectuer une classification inter-espèces des LABs. Les lactocoques apparaissent alors comme les bactéries les plus réductrices [Brasca et al. 2007], et une diversité intra-espèce a été également mise en avant [Cachon et al. 2002]. Caractériser cette diversité est d'autant plus important qu'il a été démontré que la production et la stabilité des arômes sont dépendantes des conditions d'oxydoréduction, en particulier pour l'orientation des voies de catabolisme des acides aminés. Des conditions oxydantes favorisent la production d'aldéhydes et de composés soufrés par l'oxydation chimique des α -céto-acides issus de la transamination. A l'inverse, des conditions réductrices stimulent la production d'acides carboxyliques [Kieronczyk et al. 2006]. Ces constatations font de l'Eh et de l'utilisation d'atmosphère réductrice des leviers pour modifier les qualités sensorielles des produits [Cachon et al. 2004].

L'autolyse

L'autolyse est un phénomène de lyse cellulaire sous l'action d'hydrolysases bactériennes, qui conduit à la mort cellulaire et à la libération du contenu cytoplasmique dans le milieu environnant. Ce mécanisme représente une propriété intéressante pour accélérer le développement aromatique lors de l'affinage des fromages notamment [Law et al. 1974; Chapot-Chartier et al. 1994; Crow et al. 1995]. En effet, la libération du contenu cytoplasmique dans le milieu, et en particulier des enzymes du système protéolytique (peptidases, enzymes du catabolisme des acides aminés, estérases) augmente la production aromatique [Law et al. 1974]. L'ensemble des travaux concernant la lyse des

lactocoques révèle pour une même souche une élévation d'un facteur 2 à 3 de la concentration finale en acides aminés libres dans le produit et une réduction de l'amertume par rapport à une souche non lysée [Crow et al. 1995; Lepeuple et al. 1998; Ledebøer et al. 2002; O'Sullivan et al. 2000]. De plus il semblerait qu'un lien entre autolyse et lipolyse existe, du fait de la libération d'estérases dans le milieu lors de la lyse cellulaire (Figure 15).

Ces enzymes, capables d'hydrolyser la paroi bactérienne, sont synthétisées pendant la croissance cellulaire, et interviennent dans différentes fonctions cellulaires essentielles : séparation des cellules, recyclage du peptidoglycane ou encore expansion de la paroi au cours de la croissance. Sous l'action de facteurs environnementaux défavorables à la physiologie cellulaire, l'action de ces enzymes peut conduire à la lyse bactérienne, elles sont alors appelées autolysines. De nombreux agents comme le pH, la température, la source de carbone, ou encore la concentration en sel, apparaissent comme importants dans l'induction de l'autolyse [Riepe et al. 1997]. Chez *L. lactis*, 5 hydrolases du peptidoglycane ont été identifiées, dont quatre *N*-acétyl-glucosaminidases, AcmA, AcmB, AcmC et AcmD, ainsi qu'une endopeptidase YjgB [Gasson 1996; Buist et al. 1995; Inagaki et al. 2009; Steen et al. 2008; Meyrand et al. 2007]. De plus, il a été démontré une relation entre la lysogénie des souches de LAB, la présence dans leur génome de l'ADN de prophage(s), et l'induction du phénomène de lyse [O'Sullivan et al. 2000]. Cette induction peut survenir en présence d'un agent mutagène ou d'un stress environnemental, et provoquer la synthèse d'endolysine et d'une deuxième protéine, l'holine, qui en s'agglomérant, forme des pores dans la membrane plasmidique, permettant à l'endolysine d'accéder au peptidoglycane. Ce phénomène est à l'origine de la forte activité de lyse démontrée au sein de la souche autolytique *L. Lactis* ssp. *cremoris* AM2 [Lepeuple et al. 1998].

Des travaux récents ont montré le développement par ingénierie moléculaire de souches de *Lactobacillus casei* au système autolytique contrôlé [Xu & Kong 2013]. Cependant, les mécanismes moléculaires de l'induction et du contrôle de l'autolyse en matrice laitière restent à approfondir, tout comme l'origine des variations constatées dans l'aptitude de la lyse, très souche-dépendante [O'Donovan et al. 1996; Boutrou et al. 1998; Govindasamy-Lucey et al. 2000]. Plusieurs hypothèses ont été avancées pour expliciter ces différences : variations de composition pariétale, d'efficacité des hydrolases, de sensibilité et de présence de prophages ou encore d'efficacité de la protéine de paroi PrtP dans l'hydrolyse des autolysines [Govindasamy-Lucey et al. 2000].

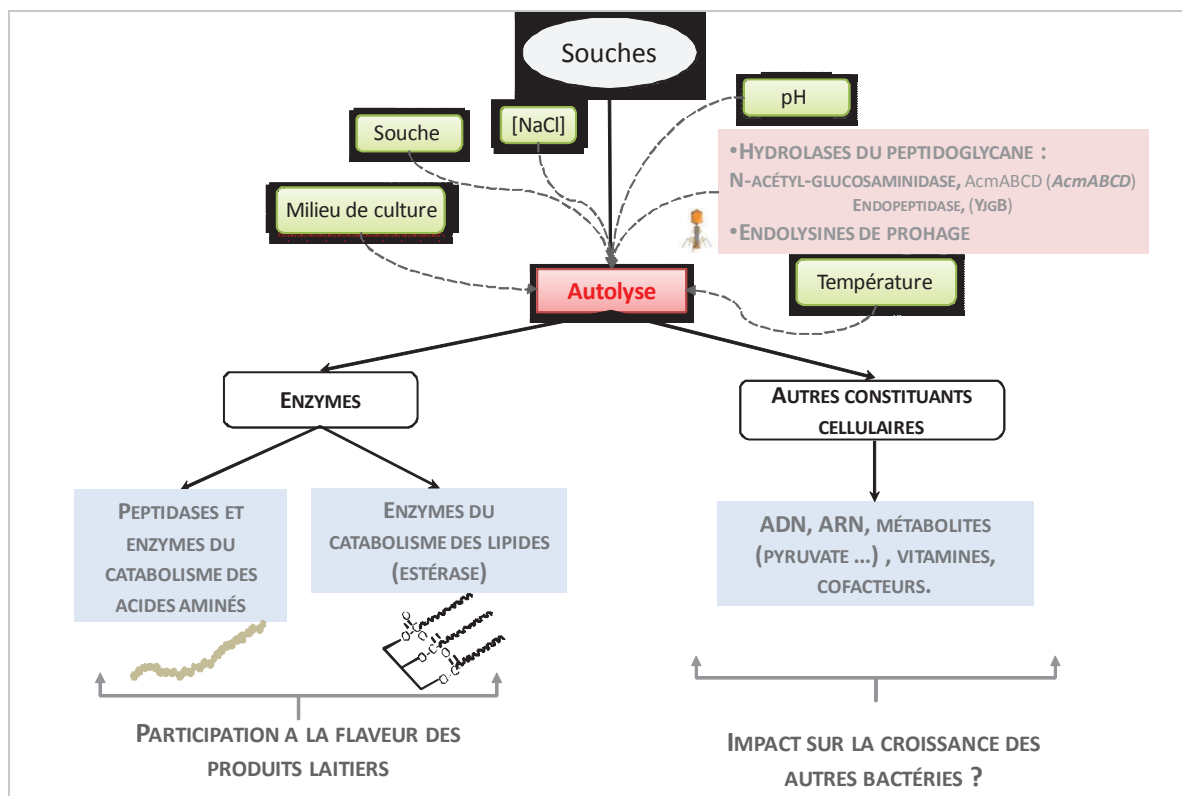


Figure 16 : Illustration des principaux facteurs intervenant dans le phénomène d'autolyse des souches, et des conséquences de la libération des constituants cellulaires. Adapté de [Corrieu 2008].

Les propriétés texturantes

Les propriétés texturantes des LAB en tant que producteurs d'exopolysaccharides (EPS) sont principalement utilisées pour améliorer les qualités organoleptiques des produits laitiers frais et répondre aux demandes des consommateurs pour des produits « doux et crémeux ». Ainsi, la production d'exopolysaccharides par les bactéries représente une solution économique et naturelle pour épaissir et stabiliser les solutions en lieu et place des traditionnelles augmentations du taux de matières protéiques ou de rajout d'additifs (texturants ou épaississants) [Jolly et al. 2002; Welman & Maddox 2003]. Bien que les EPS ne possèdent pas de saveur en tant que telle, ils peuvent augmenter le temps que le produit reste en bouche et ainsi produire une meilleure perception.

Les EPS sont des polysaccharides d'unités glucidiques répétées qui sont classés en fonction de leur localisation cellulaire (paroi, capsule ou libérés), mais également selon leur composition en sucres (homopolysaccharides ou hétéropolysaccharides). Les homopolysaccharides sont constitués d'un seul type de monosaccharide, comme le dextrane (D-glucose) ou les levanes (D-fructose). Les hétéropolysaccharides, comme le xanthane, représentent la majorité des EPS bactériens, et sont constitués de différents types de monosaccharides. Il existe une grande diversité de compositions, de charges, et de structures (chaînes linéaires ou ramifiées) des EPS parmi les LAB [De Vuyst et al. 1998;

van Kranenburg et al. 1999; Broadbent et al. 2003; Donot et al. 2012]. Les unités glucosidiques les plus répétées sont le glucose, le galactose, et le rhamnose. La biosynthèse des polysaccharides produits par les lactocoques débute avec la synthèse des précurseurs, les sucres nucléotidiques, et se poursuit par l'assemblage des sous-unités répétitives sur une ancre lipidique cytoplasmique. Les EPS produits par *Lactococcus lactis* ssp *cremoris* NIZO B40 sont constitués d'une unité pentasaccharidique composée des sucres nucléotidiques suivants : glucose, galactose, rhamnose, et de phosphate selon un ratio de 2:2:1:1 [van Casteren et al. 1998; Looijesteijn et al. 1999; Ruas-Madiedo et al. 2002]. Les dernières étapes comprennent l'exportation et la polymérisation de centaines de sous-unités à l'extérieur de la cellule (pour revue sur la synthèse [Welman & Maddox 2003]). D'un point de vue physiologique, les EPS sont impliqués dans la prévention de la dessiccation, la protection contre divers stress, l'adhérence aux surfaces, et dans la sensibilité et la résistance aux bactériophages [Whitfield & Valvano 1993].

Les quantités d'EPS produites dans le lait par les différentes souches de *L. lactis* varient de 80 à 600 mg.L⁻¹, même si de plus faibles rendements ont été également reportés [Cerning 1995; Knoshaug et al. 2000]. Cependant, en plus d'une forte diversité intra-espèce, il est souvent constaté une instabilité de la production d'EPS, en particulier lorsque les gènes codant pour la production d'EPS sont plasmidiques et peuvent donc être perdus, comme cela peut-être le cas avec les LAB mésophiles. Chez *L. lactis* ssp. *cremoris* NIZO B40, les gènes *eps* couvrent 12 kpb (*epsRXABCDEFGHIJKL*) sur le plasmide pNZ4000, tout comme chez *L. lactis* ssp. *cremoris* HO2 où 21,8 kpb contiennent 23 gènes potentiellement impliqués dans la synthèse des EPS sur le plasmide pCI658 [Kranenburg et al. 1997]. Cependant, chez la souche ssp *cremoris* SMQ-461, une région chromosomique de 13,2 kpb a été également mise en évidence pour ces mêmes gènes [Dabour & LaPointe 2005]. Les conditions de croissance comme la composition du milieu (vitamines, minéraux) affectent également fortement la productivité [Degeest & Vuyst 1999; Degeest et al. 2001]. De manière générale, les ferments mésophiles ont tendance à atteindre leur productivité maximale en condition non-optimale pour la croissance, à la différence des espèces thermophiles où la production est associée à la croissance.

La production de bactériocines : la nisine

La dernière propriété fonctionnelle recherchée pour les ferments de lactocoques est la synthèse, chez certaines souches, de nisine. La nisine appartient à la classe I des bactériocines. Ce sont des peptides antimicrobiens sécrétés par les bactéries et présentant une activité inhibitrice contre des souches phylogénétiquement proches de la souche productrice [Tagg et al. 1976; Klaenhammer

1993]. La nisine présente un intérêt industriel de par son innocuité envers l'homme et ses fonctions de conservateur naturel. En effet, la nisine a une action contre un large spectre de bactéries Gram positif (*Lactococcus*, *Lactobacillus*, *Listeria*, *Enterococcus*, *Clostridium*, *Staphylococcus* et *Mycobacterium*) [Meghroun et al. 1999]. Cette classe, aussi appelée classe des lantibiotiques, regroupe les composés protéiques d'une taille inférieure à 5 kDa et contenant les acides aminés modifiés thioéther lanthionine et méthyllanthionine, et des acides aminés déshydratés tels que les didéshydrobutyrine (Dhb) et didéshydroalanine (Dha).

La nisine est un peptide de 34 acides aminés dont la structure est composée de cinq cycles successifs élaborés lors de la maturation post-traductionnelle, et d'une région centrale fixe [Van de Ven et al. 1991]. Les gènes responsables de la synthèse de nisine ont été identifiés au sein de trois opérons chez *L. lactis*, codant une douzaine de gènes. Le gène *nisA* code la protéine responsable de la structure de la nisine, les gènes *nisB*, *nisC*, *nisD* et *nisM* codent des protéines impliquées dans les modifications post-traductionnelles, et le gène *nisP* code une protéase qui participe à la maturation de la bactériocine. Le gène *nisT* est impliqué dans la sécrétion du peptide mature avec la participation du transporteur ABC NisFEG (*nisF*, *nisE* et *nisG*). L'immunité à la bactériocine de la souche productrice est garantie par une lipoprotéase NisI. Enfin, les gènes *nisR* et *nisK* codent des facteurs de régulation de la transcription [Kleerebezem et al. 2001].

L'action bactéricide de la nisine est fonction de sa concentration dans le milieu (pour revue, [Reisinger et al. 1980; Abee et al. 1995]). Un premier mode opératoire, en présence de concentrations nanomolaires, consiste en une inhibition de la synthèse de la paroi [Reisinger et al. 1980]. Cette inhibition est le résultat de la formation d'un complexe entre le précurseur lipidique membranaire du peptidoglycane, le lipide II, et les molécules de nisine [Wiedemann et al. 2004]. Cette association engendrerait des pores stables et uniformes, ce qui conduirait à un efflux de protons, et donc à la dissipation de la FPM, d'où la mort cellulaire. Le second mode d'action, dit de Wedge-model, à concentration micromolaire, résulte de la fixation de la nisine (chargée positivement) aux polymères de paroi (chargés négativement) par attractions électrostatiques [Breukink et al. 1997; Breukink et al. 1998], et entraînant la formation de pores. Là encore, la formation de pores provoque la dissipation de la FPM. De fortes variations entre les souches de *L. lactis* existent tant sur la capacité de production de la nisine que sur l'impact des conditions régulatrices, nécessitant des études complémentaires pour être élucidées.

Identification de la signature d'une souche

Comme nous venons de le voir, il existe une importante diversité intra-espèce entre les souches de *L. lactis* concernant l'ensemble des propriétés fonctionnelles recherchées dans le cadre des ferments. Pour exploiter au mieux cette diversité, il est nécessaire de caractériser et de comparer le comportement des souches au plus près du procédé pour identifier les souches à la signature désirée.

Traditionnellement, cette étape de sélection de souches était liée au terroir des produits et basée sur un savoir empirique. Le besoin de rationalisation des procédés industriels a contribué à développer des outils de microbiologie fondés sur l'analyse de critères microbiologiques simples. Il est possible de citer notamment la gélose FSDA (*Fast Slow Differential Agar*), développée par Huggins & Sandine (Huggins & Sandine 1984) et utilisée pour différencier les bactéries lactiques en fonction de leur capacité à acidifier. Ainsi, l'acidification rapide du milieu par les souches Lac⁺ se traduit par l'apparition, après 24 heures d'incubation, d'un halo rouge intense autour des colonies, à la différence des souches Lac⁻. Cependant, cette approche présente de nombreuses limites comme nous allons le voir au travers de l'exemple de la sélection de souches de *L. lactis* ssp *lactis* biovar *diacetylactis*. Ces souches sont recherchées pour leur capacité à produire du diacétyle (D), arôme aux notes « beurrées » d'intérêts. Les souches ont été principalement criblées sur la base de critères phénotypiques. La capacité des souches à consommer du citrate, propriété métabolique traditionnellement associée aux souches du biovar *diacetylactis*, est testée par croissance sur gélose Kempler&McKay à 30°C après 18 à 72 h d'incubation [Kempler & McKay 1980]. Ainsi, les souches fermentant le citrate se colorent en bleu (KMK⁺ phénotype) alors que les souches incapables d'utiliser le citrate forment des colonies blanches (KMK⁻ phénotype). La réaction de Voges-Proskauer (VP) [Speckman and Collins, 1982] est également utilisée pour identifier les souches productrices d'acétoïne (formation d'un anneau rouge à la surface, VP phénotype). La limite de détection, le manque de spécificité et le temps nécessaire pour préparer ces essais cultures-dépendantes constituent les freins majeurs à l'utilisation de ces milieux différentiels. Pour pallier à ces inconvénients, la détection du gène plasmidique *citP* codant la citrate perméase a été développée par amorce PCR, et est devenu la nouvelle référence. Cependant, le criblage d'une collection de souches productrices de D/A a révélé que des souches ne possédant aucun des gènes du catabolisme du citrate sont capables de produire des quantités équivalentes de D/A [Passerini et al. 2013]. Ce résultat remet en question les procédures classiques d'identification de ces souches et démontre la nécessité de mettre en œuvre une approche intégrative multi-échelles (génétique vs métabolique) pour comprendre l'origine de cette diversité phénotypique, et éventuellement pour prédire la spécificité des souches au sein des différentes applications. De plus, le caractère souche-dépendant

des bactéries pose la problématique de l'identification et de la caractérisation de signatures des souches face à un vaste panel de réponses, elles-mêmes multicritères.

Pour répondre à ces défis, l'utilisation d'études comparatives à haut débit, du génotype au phénotype, se développe, de même que les méthodes statistiques pour en permettre une analyse intégrée des données sous le nom de biologie des systèmes. Ainsi différentes techniques à « haut-débit » ont-elles été regroupées sous le nom de « omiques » par analogie à la génomique. Il s'agit de la transcriptomique (étude des profils d'ARNm), de la protéomique (identification et quantification des protéines) et de la métabolomique (dosage des métabolites). Regroupées avec la génomique, ces différentes approches permettent de rendre compte de l'intégralité des niveaux de régulation intervenant dans l'expression génique, pour donner un phénotype. La démarche de biologie des systèmes suit ces principes en mettant en place une modélisation mathématique pour décrire l'effet dynamique de l'adaptation des bactéries à leur environnement (voir Figure 17). L'apport de ces techniques a notamment été utilisé pour définir de manière précise l'action de régulateurs lors de réponses complexes dans les aliments [O'Flaherty & Klaenhammer 2011].

Approche intégrative de la spécificité

Afin de comprendre l'origine de cette diversité fonctionnelle des souches, des études comparatives sur la sous-espèce *lactis* ont été menées du génotype au phénotype. L'avancée des outils de séquençage et le développement des outils d'annotation de séquences ont permis l'identification des gènes spécifiques de la sous espèce *lactis* [Makarova et al. 2006; Siezen et al. 2011]. Les techniques de génotypage (MLST, voir Figure 3) ont appréhendé l'organisation de la diversité génétique de la sous-espèce *lactis* avec une séparation des souches en fonction de leur écotype : « domestiqué » vs « environnemental » [Passerini et al. 2010]. Les souches environnementales concentrent la majorité de la diversité génétique, alors que les souches isolées de matrice laitière sont considérées comme de pauvres contributrices à la biodiversité, mais adaptées à leur niche écologique [Habib et al. 2007; Kelly et al. 2010; Passerini et al. 2010; Taïbi et al. 2010; Siezen et al. 2011]. En particulier, les études de comparaison de génomes par hybridation (CGH) réalisée ssur des levains laitiers des sous espèces *lactis* ou *cremoris* ont démontré qu'elles partagent un large génome cœur commun [Tan-a-ram et al. 2011; Taïbi et al. 2011]. Des travaux récents ont tenté d'utiliser ces données pour discriminer les souches et prédire leur phénotype [Bayjanov et al. 2012; Bayjanov et al. 2013]. Bien que de nouvelles fonctions pour des clusters de gènes aient pu être mises en évidence, et l'annotation fonctionnelle améliorée, l'importance des régulations post-traductionnelles de l'expression génique démontre que la présence/absence de gènes spécifiques ne se traduit pas nécessairement au niveau des phénotypes [Raynaud et al. 2005; Bachmann et al. 2009b; Taïbi et al. 2011; Tan-a-ram et al. 2011; Cretenet et al. 2011].

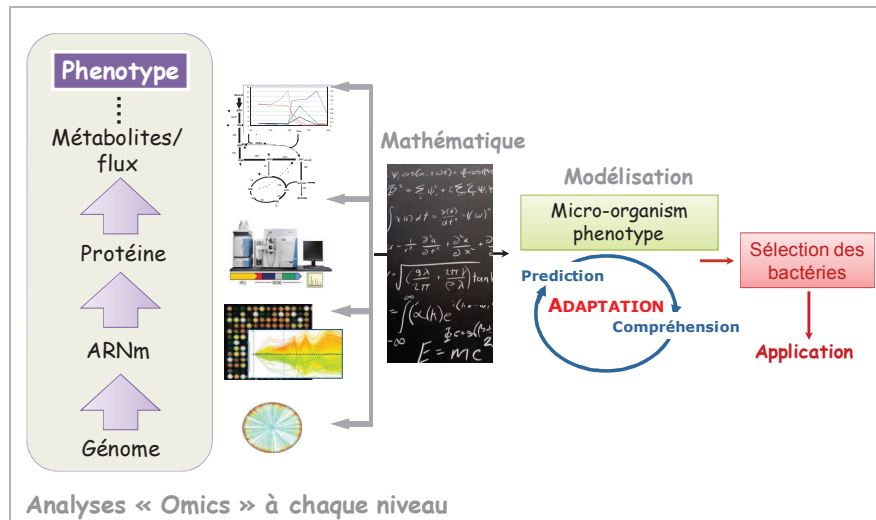


Figure 17 : Représentation de la démarche d'analyse de biologie des systèmes appliquée à la sélection de souches.

a. Transcriptome

Introduit par Velculescu en 1997, le terme de transcriptome se définit comme la mesure de l'expression des gènes et se présente comme un maillon essentiel entre le génome d'un organisme et sa physiologie [Velculescu et al. 1997]. Nous ne reviendrons pas ici sur les aspects techniques de la mesure du transcriptome (pour revue, [Blow 2009]), mais présenterons les avancées majeures apportées par cette technique dans la compréhension des réponses des lactocoques aux différents stress. La réponse transcriptomique de *L. lactis* IL1403 lors d'une carence en carbone [Redon et al. 2005a; Redon et al. 2005b], en azote (isoleucine) [Dressaire et al. 2011], et d'une régulation du taux de croissance [Dressaire et al. 2008] a été analysée en milieu synthétique, et l'importance des paramètres relatifs à l'abondance et à la stabilité des transcrits, estimée. Ainsi, l'entrée en phase stationnaire se traduit-elle par une augmentation de la stabilité des messagers. La réponse à un stress acide en matrice laitière est une réponse adaptative de grande importance pour les LAB. Une étude combinant transcriptomique et mesure de métabolites a été menée sur un levain industriel *L. lactis* LD61, permettant ainsi de préciser les effets du stress acide sur le métabolisme carboné et de mettre en évidence des réponses originales (induction du métabolisme de voies de sauvetage des purines, induction de fonctions phagiques) [Raynaud et al. 2000]. Des études visant à élucider les principaux réseaux de régulation de *L. lactis*, relatifs notamment à la biosynthèse des purines et pyrimidines [Kilstrup & Martinussen 1998; Kilstrup et al. 2005; Fernández & Zúñiga 2006], à la dégradation des peptides [Doeven et al. 2005], aux régulateurs transcriptionnels [Kok et al. 2005; van Hijum et al. 2009] et à des régulateurs transcriptionnels globaux tels que CcpA [Zomer et al. 2007] et CodY [Guédon et al. 2005; den Hengst et al. 2005], ont été menées. L'analyse globale de la dynamique d'adaptation en milieu lait a également été réalisée récemment sur la sous-espèce *cremoris* et a permis la reconstruction du réseau global de régulation transcriptionnelle en lait

[de Jong et al. 2013]. D'autres études transcriptomiques se sont attachées à comprendre les interactions de *L. lactis* avec d'autres micro-organismes de la flore fromagère, comme les levures [Maligoy et al. 2008], les bactéries lactiques non comprises dans les levains (NSLAB) [Desfossés-Foucault et al. 2013] ou potentiellement néfastes, comme *Staphylococcus aureus* [Nouaille et al. 2009; Cretenet et al. 2011] au sein de la matrice alimentaire.

b. Protéome

Défini comme le complément protéique du génome [Wasinger et al. 1995], le protéome a donné naissance à la protéomique qui, par analogie, correspond à de la génomique fonctionnelle mais au niveau des protéines. En effet, l'étude de la régulation des ARNm n'est qu'un premier niveau dans l'analyse du fonctionnement cellulaire. La mesure de l'abondance des protéines devrait être théoriquement plus riche en informations, puisque dans la cellule les protéines sont responsables de la plupart des réactions alors que les ARNm ne sont que des intermédiaires de synthèse. De plus l'existence de régulations post-transcriptionnelles vient perturber la corrélation entre quantité d'ARNm et protéines [Picard et al. 2009; Dressaire et al. 2010; Picard et al. 2013].

Même si l'analyse du protéome s'est longtemps faite, et se fait encore, grâce à la séparation des protéines par électrophorèse en deux dimensions (2D) [Kalia & Gupta 2005], de nombreuses améliorations ont été apportées en termes de marquage, d'identification des protéines et d'analyse d'image afin de rendre cette technique reproductible et quantitative avec la création de carte protéique. Cependant, afin d'éviter la séparation sur gels et augmenter le nombre de protéines identifiées, des techniques alternatives ont été mises au point (pour revue, [Gevaert et al. 2007]). Ces méthodes appartiennent à la famille dite « shotgun » qui est basée sur la digestion *in vitro* par une enzyme (souvent la trypsine) d'un mélange complexe de protéines, suivie par une analyse en spectrométrie de masse des peptides obtenus afin d'identifier les différentes protéines du mélange. Différentes techniques existent pour réaliser une quantification absolue ou relative de ce mélange de protéines : i) les méthodes fondées sur le marquage isotopique comme l'ICAT (Isotope-Coded Affinity Tag) [Gygi et al. 1999] ou encore iTRAQ (Isobaric tags for Relative and Absolute Quantification) [Ross et al. 2004] et ii) des méthodes dites « sans marquage ». Les caractéristiques de chaque méthode ainsi que leurs domaines d'applications ont été largement discutées dans la littérature [Wu & MacCoss 2002; Taylor et al. 2007; Yates et al. 2009; Bantscheff et al. 2012].

Malgré ces avancées technologiques, il n'existe à l'heure actuelle et d'après nos connaissances, que des protéomes obtenus par électrophorèse bidimensionnelle pour déterminer l'abondance et les modifications post-traductionnelles de *L. lactis*. En moyenne, la carte protéomique de *L. lactis* IL1403

comprend 238 protéines identifiées, soit environ un quart des protéines cytoplasmiques à pH acide [Guillot et al. 2003; Gitton et al. 2005; Kilstrup 2005]. Des analyses différentielles comparatives en carence de purines ont caractérisé l'induction de protéines de la voie de synthèse des nucléotides, la réduction des protéines du métabolisme général (synthèse protéique, métabolisme carboné ou des GTPases), puis une induction des enzymes glycolytiques et des facteurs d'élongation. De manière similaire à l'analyse transcriptomique, les études d'adaptation protéique au stress acide ont révélé que l'expression de 90 protéines était modifiée lors de l'autoacidification à pH 5 [Budin-Verneuil et al. 2005], alors que l'ajout d'acide chlorhydrique n'induit que peu de protéines pour cette même valeur [Frees et al. 2003]. L'étude des protéines positivement dérégulées en condition acide, sur des mutants résistants au stress acide, a révélé les protéines suivantes : RecA, PycA, PyrG, GltX, RpsA, et DnaN [Budin-Verneuil et al. 2007]. Des études comparatives, là encore, ont cherché à comprendre l'adaptation au stress thermique chez différentes LAB [Wouters et al. 2001; Derzelle et al. 2003]. Enfin, une étude protéomique a prouvé la surexpression d'une cinquantaine de protéines impliquées dans la nutrition azotée (peptidolyse) et la synthèse des purines (synthèse de novo et voies d'interconversion) lors de la croissance des souches sur lait comparativement à une culture sur milieu complexe [Gitton et al. 2005].

c. Métabolome

Par extension des termes en « -ome », celui de métabolome a été avancé pour définir l'ensemble des métabolites présents dans une cellule et leur quantification. Résultant de l'activité des protéines dans l'environnement cellulaire, les métabolites présentent une diversité considérable par comparaison au nombre limité de gènes et de protéines. Du fait de la diversité de leurs propriétés physico-chimiques (volatilité, polarité, ou encore tailles diverses) il n'existe pas une technique de choix pour leur mesure. De plus, l'application des cribles en matrice laitière nécessite d'adapter les méthodes traditionnelles pour obtenir des analyses sensibles de qualité. La complémentarité des méthodes entre couplage d'une technique chromatographique pour la séparation (chromatographie gazeuse (GC), liquide (LC), électrophorèse capillaire (EC)) et la spectrométrie de masse a démontré son efficacité aussi bien pour les composés volatiles [Alonso & Fraga 2001; Goupy et al. 2000; Ligor et al. 2008] que pour les métabolites exocellulaires ou intracellulaires en matrice alimentaire [Herrero et al. 2012]. Bien que moins sensible, la spectroscopie par résonance magnétique nucléaire (RMN) a été également employée avec succès pour décrire la dynamique d'assimilation des différents sucres de *L. lactis* [Neves et al. 2002; Neves et al. 2005]. Les interactions entre métabolisme carboné/azoté, anabolisme/catabolisme ont été investiguées à travers différentes études qui ont démontré la très faible proportion de source carbonée incorporée dans la biomasse (5 %) [Novák & Loubiere 2000], ou encore l'augmentation de l'efficacité énergétique du métabolisme

à taux de croissance élevé ($0,6 \text{ h}^{-1}$) [Lahtvee et al. 2011]. L'acquisition de ces données métaboliques en parallèle des mesures physiologiques et de l'analyse génomique a été utilisée pour proposer un modèle métabolique de la souche *L. lactis* ssp. *lactis* IL1403 comprenant 621 réactions et impliquant 509 métabolites [Oliveira et al. 2005].

d. Approches combinées

Quelques rares études combinant des données génomiques, transcriptomiques et phénotypiques ont été conduites ces dernières années. L'analyse de souches d'isolats fromagers traditionnels par approche génétique (PFGE et MLST) et phénotypique (utilisation d'hydrates de carbone, profil d'activités enzymatiques, production de COVs, et résistance aux antibiotiques) a permis d'évaluer la diversité de l'espèce *Lactococcus lactis* [Fernández, et al. 2011]. De plus, cette étude a démontré la partition des souches en deux sous-espèces, *lactis* et *cremoris*, et la proximité de ces deux groupes aussi bien sur la base de critères technologiques que génétiques, et cela à la différence d'autres auteurs qui identifient des souches au génotype *cremoris* mais au phénotype *lactis* [Felis et al. 2007; Rademaker et al. 2007]. Cependant, aucune intégration de ces deux niveaux de diversité n'a été réalisée pour quantifier la part de la diversité génotypique face à celle phénotypique. De plus, les déterminants de cette diversité phénotypique n'ont pu être identifiés. Pour répondre à cette question, la recherche de relations entre la présence de gènes et des caractéristiques phénotypiques de souches de *L. lactis* a été menée [Bayjanov et al. 2012; Bayjanov et al. 2013]. Ces études ont défini près de 1388 relations gènes-phénotype et amélioré l'annotation fonctionnelle de certains clusters (utilisation de mélibiose ou d'arabinose). Cependant, de précédentes recherches ont démontré que le partage d'un large génome cœur (plus de 2000 gènes pour six souches laitières de *L. lactis*) n'exclue pas la présence de profils d'expression génétiques souches-dépendant [Tan-a-ram et al. 2011]. Ce résultat met ainsi en lumière l'importance de la prise en compte des régulations transcriptionnelles [Guédon et al. 2002; Klaenhammer et al. 2007; Dressaire et al. 2009; Eckhardt et al. 2013] et post-transcriptionnelles [Kok et al. 2005; Dressaire et al. 2010; Picard et al. 2009; Picard et al. 2012] au-delà de la simple présence de gènes pour expliquer les différences phénotypiques. Ces études ont notamment permis d'identifier des réponses originales de la bactérie, comme la mise en place d'une réponse à la carence en lactose en matrice fromagère chez *L. lactis* alors que la concentration résiduelle est encore importante, mettant ainsi en évidence une limitation vraisemblable de la diffusion du lactose dans la matrice [Cretenet et al. 2011].

Cependant, de nombreux points restent encore à éclaircir pour décrire les liens fonctionnels entre le génotype d'une souche laitière et son expression pour conduire au phénotype en environnement laitier. Le principal frein réside dans l'intégration « mathématique » et « biologique » des différents niveaux « -omiques » afin de tendre vers une évaluation prédictive des potentialités d'adaptation

d'une souche dans un environnement donné. En particulier, l'acquisition et l'intégration de données phénotypiques multicritères (acidification, production aromatique, activité réductrice, rhéologie ..) représentatives de la fonctionnalité des souches dans leur environnement laitier d'utilisation restent un challenge. Cette difficulté s'explique principalement par la diversité de l'appareillage analytique à utiliser et l'absence de technique de criblage à haut débit en matrice lait.

Méthodes statistiques

L'utilisation d'études comparatives à haut débit n'aurait pu se développer sans les progrès concomitants des méthodes statistiques pour en permettre l'analyse. Le traitement statistique des données suit le schéma classique présenté Figure 18 : vérification de la reproductibilité des données (valeurs aberrantes, manquantes, distributions « anormales »...), traitement de normalisation par transformation (surtout quand les variables sont hétérogènes ; voir pour revue des traitements, [Berg et al. 2006], analyse unidimensionnelle avec des statistiques descriptives (moyenne, écart-type) et enfin analyse multidimensionnelle ou différentielle des données. Dans cette partie, nous nous attacherons à introduire les principes des méthodes d'analyses multidimensionnelles utilisées en raison de leur rôle clef dans la caractérisation d'échantillons.

L'analyse multidimensionnelle, et les méthodes qui en découlent, s'applique lorsque l'on souhaite analyser un ensemble d'objets (différentes conditions expérimentales ou souches par exemple) à partir d'un vaste ensemble de variables quantitatives (métabolites, transcrits, protéines..) ou qualitatives (appartenance à une espèce, un écotype..). Intuitivement, pour étudier ces objets, il pourrait-être envisagé de les modéliser dans un espace à p dimensions, avec p étant le nombre de variables. Mais dès lorsque p est supérieur à 3, les représentations graphiques deviennent rapidement complexes et peu explicites. Afin de simplifier l'interprétation des données, les méthodes d'analyse multidimensionnelle proposent de réduire le nombre de dimensions à étudier pour résumer l'information. Leur développement est lié aux deux grands avantages qu'elles présentent [Munck et al. 1998]:

- Caractérisation des liens globaux existant entre un nombre important de variables et des objets afin de décrire de manière succincte les principales informations structurantes du jeu de données. Ceci est particulièrement intéressant en métabolomique, protéomique, et transcriptomique où l'information recueillie est vaste, et où il est nécessaire d'opérer un tri pour se focaliser sur les principaux messages biologiques permettant d'expliquer les différences entre plusieurs conditions.
- Utilisation dans la plupart des cas d'une approche géométrique avec des représentations graphiques pour les variables et les objets étudiés qui facilitent l'interprétation du résultat.

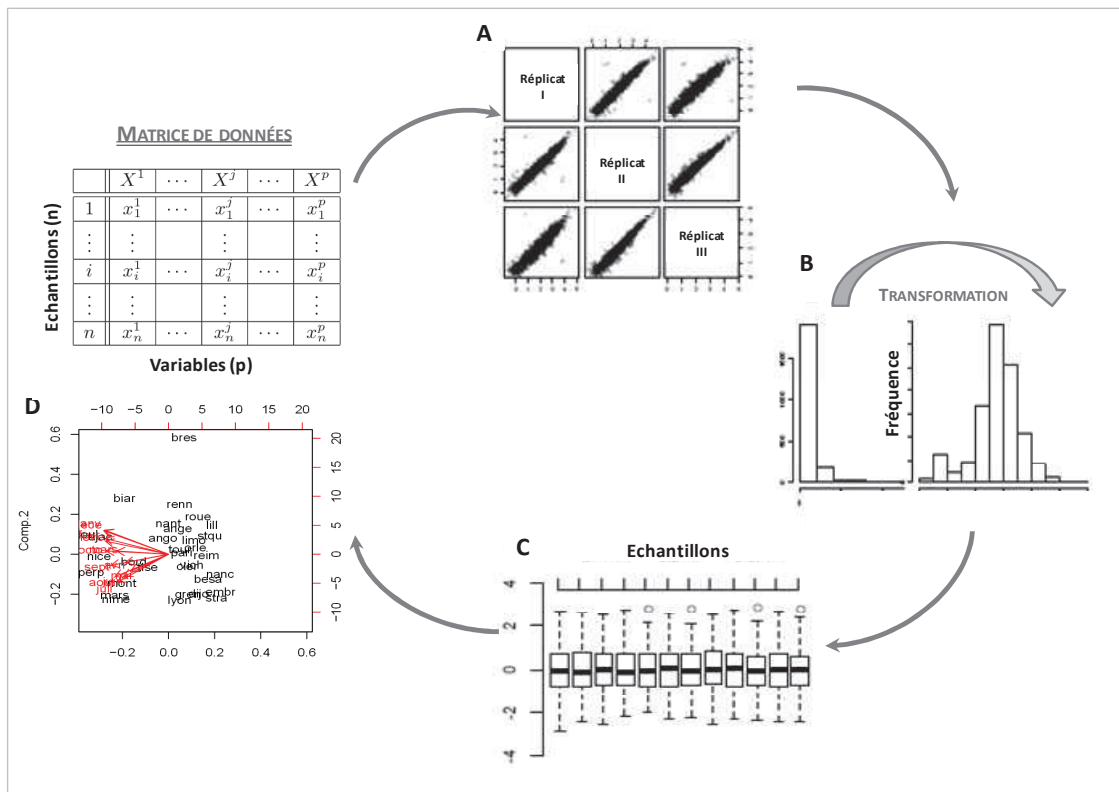


Figure 18 : Représentation graphique du processus de traitement et d'analyse des données. A : Vérification de la reproductibilité des données (Nuage de points). B : Analyse de la distribution des données et traitement de normalisation (histogramme). C : Analyse descriptive unidimensionnelle (Boîtes à moustaches). D : Analyse exploratoire multidimensionnelle (Biplot d'une Analyse en Composantes Principales).

Il existe deux grandes familles au sein de l'analyse de données multidimensionnelles, les méthodes dites descriptives non supervisées qui visent à résumer, structurer et classifier l'information, et les méthodes dites supervisées qui tendent à expliquer une réponse dite dépendante par un ensemble de variables [Parvin 2004]. Le diagramme décisionnel présenté Figure 19 permet de présenter les cas d'application des différentes méthodes recensées.

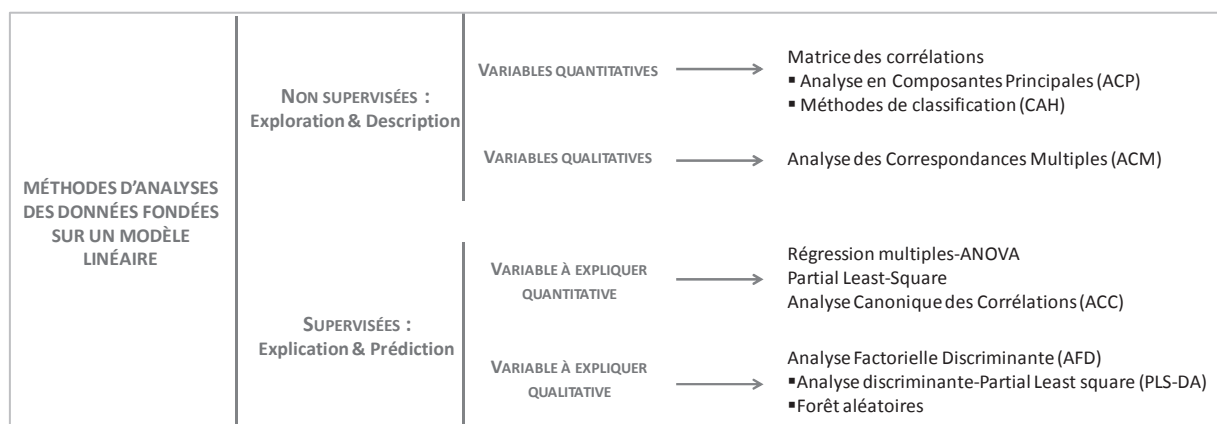


Figure 19 : Diagramme décisionnel pour le choix de la méthode multidimensionnelle à appliquer. Les méthodes précédées d'un carré seront développées dans ce manuscrit.

a. *Analyse multi-variables non supervisée, étude de la diversité*

Développées dans les années 60, ces méthodes d'analyses non supervisées sont des méthodes de choix pour étudier la diversité des échantillons (regroupement éventuel des échantillons en classes en fonction de certains critères) et comprendre les interactions entre les gènes ou métabolites et les différentes conditions, cela sans aucun a priori sur les éventuelles liaisons existantes.

▪ Analyse en Composantes Principales (ACP)

L'ACP est une méthode factorielle de réduction de dimension pour l'exploration statistique de données quantitatives complexes (n individus $\times$ p variables). Cette réduction de dimension s'effectue en définissant un nouveau set de variables (ou axes factoriels), de dimensions q inférieures aux données initiales ($q < n$) tout en tentant de conserver le maximum d'information (e.g. la variance du jeu de données ou inertie). Ces nouvelles variables sont appelées composantes principales (CPs), elles sont non corrélées entre elles, orthogonales et hiérarchisées par la fraction totale d'information qu'elles résument (Figure 20). Ces CPs sont les vecteurs propres de la matrice de covariance des données, elles correspondent à des combinaisons linéaires des variables initiales qui maximisent l'inertie (détails mathématiques, voir [Anderson 1963; Jolliffe 2002]). La représentation des individus dans ce nouvel espace (majoritairement sur les 3 premières PC) permet de mettre en évidence les éventuelles tendances au sein des individus et identifier si une structure, non connue a priori au sein des données, se révèle, ou si des individus se regroupent « naturellement » suivant telle ou telle condition expérimentale. La représentation des variables révèle les corrélations entre les variables et les CPs permettant ainsi d'interpréter les CPs.

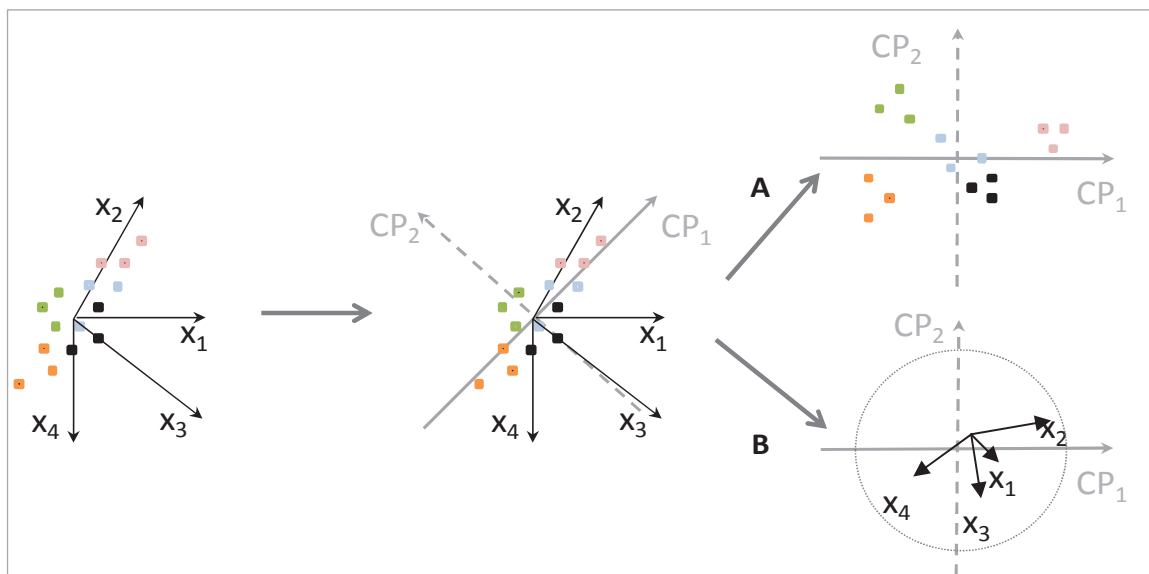


Figure 20: Principe de l'ACP. Réduction de la dimension du jeu de données de $n=4$ par construction de deux Composantes Principales (CP) maximisant l'inertie, et projection des objets (A) et des variables (B) dans le nouvel espace factoriel défini par les CPs. B. Cercle des corrélations (entre +1 et -1 si données centrées et réduites) qui indique la corrélation entre la variable et la CP.

L'efficacité de l'ACP pour extraire l'information principale d'un jeu de données explique ses nombreuses utilisations dans toutes les déclinaisons des techniques «-omiques», telles que : i) l'analyse de données transcriptomiques [Yeung & Ruzzo 2001], ii) la recherche de bio marqueurs en réponse aux problématiques de diagnostics santé par l'aide à l'interprétation d'empreintes métaboliques dans le plasma ou les urines [Ellis et al. 2007; Ellis & Goodacre 2006], iii) le contrôle qualité et réglementaire des produits alimentaires [Arvanitoyannis et al. 1999], ou encore, iv) l'étude de la dynamique des interactions métaboliques dans une co-culture [Howell et al. 2006]. Des caractérisations de la spécificité de produits fermentés (fromages, vins et bières) en fonction de leur composition en micro-organismes ou de la notion de terroir ont pu être menées par l'utilisation de l'ACP [Ayad et al. 1999; Howell et al. 2006; Genovese et al. 2007; Pope et al. 2007].

- Classification Ascendante Hiérarchique (CAH)

La classification est un outil qui dépasse le cadre strict de l'exploration puisque l'objectif est la recherche d'une présence de segmentation dans le jeu de données, c'est-à-dire une répartition des individus en classes. Il existe de nombreuses techniques de classification qui visent toutes à répartir n individus caractérisés par p variables en un nombre m de classes [Hand 2012]. Chacune de ces classes a pour vocation d'être la plus homogène possible, et les différentes classes entre elles, les plus distinctes. Dans le cadre d'une classification non supervisée, il n'y a pas d'a priori connu sur la typologie. Le principe de la CAH, ou « clustering » en anglais, repose sur l'agrégation deux à deux des individus les plus proches pour créer une partition, et ainsi de suite par itération jusqu'à regrouper les deux groupes les plus éloignés. Le résultat des différentes partitions créées est présenté sous forme d'un arbre ou dendrogramme, contenant l'historique de la classification entre les différents individus et permettant de retrouver $n-1$ partitions effectuées. Il est nécessaire, si cela n'est pas déjà fourni, de calculer au préalable une matrice de distances entre les individus et donc de se munir d'une métrique pour définir ces distances inter-individus (distance euclidienne, associée à la corrélation...), et de fixer une règle pour agréger un groupe d'individus à un autre groupe, appelé critère d'agrégation (minimum, maximum, moyen ou critère de Ward [Ward 1963] par décomposition de l'inertie). Une fois le dendrogramme construit, il est nécessaire de se donner « une hauteur de coupe » pour obtenir une classification. C'est un choix arbitraire de l'utilisateur qui est d'autant plus pertinent que le nombre de classes sélectionnées correspond à une distance inter-nœuds importante. Les CAH réalisées à partir des distances euclidiennes et par sélection du critère de Ward donnent des résultats complémentaires à l'interprétation de l'ACP, ce sont les principaux paramètres sélectionnés.

b. Analyse multi-variables supervisée, analyse de la spécificité

Aux analyses exploratoires descriptives (ACP, CAH), s'opposent les analyses supervisées, plus récentes, pour lesquelles l'objectif est d'identifier les variables participant à expliquer, voire prédire, une réponse imposée. Cette réponse peut-être quantitative (un rendement, un taux de survie d'une bactérie à un stress...) ou qualitative, et crée une partition volontaire des individus en classes (patients sains vs malades, distinction entre des souches en fonction de leur environnement d'isolement ou de leur sous-espèce). Dans le cas des classifications supervisées, les objectifs d'analyses sont doubles : d'une part, l'évaluation de la pertinence de la classification établie et éventuellement la ré affectation des individus dans de nouvelles classes si le taux d'erreur est trop important, et d'autre part l'identification des variables discriminantes de la distinction imposée. Parmi la longue liste de méthodes appartenant à cette famille, il sera introduit la régression catégorielle des moindres carrés, ou PLS-DA, et les algorithmes de forêts aléatoires ou Random forest (RF).

- Régression catégorielle des moindres carrés, PLS-DA

Développée dans les années 2000 par Enciso & Tenenhaus, l'analyse discriminante des moindres carrés, ou *Partial least squares discriminant analysis* (PLS-DA), est définie comme une régression partielle des moindres carrés (PLS) [Antoniadis et al. 2003; Boulesteix 2004] d'une variable binaire catégorielle Y à partir d'un set X de variables prédictives [Pérez-Enciso & Tenenhaus 2003]. Cette méthode représente un compromis entre la traditionnelle analyse discriminante linéaire (LDA) et la méthodologie classique d'ACP. L'analyse des résultats s'effectue à partir de représentations graphiques (individus et variables) similaires à celles décrites pour l'ACP pour apprécier la qualité de séparation des individus et identifier les variables discriminantes. Testée récemment sur de nombreux jeux de données transcriptomiques et métabolomiques où le nombre de prédicteurs (variables) était très largement supérieur au nombre d'observations (individus) et de données multi colinéaires, l'identification des marqueurs génétiques et métaboliques pour chaque classe a pu être réalisée avec succès [Pérez-Enciso & Tenenhaus 2003; Steinfath et al. 2010; Viguerie et al. 2012; Li et al. 2012; Ray et al. 2012]. En effet, les nouvelles variables créées, ou composantes de la PLS, sont choisies cette fois-ci en ne cherchant pas à maximiser l'inertie du jeu de données mais en maximisant l'adéquation à la variable à prédire. Un taux d'erreur de prédiction de classement peut-être également calculé pour estimer la qualité des données à prédire la variable qualitative.

- Forêts aléatoires, Random Forest

Les forêts aléatoires, ou « Random forest », introduites par Breiman en 2001, sont des algorithmes basés sur des modèles d'agrégation utilisés aussi bien pour les problématiques de classification que de régression [Breiman 2001]. Ils appartiennent à la famille des algorithmes d'apprentissage, ou *Machine Learning*, qui débutent par la construction d'un set de classificateurs (échantillons d'entraînement) avant de passer à la classification des nouveaux individus en procédant à un vote (pondéré) de leurs prédictions (fonction de la qualité du set de classificateurs) [Dietterich 2000b; Dietterich 2000a]. Il n'y a donc pas, à la différence des autres méthodes présentées ci-dessus, la construction de nouvelles variables (PC ou PLS composantes). Dans le cas des forêts aléatoires, l'algorithme de classification débute par une sélection aléatoire d'un sous-set de variables de l'échantillon (bootstrapping) pour chaque nœud de l'arbre de décisions afin d'assigner chaque individu à une classe préalablement définie. Cette procédure va être reproduite jusqu'à l'attribution de chaque individu dans une classe et la constitution d'un arbre de décisions. Ces arbres de décisions, ou CART (Classification And Regression Trees), ont été mis au point dans les années 70-80 par Breiman également [Breiman 1984]. Dans le cadre des forêts aléatoires, ce processus est reproduit itérativement un grand nombre de fois ($n = 5000$ classiquement) et les résultats de ces arbres de décisions sont combinés entre eux pour attribuer à chaque individu sa classe définitive en fonction de la classe totalisant le plus d'occurrence. Pour chaque forêt créée, une erreur de prédiction de classification (« Error out of Bag ») est calculée par rapport à la classification souhaitée pour évaluer l'efficacité dans le choix de variables pour prédire la réponse à expliquer. Ces algorithmes se sont révélés très efficaces, notamment sur de larges jeux de données peu discriminants vis-à-vis de la réponse souhaitée, ce qui explique qu'ils soient de plus en plus utilisés, comme en témoigne une recherche sur NCBI « Random forest analysis », avec 2492 réponses d'articles (le 3/08/2013) dont plus d'un tiers ces cinq dernières années. En particulier, on notera leur succès dans l'identification des liaisons entre gènes et phénotypes chez *Lactococcus* [Bayjanov et al. 2012], la sélection de marqueurs génétiques de la sous-espèce *lactis* et de l'écotype laitier [Siezen et al. 2011] ou encore la mise en évidence de marqueurs métaboliques spécifiques lors de co-cultures de *S. thermophilus* [De Bok et al. 2011].

c. La sélection de variables, recherche de marqueurs de spécificité

L'une des problématiques majeures de l'analyse de données issues des techniques à « haut débit » est le déséquilibre résidant entre la taille considérable des données acquises au niveau des variables mesurées (gènes, protéines, métabolites) face, bien souvent, aux quelques dizaines d'observations dans l'échantillon des individus. Bien que les méthodes multi-variables présentées soient capables de résumer l'information afin d'identifier des variables discriminantes, il est souvent intéressant de

sélectionner un nombre restreint de variables. En effet, disposer d'un sous-set de variables simplifie la constitution de modèles, facilite l'interprétation biologique, supprime les effets non significatifs et améliore la distinction des phénomènes observés. De très nombreuses techniques de sélection de variables ont été développées ces dernières années [Dash & Liu 1997; Saeys et al. 2007; O'Hara & Sillanpää 2009; Bolón-Canedo et al. 2013]. Cependant, il n'existe pas de « meilleures » méthodes, et le seul consensus au sein de la communauté scientifique est d'adapter la technique sélectionnée en fonction de la problématique biologique et du type de données. Ainsi, le premier ensemble de techniques de sélection présenté s'applique tout particulièrement aux méthodes multivariées géométriques qui définissent de nouvelles variables (PC ou PLS). Les variables peuvent donc être sélectionnées en fonction d'un seuil de contribution à chacun des axes. Il est également possible de modifier la procédure de constitution des axes, comme dans le cas de la sparse-ACP/PLS. Dans ce cas, au lieu de combinaisons linéaires de variables, les composantes principales sont définies comme un problème d'optimisation de régression où les nouvelles variables sont obtenues après imposition d'une contrainte (type *lasso*) pour fixer un nombre de variables à utiliser [Zou et al. 2004; Tibshirani 2011]. Cette méthode a été en particulier utilisée avec succès pour l'analyse de données transcriptomiques [Lee et al. 2012; Kawano et al. 2012] et métabolomiques [Allen & Maletić-Savatić 2011; Jiang et al. 2013] afin d'identifier des marqueurs de prévalence aux cancers ou l'origine géographique de différents individus. Le second type de technique est directement lié à l'utilisation des forêts aléatoires, puisque ces dernières classifient les variables selon leur rang d'importance [Breiman 2001; Karpievitch et al. 2009; Boulesteix et al. 2012]. Une variable est d'autant plus importante que le brassage des données de cette variable et la ré affectation à de nouveaux individus augmente l'EOB. Là encore, cette technique a été utilisée récemment pour sélectionner les variables discriminantes sur différents jeux à l'échelle « omique » [Bok et al. 2011; Díaz-Uriarte & Alvarez de Andrés 2006; Karpievitch et al. 2009; Chen et al. 2013].

PROBLEMATIQUE

PROBLEMATIQUE

Lactococcus lactis est la principale espèce employée en tant que ferment mésophile pour la production de fromages et de laits fermentés [Fox 2000; Hui et al. 2003; Hui et al. 2005]. Dans le cas applicatif des laits fermentés, *L. lactis* ssp. *lactis* assume, au cours de la phase de maturation, des fonctions métaboliques multiples (production d'acide lactique, dégradation de macromolécules pour la synthèse de composés aromatiques, synthèse éventuelle d'exopolysaccharides ou de bactériocines) qui vont profondément impacter la flaveur, la texture et la préservation du produit. Autrefois, les principales fonctionnalités recherchées sur un ferment étaient la qualité texturale et la flaveur des produits. Dans cet objectif, la sélection de ferment est donc longtemps restée empirique, basée sur l'évaluation gustative des produits pour identifier les souches qui feront de « bons ferments » [Leroy & De Vuyst 2004]. Désormais, les producteurs laitiers recherchent de nouvelles souches afin d'améliorer la sécurité, la qualité et l'attractivité de leur produits. En effet, certaines de ces bactéries peuvent exercer des effets bénéfiques sur la santé, directement en tant qu'organismes vivants (rôle probiotique décrit il y a déjà plus de 100 ans) [Metchnikoff & Mitchell 1908] ou indirectement par la sécrétion de métabolites secondaires aux propriétés pro-santés [Stanton et al. 2005]. De nouveaux produits sont sans cesse mis sur le marché pour répondre aux attentes des consommateurs en termes de santé (produits allégés en matière grasse, probiotique), d'aromatisation (fraîcheur, douceur, goût crème) ou encore de « naturel » (sans ajout d'excipient : arôme et gélifiant) [Leroy & De Vuyst 2004]. Satisfaire ces demandes impose de disposer d'un réservoir de souches avec une palette de signatures fonctionnelles de plus en plus vaste afin de singulariser les produits face à la concurrence [Hansen 2002; Mills et al. 2010]. De plus, dans l'objectif de rationaliser la production, les recherches se tournent également vers l'identification de ferments maîtrisables dans l'environnement de production et pouvant notamment résister aux attaques de phages, et aux différents stress rencontrés lors de la préparation (survie à la lyophilisation) [Hansen 2002; Mills et al. 2006]. Ainsi, la sélection de nouveaux ferments de *L. lactis* ssp. *lactis* représente un enjeu important pour le développement de l'industrie laitière.

Deux options sont possibles pour identifier de nouvelles souches de *L. lactis* ssp. *lactis*, d'une part créer de la diversité par ingénierie génétique, et d'autre part, caractériser la diversité naturelle des souches pour mieux l'exploiter. L'avènement du séquençage et de l'annotation fonctionnelle des génomes bactériens de l'espèce *L. lactis* (9 génomes disponibles en septembre 2013) a facilité l'ingénierie métabolique ces dernières années [Teusink & Smid 2006]. Cependant, bien que des

caractéristiques cibles aient été améliorées telles que la production de nisine [Qiao et al. 2013], la synthèse d'arômes [Tanous et al. 2002; Yvon et al. 2003; Yvon et al. 2006], ou encore la résistance aux bactériophages [Barrangou et al. 2012], la méfiance des consommateurs et les réglementations sanitaires européennes freinent la mise sur le marché de ces souches génétiquement modifiées [Pedersen et al. 2005; Sybesma et al. 2006]. Ainsi, à l'heure actuelle, il semblerait que l'identification de nouvelles souches, à la signature singulière et utilisables dans un contexte industriel, repose principalement sur une problématique de caractérisation de la diversité phénotypique naturelle de souches de *L. lactis ssp. lactis*.

La sélection de ferment est un enjeu d'autant plus important qu'il a été identifié un fort caractère souche dépendant des propriétés fonctionnelles recherchées [Ayad et al. 1999; Cachon et al. 2002; Delgado & Mayo 2004; Bachmann et al. 2009b; Fernández et al. 2011; Bayjanov et al. 2013]. Ces recherches ont également démontré que la sous-espèce *lactis* présente des niveaux de diversité variables en fonction de la niche écologique [Kelly & Ward 2002; van Hylckama Vlieg et al. 2006; Nomura et al. 2006]. Les nombreuses études conduites sur l'analyse de la diversité phénotypique de la sous-espèce *lactis* se sont concentrées sur des caractéristiques précises : activité acidifiante [Cachon et al. 2002], potentiel d'oxydoréduction [Brasca et al. 2007], texture [Kristo et al. 2003], production d'arômes [Cheng 2010; Smit et al. 2004], biosynthèse de nisine [Aly et al. 2012], ou encore autolyse [Valence et al. 2000]. Cependant, caractériser la diversité fonctionnelle de souches de *L. lactis ssp. lactis* nécessite de cribler un éventail de réponses métaboliques et technologiques afin d'obtenir une image du comportement global des souches dans leur fonction de ferment. Quelques rares études ont présenté une analyse multicritères de la diversité phénotypique de la sous-espèce *lactis*. Mais ces recherches, non intégralement réalisées en matrice lait, empêchent toute prédiction de la fonctionnalité de la souche dans l'environnement laitier. En effet, l'expression des différentes fonctions technologiques du ferment varie en fonction du procédé laitier considéré (matière première, température, interaction avec d'autres souches...) [Foda et al. 1974; Becker & Puhán 1989; Gastaldi et al. 1997; Kristo et al. 2003]. En particulier, les profondes modifications qu'implique l'adaptation à une matrice lait ont été démontrées aussi bien au niveau transcriptomique, protéomique et enzymatique [Gitton et al. 2005; Bachmann et al. 2009b; Cretenet et al. 2011; de Jong et al. 2013]. Ainsi, l'identification de souches à la signature phénotypique singulière impose une analyse multicritères, en intégrant des indicateurs phénotypiques pertinents (vitesse d'acidification, potentiel redox, concentration en acides organiques volatiles, autolyse ...) dans un environnement le plus proche possible de l'application. Cependant, l'acquisition de ces données phénotypiques multicritères en haut débit et en matrice laitière demeure un challenge

analytique et seul un nombre restreint de méthodes a été proposé [Nicolaou & Goodacre 2008; Bachmann et al. 2009b].

Des études récentes ont combiné analyses phénotypiques et caractérisation génétique pour rechercher des marqueurs génétiques prédictifs de la diversité fonctionnelle des souches laitières et ainsi remplacer les fastidieuses études phénotypiques [Bayjanov et al. 2012; Bayjanov et al. 2013]. Ces études ont défini près de 1388 relations gènes-phénotype et amélioré l'annotation fonctionnelle de certains clusters. Cependant, elles ont échoué à décrire des liens forts entre le génotype, son expression, et le phénotype laitier observé. En effet, de précédentes recherches ont démontré que le partage d'un large génome cœur (plus de 2000 gènes pour six souches laitières de *L. lactis*) n'exclue pas un polymorphisme entre les profils transcriptomiques, enzymatiques et phénotypiques de souches laitières [Bachmann et al. 2009a; Tan-a-ram et al. 2011]. De plus, le criblage d'une collection de souches productrices de diacétyle/acétoïne a révélé que des souches ne possédant aucun des gènes du catabolisme du citrate sont capables de produire des quantités équivalentes à celles en disposant [Passerini et al. 2013]. Ce résultat remet en question les procédures classiques d'identification de ces souches uniquement basées sur des déterminants génétiques. Il démontre également la nécessité de mettre en œuvre une approche intégrative multi-échelles (génétique, transcriptomique, protéomique et métabolique) pour comprendre l'origine de cette diversité phénotypique, et éventuellement prédire la spécificité des souches. Ainsi de nombreux points restent encore à éclaircir pour décrire les liens fonctionnels entre le génotype d'une souche et son expression au sein d'un environnement particulier pour conduire au phénotype laitier.

Dans l'optique d'évaluer la diversité fonctionnelle de souches de *L. lactis* ssp. *lactis* et d'améliorer la qualité prédictive de l'utilisation de ferments, nous avons, dans cette thèse, caractérisé la spécificité de souches, en culture pure en lait, en appliquant une approche de biologie intégrative. Pour ancrer cette thématique dans un environnement applicatif fort, nous avons collaboré avec un industriel (Yoplait) et centré cette étude sur l'analyse du comportement d'une souche-ferment modèle, *L. lactis* ssp. *lactis* *ND5F* (propriété de Yoplait, abrégée *ND5F*). Cette souche présente une singularité organoleptique non expliquée, aux notes « crème » et « fraîche » dans les produits laitiers fermentés. L'objectif de cette thèse vise à identifier les déterminants de la spécificité organoleptique de *ND5F* dans sa fonction de ferment laitier par une approche intégrée du génotype (gènes et génome) au phénotype (protéines, métabolites, indicateurs physiologiques et technologiques). Pour caractériser ces marqueurs, la signature de *ND5F* a été confrontée à la diversité d'un panel de souches représentatives de la phylogénie de l'écotype « domestiqué » (MLST) [Passerini et al. 2010].

L'enjeu est ainsi d'acquérir une visibilité sur l'origine de cette spécificité pour pouvoir ensuite moduler cette signature en fonction du procédé.

Pour répondre à cette problématique, cette thèse s'est articulée autour de trois axes principaux. Dans un premier temps, nous nous sommes intéressés à décrypter le comportement global de la souche d'intérêt en lait, dans une condition de référence standardisée, afin d'analyser l'impact de l'environnement sur la réponse physiologique mise en oeuvre au cours de la culture. Cette étude restreinte aux vingt-quatre premières heures de culture dressera un diagnostic macro-cinétique de l'activité de la souche modèle. Par ailleurs, pour évaluer la diversité fonctionnelle des levains, il sera réalisé, en matrice industrielle (lait enrichi en matière protéique et matière grasse), une étude phénotypique multicritères et intégrée d'un panel de souches représentatif de la variabilité génotypique. Cette étude vise à proposer une organisation fonctionnelle de cette sous-population et à identifier les principaux marqueurs de cette diversité. Cette étude sera renforcée par une analyse comparative de la diversité protéique des levains afin de proposer une classification génotype-phénotype prédictive des souches. Le positionnement de notre souche modèle au sein de cette diversité a pour objectif d'identifier les déterminants de la spécificité organoleptique de cette souche. Le troisième axe sera consacré à l'identification des leviers et verrous de la production métabolique de la souche d'intérêt afin de moduler sa signature aromatique.

MATERIELS & METHODES

I Les souches

Les 11 souches bactériennes utilisées au cours de cette étude sont des *Lactococcus lactis* ssp. *lactis* « domestiquées » sélectionnées à partir de levains commerciaux (Danisco et Yoplait), ainsi qu’au sein de diverses collections de laboratoire (Laboratoire de Microbiologie et Génétique Moléculaires, LMGM-Toulouse, France, pour les souches Sx et EIPx ; Laboratoire de Microbiologie Alimentaire, LMA-Caen, France, pour la souche UCMAx). La principale souche de cette étude, *ND5F*, propriété exclusive de Yoplait, a été fournie sous forme de deux biovariants de provenance différente désignés *ND5F-Lab* et *ND5F-Danisco*. La capacité des souches à pousser en lait, critère préalable à toute caractérisation en culture pure en lait, a été testée par la présence de deux marqueurs génétiques, le gène *PrtP* codant la protéase de paroi et le gène *lacE*, codant la lactose perméase du complexe phosphotransférase [PTS], et a été confirmée par des tests phénotypiques. De plus, quatre souches ont été identifiées comme appartenant au biovar *diacetylactis* par la présence du gène *citP* codant la citrate perméase. Les caractéristiques des 11 souches de l’étude sont présentées au Tableau 6.

Tableau 6 : Caractéristiques des souches utilisées dans cette étude.

Souche	Origine	Complexe clonal	Séquence Type	Gènes d'intérêts			Référence
				<i>citP</i>	<i>lacE</i>	<i>PrtP</i>	
ND5F-Lab	Levain (Yoplait)	CC1	15	-	+	+	Cette étude
ND5F-Danisco	Levain (Yoplait)						
S86 (S477)	Levain	CC1	10	-	+	+	[Passerini et al. 2010]
UCMA5713 (S432)	Prairie (France)	CC1	18	-	+	+	[Passerini et al. 2010]
S87 (S479)	Levain	CC1	15	+	+	+	[Passerini et al. 2010]
EIP41A (S528)	Lait cru (France)	CC1	15	-	+	+	[Dhaisne et al. 2013]
EIP33A (S540)	Lait cru (France)	CC1	15	-	+	+	[Dhaisne et al. 2013]
DIA-A	Choozit DIA-A FRO (Danisco)	CC1	15	+	+	+	[Dhaisne et al. 2013]
EIP33F (S541)	Lait cru (France)	CC1	34	+	+	+	[Dhaisne et al. 2013]
EIP37F (S525)	Lait cru (France)	CC1	34	+	+	+	[Dhaisne et al. 2013]
MC70	Choozit MC-70 FRO (Danisco)	CC2	7	-	+	+	[Dhaisne et al. 2013]

Fermentation

I.1 Milieux de culture

Plusieurs types de milieux de cultures ont été utilisés pour la culture des souches (laits, milieux complexes, ou synthétiques). Les milieux les plus utilisés ont été formulés à partir d’une base de lait de vache. Afin de reproduire au plus près les conditions industrielles, différentes compositions pour les laits ont été employées. Premier type de lait, le lait écrémé est reconstitué à partir de poudre de lait écrémé, PLE (essentiellement constituée de matière azotée et de lactose, voir le Tableau 7 pour

sa composition) à 109 g.L^{-1} puis stérilisée par autoclavage (15' à 115°C). Il a été utilisé aussi bien pour des analyses de croissance que pour la préparation de l'intégralité des précultures (dites «levains») des essais en lait.

Tableau 7: Composition de la poudre de lait écrémée (données Yoplait).

Elément	Concentration (g.kg^{-1})
Matière grasse (MG)	0,52
Matière azotée totale (MAT)	350,50
-Azote non protéique (NPN)	16,70
-Protéine	333,00
Glucides (Lactose)	513,60
Calcium	$1016,50 > 1240 < 1551,50 \text{ (mg.100g}^{-1}\text{)}$
Phosphore	$952,30 \text{ (mg.100g}^{-1}\text{)}$
Sodium	$535,00 \text{ (mg.100g}^{-1}\text{)}$

Les autres laits employés dans cette étude se différencient du lait écrémé, d'une part par le traitement thermique de pasteurisation (72°C pendant 15 secondes), et d'autre part par leurs compositions standardisées en matière grasse (MG) et matière protéique (MP). La première est ajustée à partir d'addition de crème, alors que la seconde est contrôlée par ajout de PLE. Le lait utilisé pour l'étude comparative de *ND5F* aux autres souches est un lait « classique » standardisé à 35 g.kg^{-1} de MG et 40 g.kg^{-1} de MP. Les autres laits utilisés, neuf au total, ont été formulés spécifiquement pour étudier les effets de la matière grasse et protéique sur la signature de *ND5F* selon la matrice d'essais déduite de l'application de la méthodologie des plans d'expériences (Matériels & Méthodes I.24). Les compositions en MP et MG des différents laits employés dans cette étude sont répertoriées Tableau 8.

Tableau 8 : Composition en matière protéique et grasse des 10 laits employés dans cette étude.

Type de lait	Matière Protéique (g.kg^{-1})	Matière grasse (g.kg^{-1})
Lait écrémé	38	0
Lait « classique »	40	35
Tests plan d'expériences	33	0
		20
		40
		0
	43	20
		40
		0
		20
	53	40
		40

Deux autres milieux complexes ont été employés : le milieu commercial M17 (Oxoid Ltd., Basingstoke, Hampshire, Angleterre) et un milieu à base d'extrait de levure, YE. Le milieu M17 est

reconstitué conformément aux indications du fournisseur, et stérilisé par autoclavage. Afin de se rapprocher du milieu lait, l'ensemble des cultures sur M17 et sur YE (composition du milieu présentée Tableau 9) sont réalisées avec du lactose à 50 g.L⁻¹ comme source de carbone. Ainsi les milieux sont-ils supplémentés en lactose à partir d'une solution de lactose 4X (200 g.L⁻¹) stérilisée indépendamment. Le pH de ces deux milieux est également ajusté à 6,60 à l'aide d'une solution de base de KOH 10N.

La concentration en tampon phosphate du YE est augmentée d'un facteur trois lorsque ce milieu est utilisé pour une culture à pH libre de façon à ralentir l'acidification du milieu et permettre une phase de croissance plus longue, en l'absence d'une régulation du pH.

Tableau 9: Composition du milieu extrait de levure, YE en g.L⁻¹.

Composés	Concentration (g.L ⁻¹)	
	pH libre	pH régulé
Extrait de levure	10	10
K ₂ HPO ₄	7,5	2,5
KH ₂ PO ₄	9	3
MgSO ₄	0,2	0,2
MnSO ₄	0,05	0,05
Lactose	50	50

Le dernier milieu utilisé dans cette étude est un milieu synthétique chimiquement défini, MCD [Cocaign-Bousquet et al. 1995] employé pour l'étude des besoins nutritionnels de *ND5F*. La composition détaillée est présentée dans le Tableau 10.

Trois autres milieux dérivés du MCD ont été confectionnés également (voir Tableau 10) : i) MCD-AA 3X, supplémenté à 3X dans tous les acides aminés composant le MCD, ii) MCD-Vit 3X, enrichi à 3X pour toutes les vitamines contenues dans le MCD et iii) MCD-AA/Vit 3X, augmenté à 3X pour toutes les vitamines et tous les acides aminés du MCD. Les cultures sur MCD sont réalisées avec du lactose à 50 g.L⁻¹.

Comme pour les cultures en YE, la concentration en sels phosphates est augmentée pour les cultures à pH libre. Le pH est ajusté à 6,60 également. Les milieux MCD supplémentés ou appauvris en un ou deux acides aminés cibles ainsi qu'en extrait de levure à 0,8% (HYPA-Biospringer-France) ont été également utilisés.

Tableau 10 : Composition des milieux de culture chimiquement définis, en g.L⁻¹. Concentrations modifiées pour les cultures à pH libre ou précultures (). En gras les concentrations du MCD de référence modifiées.

Constituants	MCD	MCD-AA 3X	MCD-Vit 3X	MCD-AA/Vit 3X
Lactose	50	50	50	50
Acétate de sodium	1	1	1	1
Citrate d'ammonium	0,6	0,6	0,6	0,6
KH ₂ PO ₄	3 (9)	3	3	3
K ₂ HPO ₄	2,5 (7)	2,5	2,5	2,5
Acides aminés				
Alanine	0,24	0,72	0,24	0,72
Arginine	0,12	0,36	0,12	0,36
Asparagine	0,34	1,02	0,34	1,02
Cystéine	0,17	0,51	0,17	0,51
Glutamine	0,51	1,53	0,51	1,53
Glycine	0,17	0,51	0,17	0,51
Histidine	0,11	0,33	0,11	0,33
Isoleucine	0,2	0,6	0,2	0,6
Leucine	0,47	1,41	0,47	1,41
Lysine-HCl	0,35	1,05	0,35	1,05
Méthionine	0,12	0,36	0,12	0,36
Phénylalanine	0,28	0,84	0,28	0,84
Proline	0,68	2,04	0,68	2,04
Sérine	0,34	1,02	0,34	1,02
Thréonine	0,23	0,69	0,23	0,69
Tryptophane	0,05	0,15	0,05	0,15
Tyrosine	0,29	0,87	0,29	0,87
Valine	0,33	0,99	0,33	0,99
Sels				
MgCl ₂ , 6 H ₂ O	0,2	0,2	0,2	0,2
FeSO ₄ , 4 H ₂ O	0,011	0,011	0,011	0,011
CaCl ₂ , 2 H ₂ O	0,05	0,05	0,05	0,05
ZnSO ₄ , 7 H ₂ O	0,005	0,005	0,005	0,005
COCl ₂ , 6 H ₂ O	0,0025	0,0025	0,0025	0,0025
Vitamines				
Ac. P-aminobenzoïque	0,01	0,01	0,03	0,03
Biotine	0,01	0,01	0,03	0,03
Cyano-cobalamine	0,001	0,001	0,003	0,003
Ac. Folique	0,001	0,001	0,003	0,003
Inosine	0,005	0,005	0,015	0,015
Ac. Nicotinique	0,001	0,001	0,003	0,003
Ac. Orotique	0,005	0,005	0,015	0,015
Pantothénate de calcium	0,001	0,001	0,003	0,003
Pyridoxamine	0,005	0,005	0,015	0,015
Pyridoxine	0,002	0,002	0,006	0,006
Riboflavine	0,001	0,001	0,003	0,003
Thiamine	0,001	0,001	0,003	0,003
Ac. D,L-6,8-thioctique	0,0025	0,0025	0,0075	0,0075
Bases azotées				
Déoxy-thymidine	0,005	0,005	0,005	0,005
Adénine	0,01	0,01	0,01	0,01
Guanine	0,01	0,01	0,01	0,01
Uracile	0,01	0,01	0,01	0,01
Xanthine	0,01	0,01	0,01	0,01

I.2 Techniques de culture

L'ensemble des cultures en fioles-bouteilles en YE ou lait ont été effectuées en mode statique sans régulation de pH, et majoritairement en conditions de micro-aérobiose (aération initiale du milieu puis arrêt des échanges gazeux avec le milieu). La température de la culture est fixée à 28°C. Les cultures en fermenteurs ont été réalisées sous agitation dans différentes conditions de régulations de température, de pH, d'oxygène et d'atmosphère gazeuse (aérobiose, micro-aérobiose et plus rarement anaérobiose).

I.2.1 Conservation des souches

Les bactéries sont cultivées sur le milieu riche M17 en mode statique, à 30°C et prélevées en phase exponentielle de croissance (densité optique à 580 nm, $DO_{580} \approx 2$), pour être aliquotées dans des cryotubes de 2 mL en présence de 20% volume/volume de glycérol, et stockées à -80°C.

I.2.2 Précultures

Les précultures sont réalisées dans des flacons en verre sertis et imperméables à l'oxygène, en mode statique et sans régulation de pH. Le milieu et la durée des précultures sont adaptés en fonction du milieu final utilisé pour les cultures. Ainsi, pour une croissance en milieu lait, une unique préculture (200 mL) est réalisée à partir de lait écrémé stériliséensemencé afin d'obtenir une concentration finale de $1,5 \cdot 10^9$ UFC.mL⁻¹ après 13h de maturation à 37°C. Cette préculture sera utilisée pour inoculer à 10% en volume la culture finale des bouteilles ou fermenteurs. Dans le cas des essais sur YE et sur MCD, les milieux utilisés sont respectivement du M17 et du MCD enrichi en tampon phosphate. Les fioles sontensemencées (1% vol/vol) à partir de cryotubes de conservation contenant la souche souhaitée (voir I.2.1) et incubées 12 heures à 30°C. Dans ces deux cas, suivant le volume de culture final envisagé, des précultures successives dans des fioles de volume croissant sont réalisées (20, 40, 60, 500 mL) en s'assurant que les cellules soient toujours inoculées en phase exponentielle de croissance. Les bactéries de la dernière préculture sont concentrées par centrifugation (6000 rpm, 10 min, 4°C) et lavées à l'eau physiologique afin d'éviter un phénomène de dilution ou un apport exogène en nutriments, avant d'inoculer les fermenteurs à une DO_{580} cible initiale de 0,01.

I.2.3 Cultures discontinues en bouteilles

Les précultures en lait écrémé sont utilisées pour inoculer (10%, vol/vol) 500 ml de lait pasteurisé (MG 35 g/kg, MP 40 g/kg) dans des bouteilles de 1L. Le pH est non régulé (valeur initiale à 6.60) mais suivi en ligne (Sonde pH 405-DPAS-SC-K8S, Mettler Toledo, France), grâce au module d'enregistrement (BIOSTAT, Sartorius, Allemagne), tout comme le potentiel REDOX (Sonde Redox

(Pt4805-DPA-SC-S8, Mettler Toledo, France) et la pression partielle en oxygène dissous, PO_2 (InPro6900, Mettler Toledo, France). Les fermentations sont conduites à 28°C et stoppées quand le pH atteint 4,60. La culture est alors transférée et stockée dans des contenants stériles, à 4°C, conservés de 1 à 27 jours.

I.2.4 Cultures discontinues en fermenteur

Les cultures discontinues sont réalisées dans des fermenteurs de 2 litres (BIOSTAT, Sartorius, Allemagne). La température, le pH (Sonde pH 405-DPAS-SC-K8S, Mettler Toledo, France), l'agitation et la pression partielle en oxygène dissous (Pt4805-DPA-SC-S8, Mettler Toledo, France) sont mesurés et régulés en ligne. L'agitation est fixée à 100 rpm, à l'exception des cas de régulation de la pression partielle en oxygène, PO_2 , où en fonction de la consigne de cette dernière, l'agitation ainsi que le flux d'air en entrée seront variables et ajustés pour tendre à la consigne. Lorsque les cultures sont réalisées à pH régulé (6,40 ou 5,5 en consigne), il est ajusté par ajout automatique de KOH 10 N. Le potentiel REDOX (Sonde Redox Pt4805-DPA-SC-S8, Mettler Toledo, France) est également enregistré en ligne, tout comme le volume d'addition de base. Avant utilisation, les sondes PO_2 , REDOX et pH sont préparées (ajout/vérification d'électrolyte), polarisées puis calibrées (sauf REDOX) avant stérilisation. La calibration de la sonde PO_2 est réalisée sur de l'eau osmosée. Le milieu YE est stérilisé directement dans le fermenteur, et le lactose concentré stérile est rajouté stérilement. Les milieux chimiquement définis MCD-Lactosesont stérilisés en ligne via une cartouche de filtration Sartoban de porosité de 0,2 μm (Sartorius-Allemagne) avant d'être introduits dans le fermenteur. Enfin, les laits pasteurisés et le lait écrémé sont introduits via une pompe et un tuyau stérile dans le fermenteur après leur traitement thermique. La température est fixée à 28°C, mais des essais à 33 et 40°C ont été également conduits. Dans le cas des cultures anaérobies, le milieu stérile est dégazé à l'azote pendant 30 minutes puis maintenu sous pression d'azote (20 mbars).

Pour les cultures en aérobiose, le milieu bénéficie d'une aération permanente par bullage d'air dans le fermenteur, tandis que pour les conditions de micro-aérobiose, le milieu est laissé en contact avec l'atmosphère ambiante via une sortie stérile qui est fermée après inoculation. Dans le cas des traitements initiaux du lait, de l'air et de l'azote sont envoyés dans le milieu afin d'atteindre une PO_2 stable de 100, 50 ou 0% dans le fermenteur avant de fermer les arrivées et sorties d'air du fermenteur lors de l'inoculation. Durant la fermentation, des surnageants de culture sont congelés (-20°C) après élimination des cellules par centrifugation (8000 rpm, 10 minutes, 4°C) toutes les 30 minutes jusqu'à la fin de la culture pour les analyses des métabolites.

I.3 Caractérisation de l'activité acidifiante

La mesure de l'activité acidifiante a été réalisée soit à partir du dispositif CINAC (CINétique d'ACidification) de suivi de l'acidification en continu sur plusieurs cultures en parallèle (Données CYD-Vienne), soit à partir des données de pH enregistrées en ligne sur les fermenteurs. Les mesures fournies par les sondes sont lissées et leur traitement permet de caractériser l'activité acidifiante des souches en fonction des conditions environnementales. Cinq descripteurs calculés à partir de l'étude en temps réel des cinétiques d'acidification ont été utilisés pour l'étude [Spinnler & Cordieu 1989]: $V_{\max}^a$ la vitesse maximale d'acidification en unité pH.min⁻¹; $T_{\max}^a$ le temps nécessaire en heure pour atteindre $V_{\max}^a$; $pH_{\max}^a$ le pH en unité pH correspondant à $V_{\max}^a$; T^a la durée de la phase de latence correspondant à une diminution de 0,08 unité pH. h⁻¹; et $T_{pH4,6}$ le temps en heure nécessaire pour atteindre un pH de 4,6. Chaque culture est réalisée à minima en double sur 24 h (réplicat biologique). Les pH initiaux, avant ensemencement, et à 24 h, sont vérifiés sur une même électrode externe afin de s'assurer qu'il n'y a eu aucune dérive des électrodes au cours des 24 heures de mesure.

I.4 Mesure du potentiel d'oxydoréduction

L'activité réductrice des souches a été suivie en ligne et en temps réel au cours des cultures (sonde Redox Pt4805-DPA-SC-S8, Mettler Toledo, France) et enregistrée par le module d'acquisition des fermenteurs. La mesure directe du potentiel redox (E_m) est convertie en potentiel redox par rapport à une électrode de référence en hydrogène, d'après l'équation suivante : $Eh = E_m + E_r$. Dans le cas de la sonde Ag/AgCl utilisée, le potentiel E_r peut-être calculé à partir de l'équation suivante tenant compte des modifications du potentiel en fonction de la température :

$$E_r = 225,24 - 0,7329 \times T, \text{ où } T, \text{ la température en } ^\circ\text{C de la culture (Mettler Tolédo). Equation 1}$$

Comme les valeurs du potentiel Eh sont pH-dépendantes [Jacob 1970], elles sont toutes ramenées à un pH 7 virtuel pour s'affranchir des effets de l'acidification des souches, dit Eh_7 , par l'inclusion d'un coefficient de corrélation α entre le Eh et le pH :

$$Eh_7 = Eh - \alpha(7 - pH) \text{ avec } \alpha = 38 \text{ pour du lait et } 55 \text{ pour du YE après détermination de la pente } Eh=f(pH)$$

Equation 2

Ainsi, à partir de la mesure de E_m , le Eh_7 peut être estimé à partir de l'équation suivante :

$$Eh_7 = E_m + E_r + \alpha(7 - pH)$$

Equation 3

Par analogie avec l'étude de l'activité acidifiante, les cinq descripteurs suivants décrits par Cachon [Cachon et al. 2002] ont été calculés : $V_{\max}^r$ la vitesse maximale de réduction en mV.min⁻¹; $T_{\max}^r$ et $pH_{\max}^r$ respectivement le temps (h) et le pH correspondant à $V_{\max}^r$; Eh_7^r , le Eh_7 en mV correspondant à $V_{\max}^r$; et Eh_7^f , le Eh_7 en mV correspondant à la fin de culture.

Techniques analytiques

I.5 Mesure de la biomasse

I.5.1 Turbidimétrie

La croissance est évaluée instantanément par mesure de l'absorbance différentielle à une longueur d'onde de 580 nm (DO_{580}) à l'aide d'un spectrophotomètre (Biochrom Libra S4) entre la culture et un échantillon du milieu de culture non inoculé, en diluant si nécessaire les échantillons pour rester dans la gamme de linéarité de l'appareil ($DO_{580} < 0,6$). Cette technique est applicable pour les milieux M17, YE et MCD. Le lait est, quant à lui, un milieu très opaque qui ne permet pas un suivi direct de la croissance par mesure directe de l'absorbance, même après dilution de l'échantillon. Une clarification du lait est alors nécessaire [Mercade 2002] : un volume de culture « lait » est mélangé à deux volumes d'une solution de clarification (EDTA 15 mM, NaOH 0,15 mM). Le culot cellulaire est lavé 3 fois à l'aide d'une solution d'eau physiologique (NaCl 0,9 %) puis resuspendu dans un volume d'eau physiologique. L'absorbance à 580 nm de l'échantillon ainsi obtenu peut alors être mesurée. Cette technique applicable au lait écrémé, perd en fiabilité lorsque le pourcentage en matière grasse du lait augmente.

I.5.2 Gravimétrie

La masse sèche de la biomasse contenue dans un échantillon de culture en milieu YE de volume connu est déterminée après filtration au travers d'une membrane de porosité de 0,45 μ m, lavage, séchage des filtres à l'étuve sous vide (24 h, 200 mbar, 60°C) et pesée (Balance de précision à 0,0001g-Sartorius). La masse sèche est obtenue par la différence entre les masses du filtre après et avant filtration. La corrélation entre la DO_{580} et la concentration en biomasse (X, en g/L) pour la souche *L. lactis ssp. lactis ND5F* a pu être établie : $X = 0,331 \times DO_{580}$ ($R^2=0,999$).

I.5.3 Dénombrements, viabilité

Les dénombrements sont réalisés sur boîtes M17-agar en triplicat après application de 100 μ L de solution bactérienne à la dilution adaptée. Après une incubation de 48 h à 30°C, le nombre de colonies est compté (UFC.mL⁻¹). Tout comme pour la masse sèche, une corrélation en milieu YE entre la DO_{580} et le nombre de bactéries (Y, en UFC.mL⁻¹) pour la souche *L. lactis ssp. lactis ND5F* a pu être établie : $Y = 5.10^8 \times DO_{580}$ ($R^2=0,997$). Le calcul du pourcentage de viabilité cellulaire est déduit de la différence entre le nombre de bactéries comptées sur les échantillons de laits fermentés à J14 et ceux à J1.

La viabilité à 14 jours des bactéries a été qualitativement évaluée sur les culots de cultures en lait fermenté (centrifugation 5 min à 14 000 rpm, à 4°C) dilués au 10^{ème} et préparés dans une solution de tricitrate de sodium (0,03 M, pH 7) et de NaCl (0,3 M). Une double coloration avec les fluorophores DAPI (4',6'-diamidino-2-phénylindole) [Life technologies, DAPI solution 5 mg.mL⁻¹, D1306] et IP (iodure de propidium) [Invitrogen, iodure de propidium solution 1 mg.mL⁻¹, P3566] a été réalisée pour distinguer au microscope à fluorescence les cellules à la membrane lysée (PI, coloration rouge) du nombre total de cellules (DAPI, coloration bleue). Les solutions mères des deux marqueurs DAPI et PI ont été préparées respectivement à 50 µg.mL⁻¹ et 10 µg.mL⁻¹ extemporanément dans une solution de tricitrate de sodium (0,03 M, pH 7) et de NaCl (0,3 M). Ces solutions ont été conservées à 4°C à l'abri de la lumière pour une utilisation dans la journée. Le DAPI, spécifique des liaisons à l'ADN (bases adénine et thiamine) absorbe la lumière à 350 nm et émet une fluorescence bleue brillante (450-490 nm). L'IP est un agent intercalant des acides nucléiques qui ne pénètre que les cellules à la membrane lysée ; il absorbe à 535 nm et émet une lumière rouge (600-620 nm). Pour 25 µL d'échantillon dilué, l'IP a été utilisé à une concentration finale de 5 mM (incubation de 5 min, à 4°C) suivi de l'ajout de DAPI à une concentration finale de 12,5 µg.mL⁻¹ (incubation de 5 min, à 4°C). Les lames ont été préparées et scellées pour observation au microscope (Leica FW 4000 B), objectif x100 aux longueurs d'ondes d'excitation et d'émission correspondantes pour chaque marqueur.

I.5.4 Composition élémentaire de la biomasse

La composition élémentaire de la biomasse de *L. lactis* ND5F n'a pas été déterminée pour le milieu MCD utilisé ni pour les autres milieux synthétiques modifiés. Par conséquent, nous utiliserons la composition précédemment déterminée pour la souche *L. lactis* ssp. *lactis* NCDO 2118 sur MCD et M17, soit C_{4,63} H_{7,89} O_{2,35} N_{1,0} pour une masse molaire de 129 g.mol⁻¹ [Novak et al. 1997]. A noter que la variabilité relative à la composition du milieu est inexistante entre un milieu synthétique MCD et un milieu riche tel que M17. La variabilité inter souche de la composition élémentaire de la biomasse au sein de *L. lactis* est peu documentée ; elle sera estimée à 10 %.

I.6 Métabolites exocellulaires

I.6.1 Dosage des sucres, acides organiques, et éthanol par HPLC

Le suivi cinétique des concentrations en substrats (lactose, citrate) et produits (acides organiques : lactate, acétate, formiate, diacétyle, acétoïne et éthanol) a été effectué en routine par chromatographie liquide haute pression (HPLC) (Agilent Technologies 1200 series, Waldbronn, Allemagne) équipée d'une colonne échangeuse de cations (Bio-Rad HPX87H) maintenue à 48°C et éluée avec de l'H₂SO₄ 5mM à un débit de 0,05 mL.min⁻¹. Les surnageants prélevés au cours des

cultures en lait ont été traités comme décrit dans le premier article de cette thèse [Dhaisne et al. 2013]. En ce qui concerne les échantillons prélevés en MCD, l'étape de clarification des échantillons a été remplacée par une simple filtration sur filtre de porosité 0,45 µm.

I.6.2 Dosage des sucres, acides organiques, et éthanol par RMN

Ces éléments dosés par HPLC sur lait ont été vérifiés qualitativement et quantitativement par Résonnance Magnétique Nucléaire, RMN. Pour ce faire, 50 µL d'une solution d'acide triméthylsilyl propanoïque deutérée (TMSP-d₄) à 50 mM est additionnée à 350 µL de D₂O comme étalon interne, avant d'être rajoutée à 100 µL de lait fermenté. La solution est homogénéisée au vortex et est ensuite analysée par un spectromètre RMN (Bruker 500 Ultrashield, programme Topspin 3.0) en 1D-¹H. Le site Biological Magnetic Resonance Data Bank (www.bmrb.wisc.edu) renseigne sur les profils RMN obtenus pour les différents composés. A partir de l'intégrale des pics RMN des différents acides organiques et de leurs formules élémentaires, il est possible de quantifier l'élément en solution :

$$\text{Concentration de l'élément en mM} = \frac{[\text{Dilution (5)} \times \text{Nbr de proton de } [TMSP - d_4] (9) \times [[TMSP - d_4] finale] \times \text{Intégrale du pic}}{\text{Intégrale du pic de l'étalon interne}} \times \text{Conc. de l'étalon interne}$$

Equation 4

I.6.3 Détermination de la concentration en acide lactique non dissocié

Les concentrations en acide lactique non dissocié ont été calculées selon la relation des acides faibles en solution diluée :

$$[\text{acide lactique non dissocié}] = \frac{[\text{Acide lactique total}]}{1 + 10^{(pH - pK_a)}}, \text{ où } pK_a \text{ de l'acide lactique est égal à } 3,86. \text{ Equation 5}$$

I.6.4 Dosage des acides aminés par HPLC

Les acides aminés libres du milieu de fermentation sont analysés à partir des surnageants de fermentation, ceux-ci devant subir un pré-traitement afin de précipiter les protéines qui risqueraient d'endommager le système de dosage. Un volume de surnageant et quatre volumes de méthanol sont incubés pendant une nuit dans de la glace, puis centrifugés. Le surnageant récupéré, peut alors être dosé. Le dosage des acides aminés au sein des surnageants de fermentation est également réalisé par HPLC (Agilent Technologies 1200 series, Waldbronn, Allemagne) suivant le protocole décrit dans le premier article de cette thèse [Dhaisne et al. 2013].

I.6.5 Dosage des composés organiques volatils par SPME-GC/MS dans les laits fermentés

Le dosage des composés organiques volatils présents dans les laits fermentés produits a été réalisé à 14 J par Micro Extraction en Phase Solide (SPME) en chromatographie gazeuse (GC) et détection par spectrométrie de masse (MS) à la plate-forme de l'Institut du vin et de la vigne (ISVV, Bordeaux, France). Le protocole complet est présenté dans le premier article de cette thèse [Dhaisne et al. 2013]. Brièvement, 5g d'échantillon sont introduits dans un vial SPME de 20 ml en verre auquel il est rajouté 250 µl d'un mélange d'étalons internes ($[^2\text{H}_5]$ éthyl acetate (ethyl acetate- d_5), $[^2\text{H}_{15}]$ octanoic acid (octanoic acid- d_{15}) et $[2,2\text{-}^2\text{H}]$ 3-methylbutanal (3-methylbutanal- d_2)). L'injection est réalisée par SPME avec un injecteur automatique (Gerstel MPS2). Le vial est incubé 30 minutes à 55 °C et le temps d'extraction par la fibre SPME (StableFlex, référence catalogue 57299-U) est de 90 minutes à la même température. La fibre est désorbée directement dans l'injecteur du chromatographe (Agilent technologie HP5890, Waldbronn, Allemagne) à une température de 250 °C pendant 300 secondes. Le programme de température du four débute par un isotherme à 40 °C pendant 5 minutes puis atteint 240 °C à raison de 3 °C.min⁻¹. Le temps total de l'analyse chromatographique est de 82 minutes (colonne phase polyéthylène glycol polaire de type SGE BP 21). La détection est effectuée par spectrométrie de masse (Agilent technologie HP5972, Waldbronn, Allemagne) selon la méthode SIM (selected ion monitoring). Le détail des 3 fragments spécifiques choisis pour chaque composé ainsi que l'étalon interne de référence sont présentés Tableau 11. Les composés sont semi-quantifiés grâce au rapport entre l'aire du pic du composé et l'aire du pic de l'étalon interne, ramené à un gramme de yaourt.

Tableau 11: Temps de rétention, ions sélectionnés en SIM et étalon interne pour chaque composé identifié. ^a
 En gras, les ions utilisés pour la quantification, les autres sont employés pour l'identification.

No.	Composé	Temps de rétention	Ions sélectionnés ^a	Etalon interne sélectionné
1	Méthanethiol	2,51	47/48/45	Ethyl acetate-d ₅
2	Acétaldéhyde	2,56	43/44/29	3-methylbutanal-d ₂
3	Diméthylsulphide	2,72	62/47/61	Ethyl acetate-d ₅
4	Acétone	3,10	43/58-	3-methylbutanal-d ₂
5	Ethyle acétate	3,69	29/43-	Ethyl acetate-d ₅
6	Butanone	3,90	72/43/57	3-methylbutanal-d ₂
7	(2,3) Methylbutanal	4,08	86/44/57	3-methylbutanal-d ₂
8	Ethanol	4,40	31/45/29	3-methylbutanal-d ₂
9	Butan-2,3-dione (diacetyl)	5,35	43/86/42	3-methylbutanal-d ₂
10	2,3-Pentanedione	7,71	100/57/43	3-methylbutanal-d ₂
11	Butyle acétate	7,93	56/43/61	Ethyl acetate-d ₅
12	Diméthylsulphide	7,93	94/79/45	3-methylbutanal-d ₂
13	2-Méthyl propan-1-ol	9,08	41/43/42	3-methylbutanal-d ₂
14	2-Heptanone	12,48	43/58/71	3-methylbutanal-d ₂
15	3-Hydroxybutan-2-one (acétoïne)	17,53	45/43/88	3-methylbutanal-d ₂
16	3-Méthyl-2-buten-1-ol	19,29	53/71/41	3-methylbutanal-d ₂
17	Hexan-1-ol	20,61	56/55/43	3-methylbutanal-d ₂
18	2-Hydroxypentan-3-one	20,83	45/57/43	3-methylbutanal-d ₂
19	Diméthyltrisulphide	21,40	126/45/79	3-methylbutanal-d ₂
20	2-Nonanone	22,07	58/43/71	3-methylbutanal-d ₂
21	A. Acétique	25,17	45/43/71	A. Octanoïque-d ₁₅
22	Heptan-1-ol	25,26	70/56/43	A. Octanoïque-d ₁₅
23	Furfural	25,39	96/95/39	A. Octanoïque-d ₁₅
24	Benzaldéhyde	26,19	77/105/106	A. Octanoïque-d ₁₅
25	Butan-2,3-diol D ⁺	28,50	45/43/57	A. Octanoïque-d ₁₅
26	Octanol	29,11	56/70/84	A. Octanoïque-d ₁₅
27	Butan-2,3-diol-meso	29,67	45/43/57	A. Octanoïque-d ₁₅
28	Undécan-2-one	30,72	58/43/71	A. Octanoïque-d ₁₅
29	A. Butanoïque	31,25	60/41/73	A. Octanoïque-d ₁₅
30	2-furfuranmethanol	32,73	98/81/97	A. Octanoïque-d ₁₅
31	Nonan-1-ol	34,26	56/70/83	A. Octanoïque-d ₁₅
32	Décan-1-ol	33,93	55/97/112	A. Octanoïque-d ₁₅
33	2-Tridecanone	37,83	58/43/71	A. Octanoïque-d ₁₅
34	A. Hexanoïque	39,58	60/73/87	A. Octanoïque-d ₁₅
35	Diméthylsulfone	41,11	79/94/81	A. Octanoïque-d ₁₅
36	A, Heptanoïque	43,64	60/73/87	A. Octanoïque-d ₁₅
37	δ-Octalactone	44,93	99/71/70	A. Octanoïque-d ₁₅
38	2-Pentadecanone	45,60	58/59/71	A. Octanoïque-d ₁₅
39	A. Octanoïque	48,53	60/73/87	A. Octanoïque-d ₁₅
40	Hexadecanal	48,73	82/57/96	A. Octanoïque-d ₁₅
41	A, Nonanoïque	50,86	60/73/87	A. Octanoïque-d ₁₅
42	δ-Decalactone	52,32	99/71/114	A. Octanoïque-d ₁₅
43	A. Décanoïque	53,34	60/73/87	A. Octanoïque-d ₁₅
44	A, 9-Décenoïque	55,70	55/69/110	A. Octanoïque-d ₁₅
45	δ-Dodecalactone	57,68	99/71/114	A. Octanoïque-d ₁₅
46	A. Benzoïque	60,47	105/122/77	A. Octanoïque-d ₁₅
47	A. Dodecanoïque	60,55	60/73/87	A. Octanoïque-d ₁₅

Analyse génétique

I.7 Extraction d'ADN

Après avoir été cultivées à 30°C sur milieu M17 pendant 12 h, l'ADN génomique des cultures a été extrait à partir du kit « DNeasy™ tissue » selon les instructions du fabricant (Qiagen, Hilden, Germany). Cet ADN a ensuite été utilisé pour réaliser les étapes d'amplification. Les plasmides ont été extraits à partir du kit « QIAprep Spin » (Qiagen, Hilden, Germany).

I.8 Recherche de gènes d'intérêts

La présence de marqueurs génétiques, gène *PrtP* codant la protéase de paroi, gène *lacE* codant la lactose perméase du système PTS, et gène *citP* codant la citrate perméase, a été recherchée par amplification PCR et électrophorèse en gel d'agarose en condition standard. Pour connaître le détail des séquences des primers utilisées ainsi que les conditions d'amplification et révélation, se référer à la publication de cette thèse [Dhaisne et al. 2013].

I.9 MLST

L'analyse MLST a été effectuée à partir du schéma MLST à six loci décrit par Passerini [Passerini et al. 2010]. Il comprend trois gènes de ménage (codant pour des protéines impliquées dans des fonctions cellulaires de base) *glyA*, *pgk* et *pdp*, deux gènes cataboliques, *bcaT* et *pepXP*, et un gène de régulation du système SOS, *recN*. Le logiciel MEGA 5 a été utilisé pour générer et analyser les séquences concaténées support de l'analyse phylogénétique.

I.10 PFGE

Pour l'analyse PFGE, l'ADN génomique extrait des souches est inclus dans une matrice d'agarose, digéré par des endonucléases de restriction, et une électrophorèse en champs pulsé est réalisée sur les fragments d'ADN génomique selon les conditions de la référence [Le Bourgeois et al. 1995]. Pour comparer les profils de restriction, une matrice binaire basée sur la présence ou l'absence de chaque bande est construite et le coefficient de Dice (S_{DS}) pour chaque comparaison deux à deux est calculé. Un dendrogramme basé sur le critère d'agglomération UPGMA a été réalisé pour positionner chaque souche. Le détail du protocole est présenté dans l'article [Dhaisne et al. 2013].

Protéome

L'identification et la quantification des protéines intracellulaires ont été réalisées aussi bien à partir d'échantillon sur milieu synthétique MCD, que sur laits, par la technique dite label free LC-MS, avec la collaboration de la société Phylogène (Bernis, France) pour la partie expérimentale. Dans le cas des essais sur matrice MCD, l'effet de différentes conditions environnementales a été analysé sur la souche *ND5F* par l'expression différentielle de 19 protéines cibles (en aval du pyruvate, impliquées dans les étapes de synthèse de composés aromatiques discriminants, aminotransférases ou impliquées dans la lipolyse). Les expérimentations en lait (composition « classique » avec 40 % de MP et 35 % de MG), ont été conduites sur les onze souches de l'étude avec l'objectif de quantifier l'intégralité des protéines identifiées afin d'analyser si des signatures protéiques existent entre souches. Les principales étapes du protocole sont introduitesci-dessous. Trois réplicats biologiques au minimum sont réalisés pour chacune des conditions, avec deux à trois réplicats techniques.

I.11 Prélèvements et traitements des échantillons

Les étapes concernant le prélèvement et le traitement des échantillons ont été adaptées en fonction des matrices des essais.

I.11.1 Prélèvements et lavage des culots

a. Echantillons sur milieu synthétique

Lorsque les conditions à analyser sont atteintes, des suspensions bactériennes prélevées en MCD, correspondant à une concentration minimale de 50 mg de cellules, sont additionnées de 2 volumes d'EDTA (15 mM-pH 12,9) refroidi à 4°C et immédiatement centrifugés (10 min, 8 000 rpm, 4°C). Le culot bactérien est repris dans 10 mL de tampon refroidi à 4°C PBS (0.2 M pH 6,4 et EDTA 1 mM), puis centrifugé (10 min, 8000 rpm, 4°C). Cette étape est réalisée une nouvelle fois, puis après élimination du surnageant, les culots cellulaire sont stockés à -80°C.

b. Echantillons sur matrice lait

Lors de l'arrêt de la culture à pH 4,60, un volume de lait fermenté correspondant à 50 mg de cellules est prélevé, et 0,33 volume de trisodium citrate (1 M) suivi de 0,13 volume de tampon salin (NaCl 0,145 M, Beta glycerophosphates 0,016 M et 0,1% de Tween 80) sont rajoutés. Les deux tampons sont préalablement refroidis à 4°C. Après homogénéisation par retournement, les tubes sont laissés à 4°C pendant 15 minutes puis centrifugés (10 min, 8000 rpm, 4°C). Ces étapes sont répétées jusqu'à élimination du précipité de caséine. Le culot est ensuite traité comme dans le cas d'un échantillon sur milieu synthétique.

I.11.2 Lyse cellulaire, extraction et digestion des protéines, purification des peptides.

Les culots bactériens sont repris dans 1 mL de tampon ammonium bicarbonate 0,1 M, puis soniqués sur glace (2,5 min à puissance 50 % par pas de 30 s, puis 10 s de refroidissement, puis 30 s à

puissance 70 %). L'étape suivante est le dépôt du lysat sur Amicon Ultra (10 kDa), et la réduction des protéines en présence de DTT. L'alkylation des protéines est réalisée en présence d'iodoacétamide. Les protéines sont alors digérées par la trypsine, puis les peptides sont dessalés sur colonnes C18 (Sep Pak).

I.12 Analyse par LC-MS/MS

I.12.1 Préparation des extraits peptidiques

Les peptides sont dosés par la technique d'acide bicinchonique (Pierce BCA Protein Assay kit, Thermo Fisher) et la concentration en peptides est ajustée à 0,1 $\mu\text{g} \cdot \mu\text{L}^{-1}$, le volume d'injection est de 6 μL .

I.12.2 Conditions chromatographiques et de spectrométrie de masse

Les colonnes chromatographiques utilisées sont des colonnes C18 (75 μm x 15 cm). Un gradient de 5 à 40 % d'acétonitrile est appliqué pendant 95 minutes. Le débit est de 300 $\text{nL} \cdot \text{min}^{-1}$. Les données sont acquises avec un spectromètre de masse de type TripleTof 5600 (ABSciex). Le scan MS est réalisé en condition de haute résolution avec un temps d'accumulation de 250 ms, le scan MS/MS en condition de haute sensibilité sur les 25 ions les plus intenses temps d'accumulation de 100 ms. 2036 cycles ont été réalisés soit 11 cycles (mesures) par pic chromatographique.

I.13 Identification des protéines

Les protéines ont été identifiées contre une base de données issue de NCBI et constituée des séquences de *Lactococcus lactis* ssp. *cremoris* et ssp. *lactis* disponibles sur le site en date de novembre 2012, des séquences de la trypsine de porc et des principales séquences des protéines de caséines dans le cas d'échantillons en lait. L'algorithme PARAGON [Shilov et al. 2007] a été utilisé pour identifier les peptides par comparaison de la base « virtuelle » NCBI avec les données brutes de spectres MS et MS/MS. A ce stade, il est pris en compte d'éventuelles modifications post traductionnelles (alkylation de la cystéine, désamination) pour finir l'attribution de chaque peptide à une protéine. Pour chaque peptide identifié, il est également calculé un taux de faux positif, FDR, à partir de peptides aléatoires générés en inversant la base de données utilisée et en ajoutant en supplément 1.000.000 de peptides de séquences aléatoires (ayant des masses proches de celles observées). La probabilité des peptides de « matcher » avec cette base de données permet de déterminer un taux de FDR pour chaque peptide. Pour l'analyse des protéines cibles du MCD, seules les protéines identifiées avec un seuil de FDR inférieur à 1 % ont été considérées.

I.14 Quantification des protéines

La quantification relative des protéines a été réalisée en utilisant le logiciel Marker View (AbSciex). Après avoir généré une bibliothèque de peptides identifiés par leur temps de rétention et leur m/z (peptides identifiés contre la base de données NCBIInr avec un score de confiance supérieur à 90 % et n'étant ni redondants ni modifiés), les aires des ions parents de la MS sont extraites. C'est sur ces aires qu'il est effectué la quantification relative. Deux séries de données ont été générées, i) aires de chaque peptide identifié, et ii) aire de chaque protéine identifiée : elle est équivalente à la somme des aires des peptides identifiés pour chaque protéine.

I.15 Normalisation

Pour pallier aux différences expérimentales et à la variabilité d'ionisation des différents essais (notamment en fonction du pourcentage de caséine pour les échantillons en lait), les données ont été normalisées par l'aire totale du TIC (Total Ion Current). Pour chaque injection, un TIC_{injection} a été calculé, et la moyenne de ces TIC a permis de définir un TIC_{moyen}.

$$Aire_{protéine, normalisée} = Aire_{protéine, brute} \frac{TIC_{moyen}}{TIC_{injection}}$$

Equation 6

Dans le cas des données en lait, la normalisation a été complétée par une transformation par le logarithme en base de 2 afin de rendre la distribution des données normale. Pour finir, les résultats des différents réplicats techniques ont été moyennés pour chaque réplicat biologique. Pour la suite de l'analyse, les données liées à la trypsine et à la caséine (e.g à 5 protéines) ont été supprimées du « data set » final.

Analyses des données

Le traitement statistique des données génétiques, métaboliques et protéomiques a été intégralement réalisé à partir de l'utilisation du logiciel de statistiques R en accès libre (<http://www.r-project.org/>).

I.16 Normalisation

Afin de procéder à l'analyse statistique, la distribution des données a été explorée afin de déterminer la proximité des réplicats techniques et biologiques (le cas échéant, supprimer des valeurs aberrantes) et choisir si nécessaire le traitement de normalisation le plus adapté. Ainsi, pour être comparées sans biais lié à l'amplitude des valeurs, les données génotypiques et phénotypiques ont été centrées (soustraction par leur moyenne) et réduites (division par la variance).

I.17 Recherche de corrélations

La recherche de corrélations entre données quantitatives s'est faite par le calcul des coefficients de corrélation de Pearson et de Spearman. Le premier mesure la force d'un lien linéaire entre les variables étudiées tandis que le second reflète la corrélation entre leurs rangs. Un seuil d'acceptation tenant compte des erreurs pour l'ensemble des données et non pour chacune individuellement, appelé taux de faux positif (FDR), a été calculé selon la méthode proposée par Benjamini et Hochberg [Benjamini & Hochberg 1995] pour corriger les erreurs liées aux tests multiples. Le seuil de significativité a été fixé à FDR 5 %.

I.18 Analyse multi-variables non supervisée

Afin d'explorer le jeu de données, des Analyses en Composantes Principales (ACP) ont été menées sur les données génétiques, phénotypiques et protéiques en utilisant le package *FactoMineR* de R [Husson et al. 2013] pour examiner les regroupements d'individus (détermination d'ellipses de distribution à 95 %) et pour explorer la proximité des variables. Les dendrogrammes présentés (aussi bien sur les variables que sur les souches) ont été réalisés par Classification Ascendante Hiérarchique (CAH), à partir de la fonction *hclust* et en choisissant comme paramètres de distance inter individus la distance euclidienne et le critère d'agglomération de Ward [Ward 1963].

I.19 Analyse multi-variables supervisée

La première des deux méthodes d'analyses multi-variables supervisées utilisée est la régression catégorielle des moindres carrés, ou PLS-DA, au travers du package *mixOmics* [Lê Cao, González, et al. 2009a; Dejean et al. 2013]. Elle a été employée aussi bien pour l'analyse du degré de taux d'erreur de prédiction de classement des souches à partir des données phénotypiques ou protéomiques que pour l'évaluation de la pertinence d'un set de 14 variables discriminantes de la signature de *ND5F*.

Seconde méthode employée, l'analyse par forêts aléatoires, ou Random Forest, a été conduite à partir du package R du même nom [Cutler & Wiener 2012] sur les données de protéomiques afin d'estimer si des signatures protéiques pour certaines souches peuvent-être identifiées, et d'évaluer la qualité d'une telle classification. Les paramètres utilisés pour paramétrer cette analyse ont été spécifiquement optimisés pour assurer la robustesse et la précision de l'analyse :

- *ntree*, le nombre d'arbres est fixé à 5 000,
- *mtry*, le nombre de variables explicatrices à prendre en compte à chaque nœud est choisi comme égal à la $\sqrt{\text{Nombre de variables du dataset}}$ (paramètre par défaut),
- Nombre de forêts (e.g au nombre de simulation) pour chaque analyse a été imposé à 1000 pour l'analyse de robustesse et la sélection de variable, et 10 en analyses de routine.

I.20 Sélection de variables

Comme déjà mentionné, l'une des problématiques majeures de l'analyse de données issues des techniques à « haut débit » est le déséquilibre résidant entre la taille considérable des données acquises au niveau des variables mesurées (gènes, protéines, métabolites) face aux quelques dizaines d'observations des individus. Afin de sélectionner un nombre restreint de variables, différentes techniques ont été testées. Premièrement, un tri des variables en fonction de leur contribution en ACP en sélectionnant les variables qui sont parmi les 25 % des variables les plus contributives a été expérimenté (pour détails voir matériels & méthodes de l'article). La deuxième approche mise en œuvre consiste en une sélection d'un nombre fini de variables (nombre arbitraire à $1/10^{\text{ème}}$ du dataset utilisé) pour chaque dimension de travail lors de la réalisation d'une ACP ou PLS-DA par la méthode dite *sparse* ACP/PLS-DA de pénalisation *Lasso* des variables [Lê Cao, González, et al. 2009b] (package mixOmics). La dernière technique de sélection de variables est spécifique à l'utilisation des forêts aléatoires et se base sur le critère hiérarchisé de l'importance des variables pour identifier un nombre restreint de variables concentrant la majorité de l'information. Il a été ainsi choisi de sélectionner les 100 variables les plus importantes pour chaque simulation, et de ne retenir que les variables communes à 70 % sur ces 1000 simulations pour définir les variables discriminantes.

I.21 Analyse différentielle

Dans le cas de l'analyse du protéome, la significativité des variations observées entre deux conditions a été évaluée à partir d'un test de Student modifié pour intégrer la méthode empirique Bayes [Smyth 2004]. Ce test estime pour chaque ratio (ou contrastes) la probabilité (p_{value}) qu'il soit égal à 1. Dans le cadre de l'analyse en lait, les valeurs des protéines inférieures au quantile 0,1 entre les deux conditions testées ont été supprimées pour s'affranchir des variations estimées appartenant au bruit de fond. En effet, la quantification relative est effectuée sur les ions parents en MS. Dans la mesure où il y a toujours un bruit de fond en MS, il n'y a jamais de valeur 0 et une aire est mesurée même si ce peptide/protéine n'a pas été identifié à partir des spectres en MS-MS. Un critère de sélection sur le ratio d'expression ou *Foldchange*, *FC*, a été appliqué ($> 1,5$ ou $< -1,5$), suivi de l'ajustement des p-values selon la méthode Benjamini & Hochberg [Benjamini & Hochberg 1995]. La limite de significativité a été imposée à FDR 10 %.

I.22 Enrichissement fonctionnel

Afin de déterminer si des catégories fonctionnelles ont été enrichies au sein des protéines identifiées comme significativement différentielles entre deux conditions, un test d'enrichissement a été mené à partir de la classification fonctionnelle des gènes mise au point par Bolotin sur la souche *L. lactis* ssp.

lactis IL1403 [Bolotin et al. 2001] comme décrit dans [Picard et al. 2012]. Brièvement, pour un jeu de n protéines extraites d'un groupe N de protéines totales, K est le nombre de protéines ayant la caractéristique d'intérêt. Il est alors considéré que k , le nombre de protéines qui ont la caractéristique d'intérêt dans le sous-groupe n , suit une distribution hypergéométrique telle que :

$$k \sim H(K-1, (N-K), n) \text{ et la p-value d'enrichissement } p_{value} = P(N > n) \quad \text{Equation 7}$$

Analyse des surfaces de réponses par la méthodologie des plans d'expériences

Les plans d'expériences constituent une stratégie de planification d'expériences afin d'obtenir des conclusions solides et adéquates de manière efficace et économique (Norme ISO 3534-3). Alors que les plans de criblage identifient les éléments ayant une influence significative sur une réponse parmi une liste de facteurs, l'étude des surfaces de réponse vise, au travers d'un modèle polynomial (le plus souvent) à connaître les variations d'une réponse dans un domaine expérimental, voire à déterminer un optimum de fonctionnement. Les différentes étapes de l'étude d'un plan de surface de réponse sont introduites Figure 21.

ETAPE 1 : FORMALISER LE PROBLÈME

- Définir les objectifs et les réponses à analyser
- Choix des facteurs et du domaine expérimental

ETAPE 2 : CONSTRUCTION DU PLAN

- Définition du modèle empirique
- Construction de la matrice d'expériences (répétitions, points au centre)

ETAPE 3: EXPÉRIMENTATION

- Respect du plan, randomisation des essais

ETAPE 4: ANALYSE DES RÉSULTATS

- Analyse mathématique des résultats d'essais
- Analyse statistique du modèle
- Recherche d'optimum
- Analyse graphique du modèle

ETAPE 5: VALIDATION DU MODÈLE

- Réalisation d'essais de confirmation
- Analyse du modèle

Figure 21: Diagramme méthodologique des différentes étapes d'un plan de surface de réponses.

I.23 Formalisation du problème

Brièvement, l'objectif de ce plan de surface de réponses est de décrire la signature de la souche ND5F en lait en fonction de variations de conditions environnementales. 22 réponses ont été sélectionnées et quatre facteurs ont été choisis : température, composition du lait (Matière sèche et

Matière grasse), et état initial d'oxygénation du lait. L'espace expérimental comprend, pour chaque facteur, trois niveaux définis en valeur codée (-1, 0, +1) pour s'affranchir des valeurs des facteurs dans la suite de l'analyse.

I.24 Construction du plan

La matrice d'expériences désignée s'appuie sur un modèle polynomial du second ordre défini pour chaque réponse Y à partir des facteurs :

$$MS \times O_{ij}$$

Equation 8

Le plan d'expériences devra donc être constitué d'au minimum p expériences pour identifier les p coefficients, ici $p=15$. La matrice d'expérience a été construite sur la base d'un plan de Box-Behnken (qui comprend de base 27 expériences) adapté pour tenir compte des impératifs expérimentaux dont notamment une limite de production de quatre types de laits différents au maximum par semaine. Afin de prendre en compte également les variations liées aux différences de batch de production de lait et d'estimer au mieux la variabilité biologique et technique tout au long de cette période expérimentale, chaque journée d'expérimentation a compris la répétition d'une seule et même expérience qui va permettre lors de l'analyse d'estimer l'erreur expérimentale. Cette expérience correspond au point au centre, fermentation pour laquelle les 4 facteurs sont affectés de la valeur au centre de leur domaine d'étude (0, 0, 0, 0). Ainsi cette expérience a été répétée huit fois pour les journées de fermentations définies. Pour finir, l'ordre des expériences (32 au total) a été « randomisé » par l'intermédiaire d'un tirage au sort afin de rajouter une part d'aléatoire au sein du planning contraint des productions de laits (voir Tableau 12).

I.25 Analyse des résultats

L'analyse des résultats a été menée en utilisant conjointement le logiciel gratuit R (packages *rsm* [Lenth 2013] et *desirability* [Kuhn 2012]) et le logiciel Minitab, version 16.

I.25.1 Analyse mathématique

L'analyse mathématique consiste à estimer, grâce à la méthode des moindres carrés, les p coefficients du modèle et les N résidus, à savoir les écarts entre les valeurs observées et les valeurs prévues par le modèle pour chacun des N expériences du plan. La méthode des moindres carrés garantit une estimation non-biaisée des coefficients du modèle. Un plan d'expériences peut se retranscrire sous la forme :

$\{Y\} = \{X\}\{Coefficients\} + \{E\}$ Avec Y le vecteur des résultats d'essais, X la matrice du modèle, (Coefficients) le vecteur des estimations des coefficients et E la matrice d'erreur. Equation 9

Tableau 12: Organisation du plan de Box-Behnken randomisé en valeurs codées pour les 4 facteurs et 3 niveaux sélectionnés et adaptés pour l'étude. Les expériences en gras correspondent au point au centre.

N° Essais	Matière sèche	Matière Grasse	Valeur de la PO ₂ à l'inoculation	Température
1	0	0	0	0
2	0	0	-1	-1
3	-1	0	0	1
4	1	1	0	0
5	0	0	1	1
6	-1	0	1	0
7	1	-1	0	0
8	0	0	0	0
9	0	0	0	0
10	1	0	0	-1
11	0	-1	0	-1
12	0	1	0	-1
13	0	0	0	0
14	0	-1	0	1
15	0	1	1	0
16	1	0	1	0
17	0	0	0	0
18	0	0	-1	1
19	-1	0	0	-1
20	-1	-1	0	0
21	0	0	0	0
22	-1	0	-1	0
23	-1	1	0	0
24	0	0	1	-1
25	0	0	0	0
26	1	0	-1	0
27	0	-1	1	0
28	0	1	-1	0
29	0	0	0	0
30	0	-1	-1	0
31	0	1	0	1
32	1	0	0	1

La matrice (X) n'étant pas souvent une matrice carrée, pour résoudre ce problème il est fait appel à l'écriture matricielle de la méthode des moindres carrés [Khuri & Cornell 1996; Goupy 2001; Dodge & Rousson 2004] dont l'équation est rappelée ci-dessous :

$$\{\text{Coefficients}\} = ({}^tXX)^{-1}({}^tX)\{Y\} \text{ dont } c^{ij} \text{ sont les termes diagonaux de la matrice de dispersion } ({}^tXX)^{-1} \text{ que la résolution des moindres carrés cherche à minimiser}$$

Equation 10

Connaissant une estimation des coefficients du modèle, il est possible de prévoir la réponse $\hat{Y}(x_n)$ du modèle pour chacun des traitements du plan d'expérience en utilisant le produit scalaire entre le

vecteur des estimations des coefficients et le vecteur représentant la fonction du modèle, pour une combinaison donnée, x_0 , des modalités des facteurs.

Ainsi, pour une combinaison donnée du plan d'expériences, le résidu e_i est défini à partir de la relation suivante :

$$e_i = y(x_0) - \hat{y}(x_0) \quad \text{Equation 11}$$

et l'estimation de l'erreur expérimentale est donnée par $\sigma^2 = \text{var}(e_i)$ Equation 12

La méthode des moindres carrés utilisée pour calculer les coefficients des modèles repose en partie sur l'hypothèse de normalité des résidus, il sera donc nécessaire lors de l'analyse statistique des résultats de vérifier cette normalité.

I.25.2 Analyse statistique du modèle

L'équation du modèle empirique n'est qu'une approximation de la réalité. L'estimation des coefficients du modèle polynomial du second degré s'appuie sur des résultats d'essais qui sont, pour chacun des traitements du plan d'expérience, des valeurs particulières d'une variable aléatoire. La mise en œuvre de tests statistiques doit permettre de porter un jugement sur les résultats obtenus sur : i) le modèle décrivant la variation de la réponse dans le domaine expérimental (qualité descriptive et prédictive), ii) les estimations des coefficients associés aux différents monômes du modèle et iii) les résidus traduisant les écarts entre les valeurs mesurées et les valeurs calculées.

- La première étape de l'analyse statistique aboutit à la construction d'un tableau d'analyse de variance (ANOVA) et à la détermination de la qualité descriptive du modèle. L'analyse ANOVA vise à expliquer la variation totale de la réponse à partir de la somme des carrés des écarts, SCT, entre les résultats d'essais, y_i et leur moyenne $\bar{y}$:

$$SCT = \sum_{i=1}^N (y_i - \bar{y})^2 \quad \text{Equation 13}$$

Cette quantité peut-être décomposée en deux termes, la somme des carrés associée au modèle, SCM, et SCE, la somme des carrés des résidus :

$$SCM = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \quad \text{et} \quad SCE = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad \text{Equation 14}$$

L'analyse se poursuit par la construction d'un test statistique visant à affecter une probabilité à l'hypothèse nulle (H_0) : « le modèle ne permet pas de décrire la variance des résultats ». La statistique de test, F_{obs} exprime alors le rapport des deux variances :

$$F_{obs} = \frac{\frac{SCM}{p-1}}{\frac{SCE}{N-p}}, \text{ où } (p-1) \text{ et } (N-1) \text{ représentent les degrés de liberté affectés à SCM et SCE} \quad \text{Equation 15}$$

Les valeurs de F_{obs} suivent une fonction de répartition théorique appelée fonction de F (ou Snedecor), qui permet de déterminer à partir de quelle valeur, appelée seuil critique, le numérateur est significativement supérieur (donc le modèle significatif). En comparant la probabilité de dépasser ce seuil critique au seuil de significativité α , il est possible de conclure quant à l'acceptation (modèle non significatif) ou au rejet (modèle considéré comme significatif au seuil fixé) de H_0 . L'intégralité de cette analyse sera présentée sous forme d'un tableau d'analyse de variance :

Tableau 13: Tableau d'analyse de variance du modèle (ANOVA)

Source	Somme des carrés	Degrés de liberté	Carré moyen	F_{obs}	Probabilité
Modèle	SCM	p-1	SCM/(p-1)	F_{obs}	$P < \alpha ?$
Résidus	SCE	N-p	SCE/(N-p)		
Total	SCT	N			

La qualité descriptive d'un modèle est obtenue en calculant le coefficient de détermination R^2 ajusté qui tient compte d'une pénalisation en fonction du nombre de coefficients utilisés pour établir le modèle. Calculé à partir des deux équations suivantes, il permet de comparer la qualité descriptive de deux modèles qui n'ont pas le même nombre de coefficients :

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad \text{Equation 16}$$

$$R^2_{adj} = 1 - \frac{N-1}{N-p} (1 - R^2) = 1 - \frac{\frac{SCE}{N-p}}{\frac{SCT}{N-1}} \quad \text{Equation 17}$$

La qualité prédictive du modèle est estimée à partir du coefficient de corrélation prédictif Q^2 grâce à la formule suivante en tenant compte de l'erreur de prédiction ou PRESS:

$$Q^2 = 1 - \frac{PRESS}{SCT} \quad \text{Equation 18}$$

avec $PRESS = \sum_{i=1}^N \frac{y_i - \hat{y}_i}{1 - h_{ii}}$ où h_{ii} est le terme diagonal de la matrice $H = X(X^T X)^{-1} X^T$ Equation 19

- Deuxième étape, l'estimation de la significativité des p coefficients du modèle est également basée sur une hypothèse nulle H_0 , «le coefficient du modèle, a_i (constante, facteur, terme quadratique ou interaction) est nul». La probabilité associée à cette hypothèse est obtenue à partir du calcul d'une statistique de test t_{obs} :

$$t_{obs} = \frac{|a_i|}{s(a_i)} \text{ où } s(a_i) \text{ représente l'incertitude associée à la détermination de } a_i \text{ et traduit l'erreur-type ou écart-type de l'estimateur.}$$

Equation 20

Ainsi, l'erreur-type est définie comme la variance du coefficient, obtenue par la racine carrée du produit de la variance expérimentale, σ^2 , et du coefficient de variance c^{ij} .

$$t_{obs} = \frac{|a_i|}{s(a_i)} = \frac{|a_i|}{\sqrt{c^{ij} \times \sigma^2}}$$

Ainsi, la statistique t_{obs} est égale à :

Equation 21

Les valeurs de cette statistique suivent une fonction de répartition de la loi de Student, utilisée pour calculer à partir de quelle valeur critique, le numérateur de t_{obs} est significativement différent de 0. En fonction de la comparaison au seuil de significativité α choisi, le coefficient est jugé significatif et conservé, ou non significatif et éliminé. Dans les modèles présentés, seuls les coefficients étant significatifs ont été conservés. Pour chaque modèle simplifié établi par régression multiple, l'analyse de variance ANOVA ainsi que les estimateurs de la qualité de description et prédiction du modèle ont été calculés.

- Dernière étape, l'analyse statistique des résidus vient compléter l'analyse statistique des coefficients. L'objectif est de vérifier si, localement, les résidus ne sont pas anormalement importants. Pour ce faire, la normalité des résidus est testée par la droite de Henry et le test de Shapiro-Wilk. De plus, l'observation de la variance des résidus est réalisée afin de s'assurer de l'homoscédasticité de ces derniers et de l'absence de motifs particuliers.

I.25.3 Analyse graphique du résultat

Pour représenter les résultats, une surface de réponse peut être utilisée pour matérialiser la surface de régression à partir d'un graphique dans un espace à trois dimensions. Au-delà de 2 facteurs, les autres facteurs sont maintenus à un niveau constant (choisi comme le point au centre du domaine). Cette première analyse visuelle combinée à l'analyse de graphique, permet d'identifier si des optimums (maximum, minimum, ou point de selle ou dit point de fonctionnement) sont présents.

I.25.4 Recherche d'optimum-fonction de désirabilité

A côté d'une estimation graphique, la recherche d'une combinaison de valeurs pour les facteurs qui optimisent la réponse a été réalisée par résolution mathématique classique (résolution de la dérivée de l'équation du modèle égale à 0) ainsi qu'en transformant la réponse à partir des fonctions de désirabilité. Proposé par Harrington [Harrington 1965] pour la première fois, le formalisme utilisé ici a été celui décrit par Derringer et Suich [Derringer 1980] ; pour revue, [Costa et al. 2011]. L'introduction d'une fonction de désirabilité (transformation des valeurs de réponses entre 0 et 1) présente l'avantage de pouvoir réaliser une optimisation conjointe d'un ensemble de réponses en satisfaisant les conditions requises pour chaque réponse. Pour réaliser cette optimisation, il faut : i) obtenir la désirabilité individuelle (d) de chaque réponse, ii) combiner les désirabilités individuelles afin d'obtenir la désirabilité combinée ou composite (D), puis iii) maximiser la désirabilité composite et identifier les paramètres optimaux des facteurs.

Pour établir les fonctions individuelles de désirabilité d à partir de l'équation de réponse $f(x)$, un objectif est fixé (minimiser, maximiser ou atteindre une cible) et des limites (A et B) du domaine de réponses sont instaurées. Proche de l'objectif, les valeurs de d se rapprocheront de 1, alors qu'elles seront nulles loin de ce dernier. Mathématiquement, ces choix se traduisent par le formalisme suivant des fonctions :

$$d_r^{max} = \begin{cases} 0 & \text{si } f(x) < A \\ \left(\frac{f(x) - A}{B - A} \right)^s & \text{si } A \leq f(x) \leq B \\ 1 & \text{si } f(x) > B \end{cases} \quad \text{et} \quad d_r^{min} = \begin{cases} 0 & \text{si } f(x) > B \\ \left(\frac{f(x) - B}{A - B} \right)^s & \text{si } A \leq f(x) \leq B \\ 1 & \text{si } f(x) < A \end{cases}$$

$$d_r^{cible} = \begin{cases} \left(\frac{f(x) - A}{t_0 - A} \right)^{s1} & \text{si } A \leq f(x) \leq t_0 \\ \left(\frac{f(x) - B}{t_0 - B} \right)^{s2} & \text{si } t_0 \leq f(x) \leq B \\ 0 & \text{dans tout les autres cas} \end{cases}$$

Les valeurs de s , $s1$, $s2$ sont choisies par l'utilisateur pour rendre plus facile ou difficile le critère de désirabilité

Equation 22

La forme de la fonction de désirabilité entre la limite supérieure et la cible est déterminée par le choix de la pondération. Une fois la désirabilité individuelle de chaque réponse calculée, la désirabilité composite (D) est obtenue en réalisant la moyenne arithmétique pondérée des désirabilités individuelles des différentes réponses. Ici, aucune pondération n'a été prise en compte. Seules les réponses dont les désirabilités individuelles obtenues étaient supérieures à 75 % ont été retenues pour l'optimisation multi-réponses. Un algorithme de gradients ayant plusieurs points de départ et maximisant la désirabilité composite est ensuite utilisé afin de déterminer la solution

optimale numérique (paramètres optimaux des variables d'entrée). Uniquement les solutions présentant des valeurs pour les facteurs incluses dans le domaine expérimental initial ont été retenues afin d'augmenter la qualité prédictive des réponses. La solution est présentée sous la forme d'un diagramme d'optimisation, reprenant la désirabilité globale ainsi que la désirabilité individuelle de chaque réponse avec la prédiction de la valeur.

I.26 Validation du modèle

La validation du modèle, et en particulier de l'optimum de fonctionnement proposé, a été effectuée à partir de l'analyse de la variance entre la réponse prédite par le modèle optimisé et les valeurs des réponses expérimentales obtenues sur 3 réplicats biologiques lors du transfert à l'échelle industrielle. Dans ces répétitions, seules neuf réponses ont été enregistrées et suivies. De plus, 2 nouveaux points expérimentaux dans le domaine expérimental ont été réalisés afin de vérifier les qualités prédictives du modèle.

RESULTATS

CHAPITRE I :

ANALYSE FONCTIONNELLE DU COMPORTEMENT MACRO-CINÉTIQUE DE LA SOUCHE *L. LACTIS* SSP. *LACTIS ND5F* EN MATRICE LAIT.

INTRODUCTION

Décrire, analyser et caractériser le comportement de *ND5F* en condition industrielle est indispensable pour identifier l'origine de la singularité perçue lors de la dégustation des produits fermentés par cette souche. La démarche retenue porte, dans un premier temps, sur une analyse macro-cinétique complète (biomasse, acidification, réduction, consommation d'oxygène et principales activités métaboliques) de *L. lactis ssp. lactis ND5F* (version *Lab*) dans son environnement applicatif. L'objectif est ainsi de constituer une liste des principaux descripteurs phénotypiques et métaboliques de l'activité de *ND5F*, afin de la positionner ultérieurement face à d'autres souches. Ce premier chapitre de résultats vise également à acquérir une meilleure compréhension du comportement de *ND5F* en lait, et à déterminer les paramètres environnementaux qui peuvent influencer cette réponse phénotypique.

Ainsi, ce travail a débuté par un transfert du procédé à l'échelle laboratoire et par une évaluation de la reproductibilité des conditions industrielles. Il s'est poursuivi par un diagnostic fonctionnel de *ND5F* en milieu lait. Pour simplifier l'analyse, le suivi des différentes activités a été restreint aux étapes de culture jusqu'au conditionnement du lait fermenté à 4°C. Trois modifications de l'environnement ont été ensuite ciblées pour étudier leurs mécanismes et conséquences sur la fonctionnalité du ferment : l'auto acidification, le stress oxydatif et la thermisation du milieu. Enfin, le dernier volet de cette caractérisation a été consacré à l'analyse des exigences nutritionnelles de la souche.

I Diagnostic du comportement de ND5F dans la condition de référence

L'enjeu de cette section est de présenter une caractérisation macro-cinétique multicritères de la souche *L. lactis* ssp. *lactis* ND5F dans un contexte d'utilisation industrielle. Ce diagnostic fonctionnel contraint au transfert du procédé de production à l'échelle laboratoire. Ainsi, il a été nécessaire de définir au laboratoire une condition dite de référence reproduisant au plus près un procédé industriel de mise en œuvre de ND5F. Après examen des différents procédés industriels employés, il a été choisi de reproduire des cultures en mode batch à pH libre de *Lactococcus lactis* ssp. *lactis* ND5F en condition micro-aérobie sur du lait écrémé (facilité de production en condition standard au laboratoire). Les paramètres environnementaux sélectionnés sont listés dans le Tableau I. 1. Il est à noter que la nécessité de réaliser des prélèvements homogènes pour le suivi des cultures même lors de la formation du gel lactique implique l'utilisation d'une agitation (100 rpm). Les conséquences de cette agitation ainsi que des autres sources de divergences, en lien avec le transfert d'échelle, seront examinées au travers de l'étude comparative des profils macro-cinétiques de la souche.

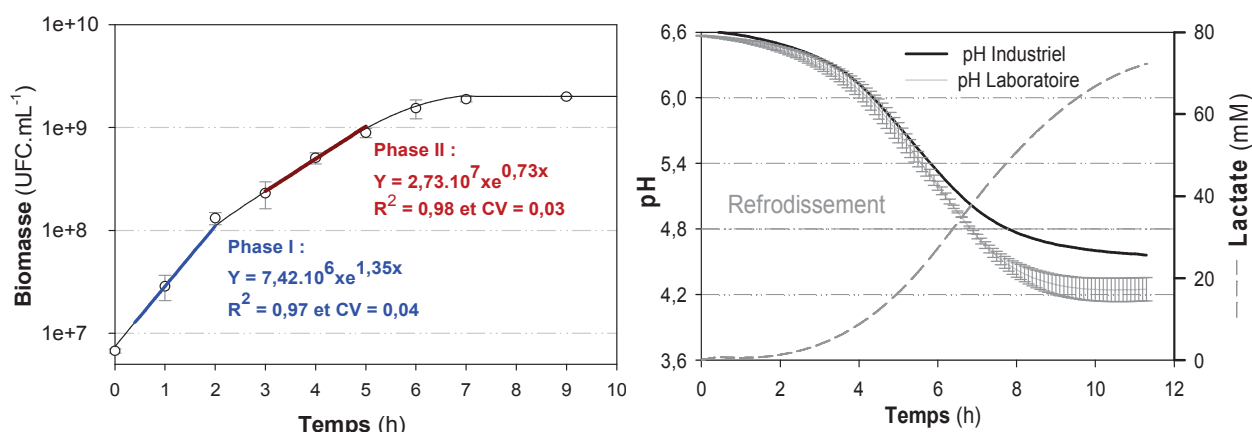
Tableau I. 1: Conditions de culture pour la culture discontinue de la souche *L. lactis* ND5F au laboratoire.

Milieu	Agitation	Température	Aération	pH
Lait écrémé	100 rpm	28 °C	Aucune aération	pH libre

I.1 Croissance diphasique et caractérisation de l'activité acidifiante

Ces deux paramètres constituent les principaux critères microbiologiques traditionnellement utilisés pour caractériser la fonctionnalité d'une souche-ferment en matrice lait. En effet, le suivi de la croissance renseigne sur l'aptitude de la souche à se développer et à atteindre une concentration cible permettant l'expression d'un profil technologique. La baisse du pH exocellulaire en lait, conséquence directe de la production d'acide (principalement d'acide lactique), est l'une des principales fonctions de l'activité cellulaire. De plus, sa cinétique conditionne la coagulation du lait.

Le profil moyen de croissance de la souche *L. lactis* ND5F dans la condition de référence est caractérisé par une croissance exponentielle diphasique, avant l'entrée en phase stationnaire à $7 \pm 0,2$ h de culture (Figure I. 1A). Après inoculation, la croissance débute sans phase de latence témoignant d'une adéquation de la physiologie cellulaire entre les conditions de pré culture et le milieu de culture (toutes deux en lait écrémé).



B

Figure I. 1: Cinétique de croissance et d'acidification à pH libre de *L. lactis* ND5F en lait écrémé dans la condition de référence au laboratoire. A) Profil moyen de la concentration cellulaire (UFC.mL⁻¹) obtenu par le lissage de la moyenne des points expérimentaux de 4 réplicats biologiques. Evaluation de la reproductibilité des données par la représentation sur chaque dénombrement moyen (○) de l'écart-type expérimental. Chaque phase exponentielle de croissance est modélisée par régression linéaire (Equation, coefficient de détermination R^2 et coefficient de variation, CV). B) Comparatif de la cinétique de pH entre le procédé industriel et au laboratoire et cinétique moyenne de production de lactate (lactate + acide lactique sous forme non dissociée) en mM. Ecart-types établis à partir des données de 4 réplicats biologiques pour les données du laboratoire.

La première phase exponentielle de croissance s'étend du temps initial jusqu'à 2,5 heures de culture, et la seconde phase exponentielle s'étend de 3 heures à 5 h de culture. Ces deux phases de croissance (modélisables par une droite de régression) sont définies par un taux de croissance maximal de $1,35 \pm 0,04 \text{ h}^{-1}$ et de $0,73 \pm 0,03 \text{ h}^{-1}$ respectivement (Figure I. 1A). L'absence de données de suivi de la croissance à l'échelle industrielle ne permet pas de comparer la reproductibilité des profils obtenus avec l'utilisation en procédé.

En parallèle, l'acidification du lait, résultat de la production de lactate, a lieu principalement pendant la phase de croissance avec une baisse de $1,84 \pm 0,02$ unité pH du lait (Figure I. 1.B). Le pH à l'arrêt de croissance est de $4,76 \pm 0,04$ upH pour la condition laboratoire, avec une production associée de lactate total de 38 ± 2 mM. Ce taux correspond à une concentration de la forme non dissociée, c'est-à-dire d'acide lactique, de $4,02 \pm 0,5$ mM. Le pH est de 4,9 dans le procédé industriel lors de l'entrée en phase stationnaire (Figure I. 1.B). Ce profil diphasique, ainsi que le début de coagulation du lait observé, confirment phénotypiquement la présence d'un système protéolytique actif chez ND5F et notamment de la protéase de paroi, PrtP. En effet, les souches PrtP⁻ incapables de dégrader la caséine ont un profil monophasique correspondant à l'utilisation par la cellule des peptides endogènes du lait (et dans une moindre mesure des acides aminés libres). Cette consommation ne permet ni de soutenir la croissance au-delà d'une densité cellulaire de 10^8 UFC.mL^{-1} [Juillard et al. 1998], ni de produire suffisamment d'acide lactique pour conduire à une acidification rapide du lait.

Dans le cas de *ND5F*, la 1^{ère} phase de croissance s'achève à $1,05 \cdot 10^8$ UFC.mL⁻¹. La seconde phase de croissance exponentielle est caractéristique d'une nutrition azotée reposant uniquement sur l'hydrolyse de la caséine et sur l'intégration des oligopeptides par les transporteurs transmembranaires. Les taux de croissance de ces deux phases correspondent aux ordres de grandeur des données bibliographiques des souches PrtP⁺, avec une 1^{ère} phase avec une vitesse de croissance supérieure (meilleure efficacité nutritionnelle d'utilisation des oligopeptides du lait) à la 2nd [Juillard et al. 1995]. Après 7 h de culture, la biomasse atteint $2,1 \cdot 10^9$ UFC.mL⁻¹ ou $1,27 \pm 0,05$ g.L⁻¹, ce qui concorde avec les résultats de dénombrements à l'arrêt de croissance sur le procédé industriel, à $2 \cdot 10^9$ UFC.mL⁻¹.

La phase de croissance précède une phase de post-acidification particulièrement intense dans la reproduction du procédé au laboratoire (baisse de 0,56 unités pH et production de 30 mM supplémentaire d'acide lactique) par comparaison au procédé industriel (uniquement de 0,16 unités pH). Cette différence s'explique notamment par des variations de la quantité finale d'acide lactique produit (10 mM de différence à 10 h, données non présentées). L'absence d'un cycle de refroidissement lorsque la culture atteint le pH de 4,80 dans le procédé laboratoire est probablement à l'origine de ces divergences de comportement de post-acidification. En effet, le refroidissement brusque bloque les fonctions enzymatiques de la souche et permet ainsi de stabiliser autour de 4,60 le pH final du lait fermenté dans le procédé industriel.

La comparaison des cinétiques de croissance (Figure I. 1A) et d'acidification (Figure I. 1B) révèle le décalage temporel entre les arrêts des fonctions anaboliques (à 7 h) et cataboliques (à 10 h). Cet intervalle sous-entend que des divergences mécanistiques sont responsables de ces arrêts. Trois hypothèses peuvent expliquer l'arrêt de croissance observé, soit i) une carence nutritionnelle, soit ii) une toxicité du métabolite majoritaire, le lactate, soit un iii) un effet conjoint de la carence nutritionnelle et de l'inhibition par le lactate. Pour répondre à cette interrogation, des cultures sur milieux riche M17, semi-synthétique YE et synthétique MCD à 50 g.L⁻¹ de lactose et à pH libre ont été réalisées, prouvant que *ND5F* pouvait croître à pH 4,70 jusqu'à des concentrations exocellulaires en acide lactique total de 200 mM et de 6 mM concernant spécifiquement la forme non dissociée d'acide lactique, principal effecteur de l'effet inhibiteur (résultats non présentés). Puisque seulement 38 mM d'acide lactique total et 4,02 mM d'acide lactique non dissocié ont été dosés à l'arrêt de croissance sur lait et à pH équivalent, l'hypothèse d'un arrêt de croissance du fait exclusif d'une toxicité en acide lactique peut-être vraisemblablement rejetée. Cependant, sans analyse de la sensibilité de *ND5F* à la concentration en acide lactique, une participation de ce dernier au ralentissement de la croissance ne peut être exclue (analyse conduite paragraphe I.3 de ce chapitre).

Ainsi, l'arrêt des fonctions anaboliques impliquerait également une carence nutritionnelle qui proviendrait i) du lactose, principale source de carbone du lait, ii) des acides aminés, sous forme libre ou non, ou éventuellement iii) de facteurs de croissance (vitamines, bases nucléotidiques). Ces réponses seront apportées par l'analyse conjointe du suivi des concentrations de ces substrats dans le lait (présentée ci-après) et des exigences nutritionnelles de ND5F (analyse conduite paragraphe 0 de ce chapitre). En ce qui concerne l'arrêt de la phase de post-acidification, un arrêt des fonctions cataboliques lié à une acidification intracellulaire trop importante est avancé. Cette hypothèse s'appuie d'une part sur la mesure du pH exocellulaire, à 4,20 à l'arrêt du catabolisme, et d'autre part sur les données bibliographiques d'autres souches de *Lactococcus lactis* démontrant qu'à ce pH, le pH intracellulaire (pH_i) ne peut être maintenu au-dessus de 4,9 [Rallu et al. 1996]. De plus, à ce pH_i, l'activité de certaines enzymes (GLK et LDH) se trouve fortement inhibée [Raynaud et al. 2005]. Cependant, en l'absence d'un dosage de ces activités enzymatiques intracellulaire cette hypothèse, ne peut être ni confirmée ni infirmée.

Les profils d'acidification des deux procédés (industriel et laboratoire) apparaissent similaires pendant la majorité de la phase de croissance (Figure I. 1B). Cependant, afin de confirmer cette observation et de caractériser l'activité acidifiante de ND5F, un ensemble de six descripteurs vont être comparés sur les deux types de fermentations réalisées. Ainsi, les résultats suivants ont été comparés :

- vitesse maximale d'acidification, $V_{\max}^a$, en unité pH par minute,
- pH et temps de culture à $V_{\max}^a$, $\text{pH}_{\max}^a$ et $T_{\max}^a$,
- pH du milieu à 4h, $\text{pH}_{4\text{H}}$
- temps de culture auquel le pH du milieu atteint 4,6, $T_{\text{pH}4,6}$
- vitesse moyenne d'acidification pendant la phase de post-acidification, V_{post}^a ,

Les résultats présentés Tableau I. 2 permettent d'affirmer qu'il n'y a peu de différence significative entre l'activité acidifiante de la souche dans le procédé de fermentation mis au point au laboratoire et celle enregistrée pendant la phase de croissance à l'échelle industrielle (exception faite de $T_{\text{pH}4,6}$).

Tableau I. 2 : Comparaison des descripteurs de l'activité acidifiante de *L. lactis* ND5F à pH libre sur lait en culture discontinue en conditions industrielles et en procédé à l'échelle du laboratoire. Ecart-types établis à partir des données de 3 réplicats biologiques (données industrielles) et 4 données laboratoires. ND : Non déterminé.

Descripteurs	$V_{\max}^a$ ($10^3 \text{ pH} \cdot \text{min}^{-1}$)	$\text{pH}_{4\text{H}}$	$\text{pH}_{\max}^a$	$T_{\max}^a$ (h)	$T_{\text{pH}4,6}$ (h)	V_{post}^a ($10^3 \text{ pH} \cdot \text{min}^{-1}$)
Industriel	$7,4 \pm 0,1$	$5,96 \pm 0,01$	$5,41 \pm 0,26$	$5,38 \pm 0,56$	$8,13 \pm 0,15$	ND
Laboratoire	$7,5 \pm 0,2$	$5,94 \pm 0,04$	$5,52 \pm 0,16$	$5,30 \pm 0,20$	$7,78 \pm 0,25$	$-1,50 \pm 0,1$

L'absence de données supplémentaires à l'échelle industrielle restreint la comparaison aux données de croissance et de pH déjà présentées. Il est décidé sur la base de la proximité des résultats obtenus (Intervalle de confiance à 99 %) de valider la conduite du procédé mis au point au laboratoire pour analyser le comportement macro-cinétique de ND5F dans les conditions industrielles.

I.1 Cinétiques de consommation d'oxygène et activité réductrice

Afin d'élargir les critères d'étude du comportement de ND5F dans son environnement de mise en œuvre, les cinétiques de la pression partielle en oxygène dissout, PO_2 , et du potentiel d'oxydoréduction, REDOX, ont été étudiées.

Le profil de PO_2 révèle deux environnements gazeux au cours de la culture (Figure I. 2). La première phase en micro-aérobie se déroule du temps initial jusqu'à 1,8 h. Durant cette phase correspondant à la 1^{ère} phase exponentielle de croissance, la PO_2 du milieu décroît de 90 % à 0 %. Il est à noter que la PO_2 initiale n'est pas égale à 100% mais à 90%. Ce biais provient de la calibration du 100% de la sonde effectuée sur de l'eau et non sur du lait qui n'ont pas les mêmes constantes de saturation et de transfert en oxygène. Cet appauvrissement en dioxygène relatif à une consommation d'oxygène traduit une activité de métabolisation de l'oxygène chez *L. lactis* ND5F. Par conséquent, la présence d'au moins un système de réduction de l'oxygène est déduite. La NADH oxydase NoxE a été identifiée comme le principal élément responsable de la consommation de l'oxygène [Tachon et al. 2010], même si le système alkyl hydroperoxyde réductase (composé de la NADH oxydase AhpF et de la peroryredoxine AhpC) peuvent participer également à ce processus. La seconde phase de culture, de 1,5 h jusqu'à la fin de la fermentation, s'effectue avec une pression partielle en oxygène dissout nulle. Par conséquent, la croissance de la souche est considérée en extrême limitation d'oxygène sur la majorité de la culture.

Avant d'étudier l'impact éventuel de ces modifications d'atmosphère gazeuse sur le métabolisme de la souche et les conséquences en aval du pyruvate, l'activité réductrice de la souche va être examinée à partir de la cinétique du potentiel d'oxydoréduction normalisé Eh_7 . En effet, *L. lactis* est connue pour être l'une des bactéries lactiques capables de réduire le plus le potentiel d'oxydoréduction, Eh , du lait [Brasca et al. 2007; Jeanson et al. 2009]. Or, il a été démontré que l' Eh avait un rôle sur la microflore de nombreux produits laitiers, notamment des fromages [Boucher et al. 2006], ou sur l'orientation des voies de catabolisme des acides aminés [Cachon et al. 2002; Jeanson et al. 2009; Kieronczyk et al. 2006; Kersiene et al. 2008].

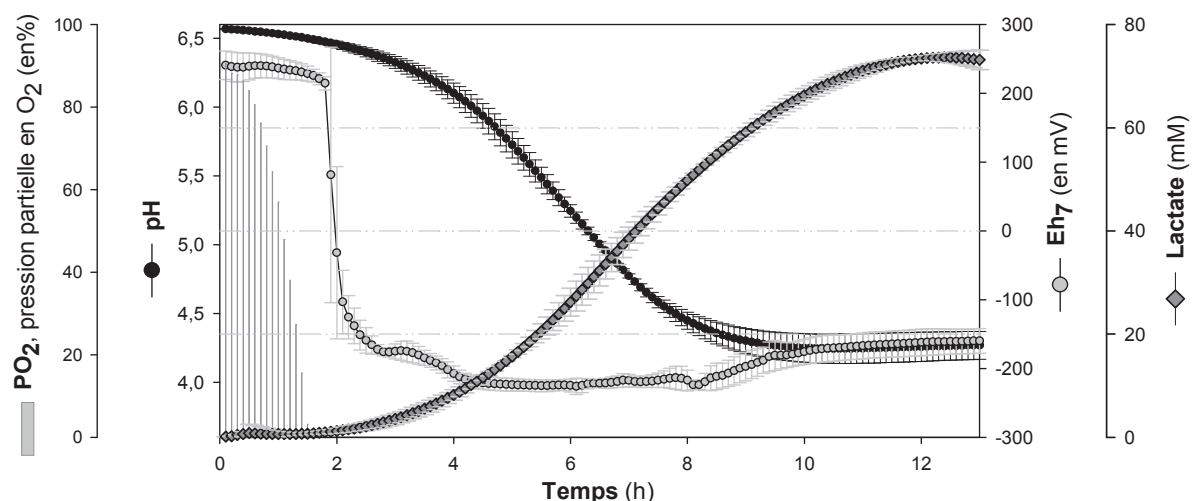


Figure I. 2: Profils moyens de décroissance de la pression partielle en oxygène dissout PO_2 (histogramme), du pH, du potentiel d'oxydoréduction Eh_7 et cinétique de production d'acide lactique pendant la culture discontinue de la souche *L. lactis ND5F* en lait écrémé à pH libre. Ecarts-types établis à partir des données de quatre répliquats biologiques.

Par conséquent, ce paramètre, souvent non considéré, va être intégré à l'analyse macro-cinétique pour caractériser le comportement de *ND5F* dans la condition de référence. Le profil du redox dans le milieu de culture est présenté Figure I. 2, en parallèle des cinétiques de pH et de dioxygène, deux facteurs à prendre en compte pour une compréhension globale du phénomène. Le potentiel Eh_7 du lait avant inoculation est à dominante oxydante (autour de +250 mV), alors qu'il devient fortement réducteur après 24 h de culture de *L. lactis ND5F*, à -160 ± 14 mV. Ainsi, à cette valeur, tous les oxydants dont le potentiel de demi-réduction est supérieur à cette valeur sont présents sous leur forme réduite.

La cinétique peut-être décomposée en 5 phases distinctes pour aider à son interprétation :

- Phase I : de $t = 0$ à $t = 1,8$ h, un début de réduction du lait est observé à la vitesse moyenne de -8 mV.h^{-1} . Simultanément à cette réduction du milieu, l'acidification du milieu débute. Cette observation concorde avec l'étude de Jeanson [Jeanson et al. 2009] qui a prouvé que chez *L. lactis*, les deux activités sont liées lorsque le niveau d'oxygène est initialement haut. L'absence d'une augmentation ponctuelle du Eh_7 pendant cette phase en présence d'oxygène, confirme l'hypothèse de l'absence de production massive d'espèces réactives de l'oxygène. En effet, ces composés plus oxydants que l'oxygène auraient augmenté le Eh_7 . La réduction observée est probablement liée à la diminution conjointe de l'oxygène (principal oxydant du milieu) et au début d'acidification (production d'acide lactique, principal réducteur).

- Phase II : de $t = 1,8$ h à $t = 2,5$ h, la vitesse de réduction du lait atteint un maximum de -1426 mV.h^{-1} et le potentiel Eh_7 décroît jusqu'à -175 mV. Cette brusque réduction de l' Eh_7 confirme la disparition pré-requise de l'oxygène pour réduire le lait. Cependant, un lait dépourvu d'oxygène (par

saturation en N_2) présente un potentiel positif (+50 mV), ce qui démontre que pour atteindre un Eh_7 négatif, ND5F doit réduire des composés oxydants présents dans le lait [Jeanson et al. 2009]. La présence de cations métalliques et de composés soufrés oxydés a été avancée comme substrats des ménaquinones membranaires pour expliquer la capacité réductrice des souches de lactocoques [Tachon et al. 2010].

- Phase III : de $t = 2,5$ h à 4 h, la vitesse de réduction du lait se stabilise à -33 mV.h^{-1} , le potentiel Eh_7 diminue à -225 mV et le pH débute sa seconde phase d'acidification (7 mM de lactate produit). Le potentiel de réduction est supérieur de 15 % par rapport au lait fermenté le plus réduit étudié lors de l'analyse de la diversité de l'activité réductrice de la sous-espèce *lactis* [Brasca et al. 2007]. Elle est cependant identique à celle enregistrée par Tachon pour les souches les plus réductrices au sein de la sous-espèce *cremoris* [Tachon et al. 2010]. Ainsi, il semblerait d'après les données de la littérature que ND5F se positionne comme une souche à l'activité réductrice intense.

- Phase IV : de $t = 4$ h à 8 h, le potentiel d'oxydoréduction reste stable autour de -220 mV et cela malgré l'acidification massive du lait (baisse de 1,70 unité pH) et la production de 54 mM d'acide lactique (voir Figure I. 2).

- Phase V : de $t = 8$ h à 10h, une remontée progressive du potentiel d'oxydoréduction du milieu, Eh_7 , selon une vitesse moyenne de $+12,5 \text{ mV.h}^{-1}$ est constatée. Cette observation traduit l'apparition dans le milieu de molécules oxydantes. A ce pH fortement acide, il est postulé que ces molécules soient des amines soit i) de la voie de l'ADI (ornithine, citrulline), ii) de la voie glutamate décarboxylase (acide- γ -aminobutyrique à partir de glutamate), ou encore ii) d'autres molécules éventuelles (histamine, tyramine...). En effet, ces molécules oxydantes sont synthétisées pour lutter contre l'acidification du milieu intracellulaire et produire de l'énergie. La présence des molécules de la voie de l'ADI et de la glutamate décarboxylase a été confirmée au sein des surnageants de culture par dosage HPLC à des concentrations de 25 mM à 10 h de culture. Aucune amine biogène toxique n'a été détectée.

I.2 Analyse macro-cinétique du métabolisme

L'évolution des concentrations en substrats potentiels de la souche *L. lactis* ND5F (lactose et citrate) ainsi que des métabolites qui découlent du nœud du pyruvate est présentée Figure I. 3. En 12 h de culture, la souche *L. lactis* ND5F ne consomme que 16,2 % du lactose initialement présent dans le lait à $51,2 \text{ g.L}^{-1}$ (soit $8,3 \text{ g.L}^{-1}$), ce qui démontre que l'arrêt de croissance observé n'est pas lié à une carence de lactose. Cette faible proportion de lactose consommé entraîne une sous-estimation lors du dosage de ce substrat estimée à 10 %. Par conséquent les capacités cataboliques de la souche seront plus sûrement évaluées d'après les concentrations de produits, dont le dosage est plus précis.

L'utilisation du citrate dans la cellule étant conditionnée chez les lactocoques par la présence de citrate perméase, CitP, aucune consommation de citrate n'est logiquement observée sur ND5F (voir Figure I. 3). Le lactose est donc la seule entrée dans le métabolisme central carboné. Aucune accumulation de glucose ou de galactose n'est observée dans le milieu exocellulaire (résultats non présentés), ce qui indique que chaque monosaccharide va suivre sa propre voie de dégradation (glycolyse ou voies de Leloir et Tagatose) sans être excrété après clivage intracellulaire du lactose. Les voies de dégradation précédemment citées aboutissent toutes au niveau du pyruvate, véritable intermédiaire clé du métabolisme chez les bactéries lactiques.

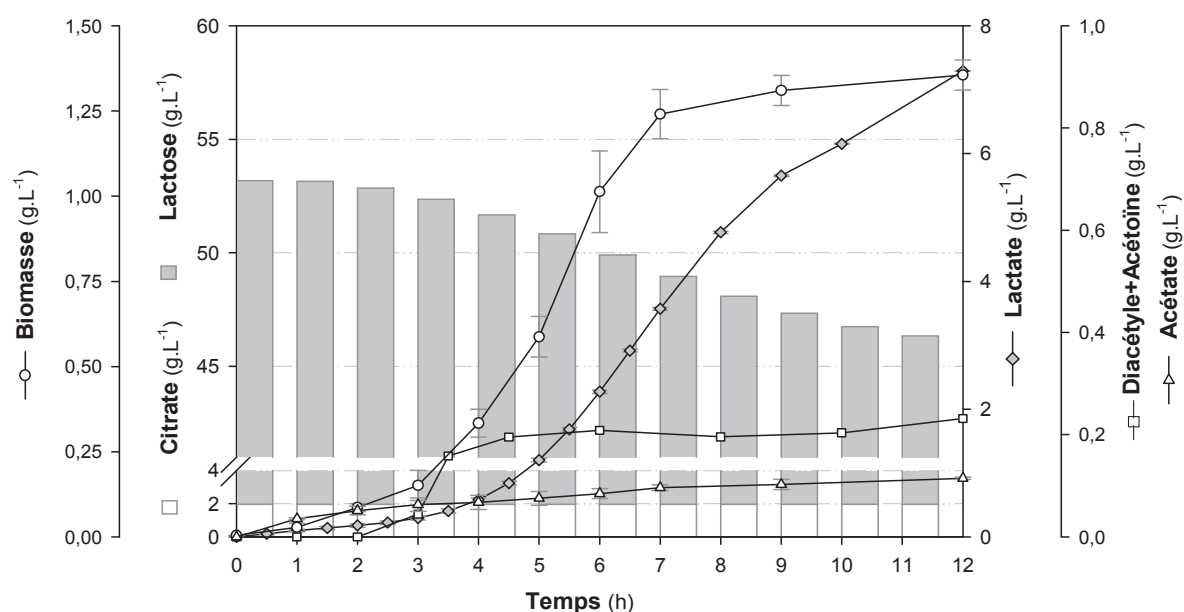


Figure I. 3 : Cinétiques des concentrations cumulées en substrats (lactose et citrate) ainsi que des métabolites produits (lactate, acétate et arômes : diacétyle et acétoïne) et de la biomasse au cours de la culture discontinue de la souche *L. lactis* ND5F en lait écrémé à pH libre. Concentrations exprimées en g.L⁻¹. Ecart-types établis à partir des données de quatre réplicats biologiques.

Le pyruvate produit à partir du lactose est principalement converti par la lactate déshydrogénase (LDH) en acide lactique. En effet, 7,3 g.L⁻¹ d'acide lactique sont produits à partir de la consommation de 8,3 g.L⁻¹ de lactose en 24 h de culture, ce qui représente plus de 84 % du devenir du pyruvate en équivalent carbone. Il est communément admis que le métabolisme de la souche est dit homolactique lorsque 90 % du pyruvate est converti en acide lactique. Par conséquent aux erreurs près de mesures de la concentration en lactose, le métabolisme de la souche *L. lactis* ND5F, en lait, est considéré comme homolactique.

Une présence d'acétate, tout comme des arômes diacétyle et acétoïne, est détectée à partir de 3 h de culture, les concentrations de ces produits étant respectivement de $0,17 \pm 0,02$ et $0,23 \pm 0,05$ g.L⁻¹. Aucune présence de formiate, d' α -acétolactate, ou d'éthanol n'a été détectée dans ces conditions. Au vu du faible taux de diacétyle et d'acétoïne produits, ces composés à l'arôme de beurre ne peuvent être seuls à l'origine de la singularité de la souche.

Afin de vérifier qu'aucun des principaux métabolites produits par la souche n'a été omis, un bilan carbone à partir des produits fermentaires et de la biomasse est réalisé selon l'Equation I. 1.

$$\text{Bilan C} = \frac{[\text{Biomasse}] \times 4,63 + [\text{Lactate}] \times 3 + [\text{Acétate}] \times 2 + ([\text{Diacétyle}] + [\text{Acétoïne}]) \times 4 + [\text{Formiate}] + [\text{CO}_{2,\text{estimé}}]}{[\text{Lactose}] \times 12}$$

$$\text{CO}_{2,\text{estimé}} = [\text{Ethanol}] + [\text{Acétate}] + 2 \times ([\text{Diacétyle}] + [\text{Acétoïne}]),$$

Biomasse de formule élémentaire $C_{4,63} H_{7,89} O_{2,35} N_{1,0}$ et de masse molaire 129 g.mol⁻¹.

Equation I. 1: Bilan carbone à partir des métabolites et de la biomasse dosés (en mM).

Ce bilan permet également d'estimer la production de dioxyde de carbone associée à la synthèse d'acétate ainsi que de diacétyle et d'acétoïne. Le bilan carbone catabolique (calculé uniquement à partir du lactose consommé et des produits du catabolisme, sans la biomasse) est égal à 0,95 pour une concentration en dioxyde de carbone estimée par le calcul de 8,3 mM. Ce bilan passe à 1,1 en prenant en compte la biomasse formée mais sans la consommation des acides aminés difficilement mesurables sur ce milieu complexe (théoriquement il devrait être égal à 1 sans les erreurs de mesures sous-estimant le lactose).

Les vitesses spécifiques de consommation de lactose et de production de lactate ont été calculées afin de compléter par une approche cinétique les processus observés (Figure I. 4). La vitesse maximale de consommation du lactose, v_{lactose} , est atteinte à $t = 4,5$ h de fermentation pendant la 2^{ème} phase exponentielle de croissance, elle est estimée à $4,5 \pm 0,1$ mmol.g⁻¹ de X. h⁻¹. La vitesse spécifique de production du lactate, v_{lactate} , est maximale à $t = 4,5$ h également, avec $18,2 \pm 3$ mmol de lactate produit. g⁻¹ de X en h⁻¹. Le rendement théorique de la voie de dégradation du lactose en métabolisme homolactique est de 4 lactates pour un lactose. Le calcul du rendement effectif à partir des vitesses spécifiques, $Y_{p/s}$ pour la fermentation de ND5F en lait à pH libre est égal à 4,04, caractéristique d'un métabolisme homolactique. Les vitesses spécifiques de production des autres métabolites ne seront pas commentées en raison de leurs faibles valeurs.

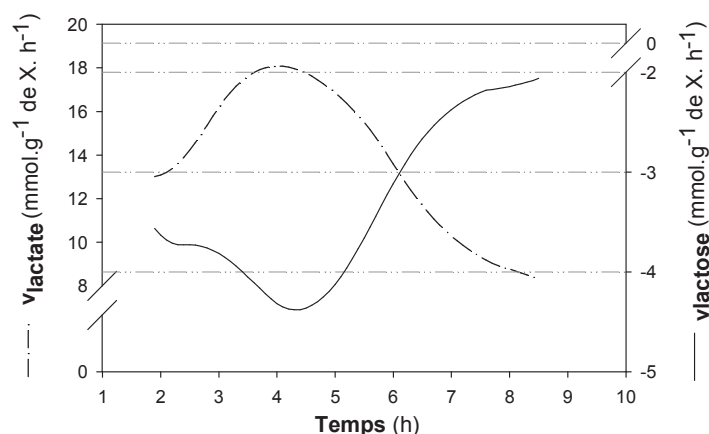


Figure I. 4 : Vitesses spécifiques moyennes de production de lactate, v_{lactate} , et de consommation de lactose, v_{lactose} , au cours de la culture discontinue de la souche *L. lactis* ND5F en lait écrémé à pH libre. Profil moyen à partir de 4 réplicats biologiques.

Les principaux descripteurs caractérisant le comportement macro-cinétique (croissance, acidification, et production de métabolites) de ND5F à pH libre sur lait sont résumés Tableau I. 3. Il sera fait référence à ce bilan, dans son format complet ou non, en fonction des besoins de l'analyse, pour les futurs essais présentés dans ce manuscrit de thèse.

Pour déterminer la composition du lait écrémé en acides aminés libres et éclairer le phénomène de consommation de ces nutriments au cours des différentes phases de croissance, les concentrations dans les surnageants de cultures ont été déterminées (voir Tableau I. 3). Le lait écrémé non inoculé présente un pool d'acides aminés libres constitué essentiellement d'acide glutamique (Glu), de sérine (Ser), de glycine (Gly), de thréonine (Thr), d'arginine (Arg), de valine (Val), de lysine (Lys) et de proline (Pro), aux concentrations équivalentes aux données bibliographiques [Juillard et al. 1995]. Cependant, il contient moins de 1 mg.L⁻¹ d'isoleucine (Ile), de leucine (Leu), de méthionine (Met), de tyrosine (Tyr), et de glutamine (Gln). Les souches de lactocoques étant généralement auxotrophes pour la majorité de ces acides aminés [Jensen 1993; Coccagn-Bousquet et al. 1995], la croissance de la souche *L. lactis* ND5F n'est donc possible que grâce aux acides aminés apportés lors de l'assimilation en parallèle des peptides initialement présents dans le lait ou issus de l'hydrolyse de la caséine (non dosés ici). Les concentrations en acides aminés présentées sont le résultat de l'équilibre existant entre leurs vitesses de production (hydrolyse de la caséine et efflux éventuel d'acides aminés par la cellule) et leurs vitesses de consommation. Ainsi, pour pouvoir être dosés, les acides aminés doivent soit s'accumuler dans le lait fermenté, ce qui traduit une vitesse de production supérieure à leur vitesse de consommation, soit être faiblement consommés et maintenir une certaine concentration résiduelle s'ils sont présents initialement dans le lait.

Tableau I. 3: Principaux descripteurs caractérisant le comportement macro-cinétique de croissance, d'acidification, et de production de métabolites de ND5F à pH libre en lait écrémé. Segmentation de la culture en trois phases : croissance, arrêt de croissance et à l'arrêt de post-acidification à 12 h. Les résultats sont présentés avec leurs écarts-types issus de l'intégration de 4 réplicats biologiques à l'exception (*) des taux de croissance indiqués avec leur erreur sur la pente. ND : Non déterminé.

Croissance de ND5F à pH libre en lait écrémé		
Phase de croissance	$\mu_I (h^{-1})^+$	$1,35 \pm 0,05$
	$\mu_{II} (h^{-1})^+$	$0,73 \pm 0,02$
	$V_{max}^a (10^3 \text{ upH} \cdot \text{min}^{-1})$	$-7,4 \pm 0,1$
	$T_{max}^a (h)$	$5,30 \pm 0,20$
	$pH_{max}^a (\text{upH})$	$5,41 \pm 0,26$
	$v_{lactate} (\text{mmol} \cdot \text{g}^{-1} \text{ de X} \cdot \text{h}^{-1})$	$18,2 \pm 3$
	$v_{lactose} (\text{mmol} \cdot \text{g}^{-1} \text{ de X} \cdot \text{h}^{-1})$	$-4,5 \pm 0,1$
Arrêt de croissance	Temps (h)	$7 \pm 0,2$
	Biomasse ($\text{g} \cdot \text{L}^{-1}$)	$1,25 \pm 0,05$
	pH (upH)	$4,76 \pm 0,04$
	[Lactose consommé] (mM)	$12,3 \pm 0,5$
	[Lactate total] (mM)	38 ± 2
	[Ac.lactique] (mM)	$4,02 \pm 0,5$
	$Y_{X/lactose} (\text{g} \cdot \text{g}^{-1})$	$0,28 \pm 0,06$
Post-acidification : Valeurs à 12 h	$Y_{lactate \text{ total}/lactose} (\text{g} \cdot \text{g}^{-1})$	$0,81 \pm 0,07$
	$pH_{12h} (\text{upH})$	$4,2 \pm 0,05$
	Baisse $pH_{12h} (\text{upH})$	$2,36 \pm 0,07$
	$V_{post}^a (10^3 \text{ upH} \cdot \text{min}^{-1})$	$-1,50 \pm 0,1$
	$[Lactate \text{ total}]_{12h} (\text{mM})$	74 ± 3
	$[Ac. \text{ lactique}]_{12h} (\text{mM})$	$21,2 \pm 0,8$
	$Y_{lactate/lactose 12h} (\text{g} \cdot \text{g}^{-1})$	$0,94 \pm 0,08$

Légende : μ_I et μ_{II} , respectivement les taux de croissance maximaux pendant la 1^{ère} et la 2nd phase exponentielle de croissance ; V_{max}^a la vitesse maximale d'acidification, T_{max}^a et pH_{max}^a le temps et le pH de culture à V_{max}^a , $v_{lactate}$ et $v_{lactose}$ les vitesses maximales spécifiques de production de lactate et de consommation de lactose ; [lactate total] concentration en acide lactique et lactate ; [Ac.lactique] concentration en acide lactique non dissocié ; $Y_{X/lactose}$ rendement en biomasse produite par gramme de lactose consommé, $Y_{lactate \text{ total}/lactose}$ rendement en lactate total produit par gramme de lactose consommé ; pH_{12h} pH mesuré à 12 h ; Baisse pH_{12h} diminution du pH en 12h et V_{post}^a la vitesse moyenne d'acidification pendant la post-acidification.

Tableau I. 4: Détermination des concentrations moyennes en acides aminés libres dans le lait avant inoculation et lors de la culture de la souche *L. lactis* ND5F à pH libre en lait écrémé, en mg.L⁻¹. Moyenne à partir de 4 réplicats biologiques. (D: composé détecté mais non quantifié (< 1 mg.L⁻¹), ND : Non détecté)

Acides aminés (mg.L ⁻¹)	Temps de culture (h)									
	Avant inoculation	1	2	3	3,5	4	5	6	7	24
ASP	2,5±0,8	ND	ND	1,9±0,5	5,2±3	2,2±1	3,6±0,5	3,1±0,4	4,3±0,5	5,9±1,2
GLU	48,7±8	50,5±2,5	49,3±2	54,2±1	48,0±1	21,5±0,5	18,5±2,5	18,8±8	30±2	73,8±5
ASN	1,1±0,3	ND	ND	1,6±0,5	ND	ND	ND	ND	2,2±0,5	ND
SER	1,1±0,8	1,0±0,8	0,9±0,6	1,9±1	D	1,9±0,5	2,0±0,3	1,9±0,7	1,8±0,2	ND
GLN	ND	1,7±0,7	1,6±0,4	1,6±0,4	1,5±0,2	1,6±0,5	1,6±0,5	1,6±0,5	1,3±0,5	ND
HIS	1,7±0,7	1,4±0,2	1,9±0,1	1,3±0,3	1,7±0,5	1,2±0,5	1,4±0,5	1,7±0,3	1,7±0,4	1,7±0,5
GLY	6,7±0,8	6,5±0,8	5,9±0,8	5,2±0,5	4,8±1,8	4,0±0,4	3,6±1,2	4,0±0,8	4,2±0,3	5,8±1,1
THR	3,6±0,5	2,5±0,3	2,0±0,7	1,4±0,7	D	ND	ND	ND	ND	4,0±0,1
ALA	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
ARG	4,5±0,4	D	D	ND	ND	ND	ND	ND	ND	1,9±0,8
TYR	ND	ND	ND	ND	ND	ND	ND	ND	ND	6,4±1,4
VAL	2,8±0,2	D	ND	ND	ND	ND	ND	ND	1,1±0,4	ND
MET	ND	ND	ND	D	D	D	D	D	D	ND
ILE	ND	ND	ND	D	D	D	D	D	1,1±0,8	1,8±0,1
PHE	2,6±0,7	2,1±0,7	2,2±0,7	2,3±0,7	2,3±0,8	2,2±0,6	2,0±0,4	1,7±0,7	1,5±0,7	1,4±0,7
LEU	ND	ND	ND	ND	ND	ND	1,1±0,8	1,1±0,8	D	ND
LYS	4,4±1	3,5±1	3,7±1	1,9±1	4,8±1	5,7±1	7,6±0,5	8,8±2	9,2±1	10,4±1
PRO	6,2±1,9	4,4±1	1,6±1,4	D	ND	D	D	4,6±0,4	8,6±0,4	72,3±3

Une diminution significative (P=0,95) de trois acides aminés (arginine, serine et thréonine) est mesurée à la fin des trois premières heures de culture, témoignant de la consommation rapide de ces acides aminés libres pour satisfaire les réactions anaboliques pendant la 1^{ère} phase de croissance.

Les besoins nutritionnels en matière protéique pour assurer une croissance en fin de 1^{ère} phase exponentielle de $1,05 \cdot 10^8$ UFC.mL⁻¹ sont estimés entre 0,26 µg.mL⁻¹ [Mills & Thomas 1981] et 0,31 µg.mL⁻¹ (calcul après utilisation des données sur la composition macromoléculaire de *Lactococcus lactis* ssp. *lactis*, thèse de Novak [Novak 1998]). Ces valeurs sont en décalage avec les apparents 6,3 µg.mL⁻¹ d'acides aminés libres consommés par ND5F, confirmant une utilisation en excédent par rapport à la croissance pendant cette phase. Entre 4 et 24 h de culture, une augmentation significative des concentrations de cinq acides aminés (Asp, Glu, Tyr, Lys et Pro) est

constatée ($P=0,99$), les autres acides aminés présentant une concentration non significativement différente au cours de la croissance. Cette production peut provenir soit de la libération ponctuelle d'acides aminés lors de l'hydrolyse de la caséine en oligopeptides par la protéase de paroi (cas démontré pour le glutamate et la proline [Juillard et al. 1995]), soit d'un efflux de ces acides aminés par les cellules. Cette observation atteste également de la fonctionnalité de la protéase de paroi à pH acide, puisqu'entre 7 h et 24 h de culture, les concentrations de glutamate et de proline augmentent alors que le pH est autour de 4,3.

Lors de l'arrêt de croissance à 7 h, seuls 4 acides aminés libres sont non détectables dans les surnageants (Ala, Arg, Tyr et Val). Cette concentration apparemment nulle résulte soit i) d'un déséquilibre entre la vitesse de consommation supérieure à la vitesse de production de ces acides aminés libres, et alors ces acides aminés peuvent-être considérés comme limitant, soit ii) d'une absence de production, et dans ce cas aucun flux de consommation ne peut être considéré. Dans ces deux hypothèses, il serait envisagé que ces acides aminés sont responsables de l'arrêt de croissance. Cependant, l'absence de dosage des acides aminés totaux dans le milieu (libres plus complexés en oligopeptides et caséines) ne permet pas d'affirmer la carence de ces derniers du milieu et donc leur réelle limitation intracellulaire. En effet, la multiplicité des sources d'azote dans la matrice lait complexifie l'analyse des besoins nutritionnels de ND5F et ne permet pas d'identifier à partir de ces données les acides aminés importants dans le déroulement du procédé actuel, ni éventuellement les autres composés (bases, vitamines) limitant. Pour répondre à la problématique des exigences nutritionnelles de ND5F, une approche complémentaire basée à la fois sur des cultures en milieu lait supplémenté, ainsi qu'un suivi des concentrations des substrats, a été menée. Les résultats de ces essais sont détaillés dans la section 0 de ce chapitre 1 dédiée à cette question.

- Cette analyse a permis de valider la condition de référence définie au laboratoire et de réaliser une étude macro-cinétique multicritères du comportement de la souche *L. lactis* ssp. *lactis* ND5F lors de sa culture sur lait écrémé. Trois phases ont été identifiées au cours de la culture : la phase de croissance, l'arrêt de croissance et la phase de post-acidification. Pour chacune d'entre elles, la cinétique de nombreux paramètres a été caractérisée, dont l'activité d'acidification du milieu. D'autres paramètres environnementaux ont été également analysés, dont les profils de consommation d'oxygène et de réduction du milieu, pour élargir la compréhension des interactions de ND5F avec son environnement. Un métabolisme homolactique de la souche a été mis en évidence, et son profil de consommation en acides aminés libres caractérisé. Cependant, une analyse ciblée sur les principaux facteurs environnementaux et les exigences nutritionnelles de ND5F est

nécessaire pour déterminer les causes de l'arrêt de croissance et de l'acidification et valider, ou non, les hypothèses formulées.

Impact des facteurs environnementaux

Afin d'approfondir la compréhension du comportement de *L. lactis* ND5F dans son environnement, les conséquences fonctionnelles de trois paramètres environnementaux ont été étudiées : i) acidification (pH exocellulaire et concentration en acide lactique), ii) changement de l'environnement gazeux (aérobie-anaérobie) et iii) thermisation différentielle du lait. Ces facteurs ont été tout particulièrement sélectionnés puisqu'en plus d'être identifiés dans la bibliographie comme ayant des conséquences sur la physiologie cellulaire, ils sont, à l'échelle industrielle, peu maîtrisés et difficilement évalués (à l'exception du pH).

I.3 Activité acidifiante et effet du stress acide sur la physiologie de ND5F

L'acidification du milieu, conséquence directe de la production d'acide lactique par ND5F, provoque une baisse du pH intracellulaire (pH_i) qui constitue, à terme, un stress sur la physiologie bactérienne. Deux facteurs, bien qu'interdépendants au cours de la culture à pH libre, sont à considérer dans l'effet toxique de cette auto-acidification [Kashket 1987; Cook & Russell 1994; Sanders et al. 1999] : (i) la production d'acide lactique, et notamment la forme acide dont la dissociation intracellulaire produit l'anion lactate et des H^+ ; (ii) l'auto-acidification du milieu, conséquence de la production d'acide lactique, qui augmente la demande énergétique pour maintenir le pH_i . Afin de caractériser la sensibilité de ND5F à ce stress acide, ces deux facteurs vont être étudiés séparément.

- Premièrement, la sensibilité des fonctions anaboliques de ND5F face à la toxicité de l'acide lactique a été déterminée. Dans ce but, des cultures statiques en tubes hermétiquement scellés et à pH régulé ont été réalisées à différentes concentrations initiales en acide lactique total, et le taux de croissance maximal a été mesuré pour chaque culture. Il est distingué la forme non dissociée qui correspond à l'acide lactique proprement dit, de la forme anionique appelée lactate. Le terme « acide lactique total » est utilisé pour désigner la somme des deux formes. Pour s'affranchir de toute inhibition nutritionnelle potentielle, ces expériences ont été conduites sur le milieu semi-synthétique YE à 50 g.L⁻¹ de lactose tamponné avec respectivement 9 et 5 g.L⁻¹ de KH_2PO_4 et K_2HPO_4 . La régulation de pH, manuelle dans ce dispositif, a été contrôlée grâce à l'établissement d'un modèle prédictif décrivant en fonction de la biomasse produite la quantité de NaOH 10 N à ajouter. Le pH cible des cultures a été fixé à $6,0 \pm 0,2$. Les concentrations initiales en acide lactique total ont été choisies pour balayer les gammes de concentrations rencontrées par la souche dans son

application en lait fermenté. Ainsi, il a été testé les concentrations suivantes : 0-0,2-0,3-0,4-0,5 et 1 M.

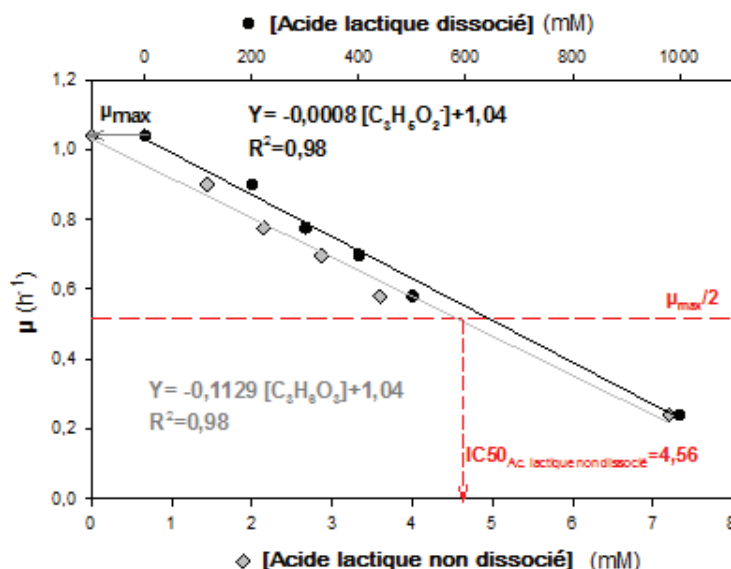


Figure I. 5 : Effet inhibiteur de l'acide lactique (total et non dissocié) sur le taux de croissance maximal en milieu YE à 50 g.L⁻¹ de lactose et à pH réglé à 6,0. Un modèle linéaire a été établi pour chaque forme et la concentration inhibitrice médiane de l'acide lactique sous forme non dissociée, IC50_{Ac. lactique}, a été déterminée. Chaque taux de croissance est le taux moyen issu de duplicats biologiques.

Pour visualiser l'effet potentiellement inhibiteur de chaque forme, le taux de croissance maximal de chaque culture est représenté en fonction des concentrations initiales respectives de lactate ou d'acide lactique (Figure I. 5). Une diminution du taux de croissance en fonction de la concentration exocellulaire en acide lactique est observée pour chaque forme (Figure I. 5). Cependant, l'établissement de deux modèles linéaires révèle que l'acide lactique sous forme non dissociée présente, à concentration égale, un effet inhibiteur 172,5 fois plus important que la somme composée du lactate produit et de l'acide lactique non dissocié. Ce résultat confirme que l'effet inhibiteur de l'acide lactique sur les fonctions anaboliques est porté presque exclusivement par la forme non dissociée au niveau exocellulaire. Pour quantifier cet effet sur la croissance de ND5F, la concentration inhibitrice médiane, IC50_{Ac. lactique}, correspondant à la concentration en acide lactique nécessaire pour inhiber de moitié le taux de croissance maximal, a été déterminée à 4,56 mM. Ces résultats permettent de prédire la réponse anabolique de ND5F à l'inhibition par l'acide lactique en YE. Ainsi, d'après le modèle établi, le taux de croissance serait nul pour une concentration en acide lactique non dissocié de 7,44 mM. Afin de valider le modèle, ce point expérimental nécessiterait d'être réalisé. Autre remarque, la quantité d'acide lactique produit par ND5F au cours de sa croissance n'a pas été intégrée aux concentrations initiales en acide lactique. Néanmoins comme la phase exponentielle de croissance a lieu dès le début de la culture, la concentration en lactate

produit est inférieure à 20 mM, ce qui représente moins de 0,15 mM d'acide lactique et justifie l'absence de correction.

La concentration en acide lactique non dissocié mesurée à l'arrêt de croissance sur la condition de référence est de 4,02 mM. A cette concentration, le taux de croissance théoriquement observé serait encore de $0,47 \text{ h}^{-1}$. Ce résultat confirmerait que l'inhibition par l'acide lactique ne peut être l'unique responsable de l'entrée en phase stationnaire, et qu'une limitation nutritionnelle ou un effet inhibiteur du pH exocellulaire sont à rechercher également.

Cependant, pour pouvoir transférer ces résultats aux essais sur matrice lait, il est nécessaire de supposer que l'effet inhibiteur de l'acide lactique est équivalent quel que soit le milieu et le pH exocellulaire. Or, le flux anabolique d'une cellule varie en fonction des disponibilités en substrats et du coût énergétique pour maintenir un gradient de pH suffisant. Ces deux critères sont dépendants du milieu puisqu'ils sont fonction des propriétés nutritionnelles et du pouvoir tampon de ce dernier. Ainsi, pour aller plus loin dans cette analyse, et évaluer si les résultats d'inhibition par l'acide lactique peuvent être appliqués sur lait, un comparatif des activités cellulaires de ND5F en phase exponentielle de croissance a été réalisé entre le milieu YE et le lait écrémé à pH libre (Tableau I. 5).

Tableau I. 5 : Comparatif des principaux critères cinétiques de caractérisation de l'activité de croissance et de production de lactate de ND5F en phase exponentielle de croissance entre le milieu YE et lait écrémé sans régulation de pH, en micro-aérobiose. Les données sur milieu YE sont issues de la moyenne de deux réplicats biologiques. Les résultats sont présentés avec leurs écarts-types issus de l'intégration de quatre réplicats biologiques à l'exception (*) des taux de croissances indiqués avec leur erreur sur la pente. Détermination de la significativité des valeurs par t-Test : p-value < à 0,05 (*), < à 0,01 (**) et < à 0,001 (***), la référence est la culture en lait écrémé.

Paramètres	YE à 50 g.L^{-1} de lactose	Lait écrémé
Rendement de production de biomasse à $t=4\text{h}$ $Y_{X/\text{lactose}} (\text{g.g}^{-1})$	0,30	$0,34 \pm 0,07$
Taux de croissance maximal à $t=4\text{h}$ $\mu (\text{h}^{-1})^*$	1,04	$0,73 \pm 0,03$
Vitesse spécifique maximale de consommation du lactose à $t=4\text{h}$ $v_{\text{lactose}} (\text{mmol.g}^{-1} \text{ de X. h}^{-1})$	-5	$-4,5 \pm 0,1$
Vitesse spécifique maximale de production du lactate à $t=4\text{h}$ $v_{\text{lactate}} (\text{mmol.g}^{-1} \text{ de X. h}^{-1})$	17	$18,2 \pm 3$

Les vitesses spécifiques de production de lactate et de consommation du lactose sont équivalentes entre les deux milieux en phase exponentielle (Tableau I. 5). Le taux de croissance maximal est, quant à lui, supérieur de 43 % sur YE, alors que le rendement de production de biomasse par rapport au

lactose consommé est identique. Il est donc probable que la différence observée s'explique par une disponibilité en acides aminés, principaux substrats anaboliques, supérieure sur YE. Des données énergétiques complémentaires, comme le pH intracellulaire, et une étude spécifique de l'effet inhibiteur de l'acide lactique en lait seraient nécessaires, mais cependant, à la vue de ces résultats, il sera considéré que l'activité inhibitrice de l'acide lactique n'est pas différente entre YE et le lait.

▪ L'influence du pH exocellulaire sur les fonctions cellulaires de ND5F, deuxième facteur identifié du stress acide, a été étudiée à partir de cultures en lait à pH régulé 6,5-6-5,5 et 5, réalisées en fermenteurs. Les conséquences de cette régulation ont été examinées sur les paramètres macro-cinétiques précédemment retenus pendant la phase de croissance, à son arrêt et en phase de post-acidification. Le choix a été fait de travailler en lait pour obtenir des conclusions plus facilement transposables pour l'interprétation des phénomènes dans la condition de référence. Des valeurs de pH supérieures ou égales à 5 ont été testées pour rester à des concentrations en acide lactique non dissocié relativement faibles, même si une production importante d'acide lactique total a lieu, et réduire ainsi l'impact de ce paramètre. Les principaux résultats de cette analyse sont résumés Tableau I. 6.

Il est mesuré sur tous les indicateurs, à l'exception du rendement en lactate total produit, une diminution avec l'abaissement de la valeur de consigne de régulation du pH (Tableau I. 6). Ces résultats permettent d'identifier un effet inhibiteur du pH du milieu dès un pH égal ou en deçà de 6. Autre observation, ces résultats montrent que le pH interagit aussi bien sur l'activité anabolique (concentration en biomasse, taux et durée de la phase de croissance) que sur l'activité catabolique (consommation de lactose et production de lactate). Ce résultat était attendu puisque l'acidification du milieu affecte le gradient de pH à travers la membrane, ΔpH , et nécessite une réponse cellulaire globale pour maintenir l'homéostasie cellulaire et un pH intracellulaire adéquat pour les activités enzymatiques anaboliques et cataboliques.

Dans toutes ces cultures, la croissance est rallongée de 1 à 2 h par rapport à la cinétique à pH libre. Néanmoins, sur ces cultures régulées, le titre en biomasse n'est jamais supérieur à la condition de référence. Les essais à pH régulé à 6,5 et 6 égalent la quantité finale de biomasse précédemment obtenue à pH libre (voir Tableau I. 3 et Tableau I. 6). Ce résultat permet donc d'écarter toute responsabilité de l'acidification du milieu sur l'arrêt de croissance de ND5F. De plus, les concentrations en acide lactique non dissocié sont restées en deçà de l' $IC_{50,Ac.lactique}$ pour les régulations de pH à 6,5-6-5 et légèrement supérieure pour la culture à pH régulé 5,5, ce qui conforte l'hypothèse d'un arrêt des fonctions anaboliques de ND5F sur lait, à pH libre, par carence nutritionnelle.

Tableau I. 6 : Analyse macro-cinétique du comportement de *ND5F* à différents pH régulés en lait écrémé. Les principaux paramètres divergeants sont surlignés en gris. Détermination de la significativité des valeurs par t-Test : p-value < à 0,05 (*), < à 0,01 (**) et < à 0,001 (***), la référence est la culture à pH 6,5.

Période	Paramètres	Valeur de pH			
		6,5	6,0	5,5	5,0
Phase de croissance	μ_I (h ⁻¹)	1,35±0,05	1,36	1*	0,21**
	μ_{II} (h ⁻¹)	0,73±0,02	0,73	0,54*	0,47**
	$v_{lactate}$ (mmol.g ⁻¹ de X. h ⁻¹)	28,8±1,5	20,7	19,4	16,9
	$v_{lactose}$ (mmol.g ⁻¹ de X. h ⁻¹)	-6,8±0,2	-5,4	-4,9	-3,9
Arrêt de croissance	Temps (h)	9±0,3	8,5±0,2	8±0,2	8±0,2
	Biomasse (g.L ⁻¹)	1,25±0,02	1,21±0,02	1,08±0,03*	0,41±0,05***
	[Lactose consommé] (mM)	22,48±2	20,25	18,35	15,84
	[Lactate total] (mM)	111,79±5	70,61	65,07	52,89
	[Ac.lactique] (mM)	0,26±0,03	0,51	1,52	3,57
	$Y_{X/lactose}$ (g.g ⁻¹)	0,16±0,1	0,17	0,17	0,08
	$Y_{lactate\ total/lactose}$ (g. g ⁻¹)	0,87±0,03	0,91	0,93	0,87
Post-acidification	[Lactate total] _{24h} (mM)	375±10	285	158	105
	[Ac. lactique] _{24h} (mM)	1,70±6	2,05	3,54	7,09
	$Y_{lactate/lactose24h}$ (g. g ⁻¹)	0,94	0,96	0,98	0,97

Malgré cette limitation nutritionnelle, un effet différentiel du pH sur les taux de croissance est mis en évidence. En effet, alors qu'aucun effet significatif de la régulation de pH n'est mesuré sur les taux de croissance des phases exponentielles I et II à pH 6,5 et 6, une diminution de ces taux de respectivement 84 et 53 % est calculée par rapport à la cinétique à pH libre pour la culture à pH 5. L'effet inhibiteur de l'acidification du milieu est donc plus marqué sur la phase d'assimilation des acides aminés et di-/tri-peptides libres du lait que lors de l'intégration des oligopeptides issus de l'hydrolyse de la caséine. Ce résultat n'est pas surprenant puisque la 1^{ère} phase de la référence se déroule à pH libre lorsque le lait est encore peu acidifié ($6,39 \pm 0,02$) alors que la 2nd phase s'accomplit jusqu'à un pH de $5,6 \pm 0,1$. Physiologiquement, cette différence de sensibilité s'explique par la dépendance énergétique des transporteurs concernés. En effet, une partie des transporteurs des acides aminés libres et des di-/tri-peptides (transporteur DtpT) sont des perméases qui nécessitent pour fonctionner une force protomotrice, alors que l'ABC transporteur Opp des oligopeptides utilise l'énergie d'hydrolyse de l'ATP. Aucun modèle linéaire prédictif de l'effet inhibiteur du pH sur le taux de croissance n'a pu être établi démontrant l'absence de proportionnalité du phénomène et la complexité de la réponse physiologique adaptative. Les

rendements en biomasse ne sont pas significativement modifiés entre les pHs 6,5 et 5,5 à la différence de la condition à pH 5 où la diminution de la valeur de rendement traduit une baisse de l'efficacité de biosynthèse.

Les concentrations en lactate total à l'arrêt de croissance sont supérieures dans les cultures régulées à celles enregistrées sur la cinétique à pH libre. Cette augmentation des concentrations s'explique à la fois par une vitesse spécifique de production de lactate, v_{lactate} , supérieure dès une régulation de pH supérieure à 5, et par une durée de croissance (donc de production) rallongée. La vitesse spécifique de consommation de lactose, v_{lactose} , est également proportionnellement augmentée, d'un facteur $0,25 \pm 0,02$ par rapport à v_{lactate} . La conservation du rapport théorique de 4 lactates pour un lactose est une nouvelle démonstration du métabolisme homolactique. De plus, la somme des rendements en biomasse et en lactate total produit est équivalente à $1 \pm 0,13$. Pendant la post-acidification, les concentrations en lactate total sont largement supérieures à la condition de référence. Ce résultat confirme l'hypothèse précédemment établie que l'acidification du milieu est responsable de l'arrêt des fonctions cataboliques. Le rendement en lactate produit pendant cette phase est très proche de 1 ($0,95 \pm 0,3$) et illustre l'orientation exclusive du flux de lactose vers la production de lactate.

- En conclusion, cette étude a permis de décrypter les conséquences de l'activité acidifiante sur la physiologie de ND5F en identifiant les responsabilités des deux facteurs inhibiteurs (l'acide lactique non dissocié et la diminution du pH du milieu) sur les fonctions anaboliques et cataboliques. Ainsi, la toxicité de l'acide lactique non dissocié sur le taux de croissance de ND5F a été modélisée, autorisant la détermination de sa concentration inhibitrice médiane, $IC_{50_{\text{Ac.lactique}}}$, à 4,56 mM. Ce paramètre, bien qu'impliqué dans la diminution du taux de croissance, n'est pas responsable de l'arrêt de croissance aux concentrations exocellulaires produites pendant la culture en lait à pH libre. Il a été également établi que la diminution du pH exocellulaire agit aussi bien sur l'efficacité de biosynthèse cellulaire que sur la production de lactate. Les conséquences macro-cinétiques de cette inhibition ont été calculées et décrites en matrice lait pour ND5F. En particulier, le pH du milieu a une action différentielle en fonction de la phase de croissance exponentielle avec une valeur seuil autour de 5,5 en deçà de laquelle les conséquences physiologiques sont beaucoup plus marquées. De plus, l'acidification du milieu est identifiée comme responsable de l'arrêt de la phase d'acidification de ND5F, et les limitations nutritionnelles comme le principal élément de l'arrêt de croissance en lait à pH libre.

La culture de référence de ND5F est réalisée dans un environnement gazeux non contrôlé et susceptible d'être modifié lors d'un transfert d'échelle du procédé. Il est donc nécessaire de caractériser les conséquences physiologiques de toute modification d'atmosphère gazeuse sur la croissance et l'activité acidifiante de ND5F.

I.4 Effet de l'environnement gazeux sur la croissance et l'acidification

Bien que dépourvus de l'un des principaux systèmes de détoxification des radicaux libres oxygénés, les catalases, les lactocoques sont des bactéries anaérobies aérotoles pour lesquelles la présence d'oxygène n'est pas létale. Dans la condition de référence, la croissance de ND5F se déroule en aérobiose entre 0 et 1,5 h de culture avant que la consommation d'oxygène par la souche n'abaisse la PO_2 à une valeur nulle pour la suite de la fermentation (voir Figure I. 2). Pour déterminer si cette consommation d'oxygène lors de la 1^{ère} phase exponentielle de croissance engendre, au travers de la production d'espèces réactives à l'oxygène (H_2O_2 , $OH^\cdot$, $O_2^\cdot$), une toxicité cellulaire et éventuellement une diminution des vitesses de croissance et d'acidification, des cultures en condition d'anaérobiose (dégazage préalable du milieu au N_2 et surpression pendant la fermentation) ont été menées en lait écrémé et YE. De plus, dans un but de caractérisation plus large de l'effet de l'oxygène, des cultures en condition aérobie ont été réalisées sur YE. En effet, le maintien d'une aération pendant la fermentation du lait provoque un moussage important, du fait du fort pourcentage en matière protéique, ce qui empêche la réalisation de ces conditions sans ajout massif d'anti-mousse. Les principaux descripteurs de la croissance et de l'acidification sont présentés Tableau I. 7.

Comparativement à l'anaérobiose, l'aération initiale du lait écrémé n'a aucun effet significatif sur les taux de croissance des phases exponentielles I et II. Cette observation est reproductible pour les vitesses spécifiques de consommation de lactose et de production de lactate qui présentent également des valeurs similaires entre les conditions de micro-aérobiose et d'anaérobiose. Ces résultats indiquent que la présence transitoire d'oxygène semble n'avoir aucun impact significatif sur la croissance de ND5F en lait.

La culture en aérobiose sur YE révèle une diminution significative du taux de croissance maximal (Tableau I. 7) et de 20% des vitesses spécifiques de consommation de lactose et de production de lactate par rapport à la culture en anaérobiose. Ainsi, la présence continue d'oxygène ralentit la cinétique de croissance ainsi que le catabolisme.

La présence initiale d'oxygène modifie, en milieu YE et lait, les descripteurs de la cinétique d'acidification de la souche. Ainsi, l'intensité de la vitesse maximale d'acidification est diminuée de près de $2.10^3 \text{ upH.min}^{-1}$ et il apparaît également un retard dans l'acidification, visible au niveau du T_{max}^a et du pH_{max}^a obtenus en présence d'oxygène. Ce résultat est cohérent avec les précédentes observations de Jeanson sur l'accélération de la phase d'acidification en absence d'oxygène [Jeanson et al. 2009], et cela, bien que la vitesse maximale de production de lactate soit équivalente dans le cas du lait

Tableau I. 7 : Comparaison des principales données macro-cinétiques au cours des cultures en fermenteurs en lait écrémé et YE à pH libre dans la condition de référence ou en condition d'anaérobiose. Les principaux paramètres divergeants sont surlignés en gris. Pour le détail, voir légende Tableau I. 3. Détermination de la significativité des valeurs par t-Test : p-value < à 0,05 (*), < à 0,01 (**) et < à 0,001 (***), la référence est la culture en micro-aérobiose en lait et en aérobiose sur YE.

Période	Paramètres	Lait écrémé		YE à 50 g.L ⁻¹	
		Micro-aérobiose Anaérobiose	Anaérobiose	Aérobiose PO ₂ à 50%	Anaérobiose
Phase de croissance	$\mu_I \text{ (h}^{-1}\text{)}$	1,35 ± 0,05	1,40	0,89 ± 0,02*	1,04 ± 0,01*
	$\mu_{II} \text{ (h}^{-1}\text{)}$	0,73 ± 0,02	0,73		
	$V_{\text{max}}^a \text{ (10}^3 \text{ upH.min}^{-1}\text{)}$	-7,4 ± 0,1	-9,8	-11,5	-13,4
	$T_{\text{max}}^a \text{ (h)}$	5,30 ± 0,20	5,00	4,5	3,5
	$\text{pH}_{\text{max}}^a \text{ (upH)}$	5,41 ± 0,26	5,80	5,80	6,20
	$V_{\text{lactate}} \text{ (mmol.g}^{-1} \text{ de X. h}^{-1}\text{)}$	18,2 ± 3	18	13,9	17
	$V_{\text{lactose}} \text{ (mmol.g}^{-1} \text{ de X. h}^{-1}\text{)}$	-4,5 ± 0,1	-4,5	-4,2	-5
Arrêt de croissance	Temps (h)	7 ± 0,2	6 ± 0,2	6 ± 0,3	5 ± 0,1
	Biomasse (g.L ⁻¹)	1,27 ± 0,05	1,21 ± 0,05	1,15 ± 0,05	0,921 ± 0,02*
	pH (upH)	4,76 ± 0,04	4,50	4,32	4,23
	[Lactose consommé] (mM)	12,3 ± 0,5	15,2	13,4	6,5
	[Lactate total] (mM)	38 ± 2	44	28	20
	[Ac.lactique] (mM)	4,02 ± 0,5	7,46	7,02	7,15
	$Y_{X/\text{lactose}} \text{ (g.g}^{-1}\text{)}$	0,28 ± 0,06	0,24	0,25	0,35
	$Y_{\text{lactate total}/\text{lactose}} \text{ (g.g}^{-1}\text{)}$	0,81 ± 0,07	0,77	0,55	0,95

Ces différences sont retrouvées lors de l'arrêt de la croissance, avec des cinétiques plus longues pour les cultures en aérobiose ou en micro-aérobiose, par rapport à l'anaérobiose. Ainsi, en lait, pour une concentration finale en biomasse équivalente, le pH d'arrêt de croissance est plus bas en anaérobiose. Cette observation illustre une activité catabolique de production de lactate supérieure en anaérobiose, comme l'illustrent les rendements de production de lactate plus faibles en présence d'oxygène. En particulier en milieu YE, le changement d'atmosphère en oxygène contribue à une

réorganisation des flux métaboliques au sein du métabolisme carboné central aboutissant au shift d'un métabolisme homolactique vers un métabolisme mixte, comme en témoigne le rendement $Y_{\text{lactate total/lactose}}$ de $0,55 \text{ g.g}^{-1}$ [Garrigues et al. 1997]. Il se traduit par une production d'acétate, d'éthanol et de diacétyle/acétoïne de respectivement 11, 10 et 5 mM dans cette condition aérée, en plus des 28mM de lactate total.

Il est probable que l'arrêt plus précoce de la croissance en milieu YE et en condition anaérobie en lait s'explique par la concentration en acide lactique non dissocié. En effet, les valeurs calculées (Tableau I. 7) sont proches de la valeur prédite d'inhibition à 100 % du taux de croissance, de 7,44 mM.

Le profil du potentiel d'oxydoréduction est également fortement influencé par la présence d'oxygène (résultats non présentés). En présence d'oxygène, le potentiel E_{h7} reste oxydant, autour de +100 mV sur YE. Au contraire, l'absence d'oxygène singularise les cultures sur lait et YE par une réduction rapide de l' E_{h7} (0,5 h) en lieu et place des 1,5 h. Le potentiel final obtenu en lait est identique à la condition de référence, avec -150 mV.

- Ainsi, le taux d'oxygène dissout dans le lait apparaît comme un levier important sur le contrôle des cinétiques de croissance, d'acidification et de réduction, en plus de la réorganisation du métabolisme central de la souche. De plus, bien que non étudié ici, il est connu que la présence d'oxygène affecte également la production d'arômes. D'une part, elle agit de manière directe en intervenant en tant qu'acteur au sein des phénomènes chimiques d'oxydation des lipides [Mc Sweeney 2004; McSweeney & Sousa 2000] et autres composés carbonés tels que le diacétyle [Bassit et al. 1993; Monnet et al. 1994]. D'autre part, la présence d'oxygène interagit indirectement en maintenant artificiellement un potentiel redox oxydant qui impacte le catabolisme des acides aminés en lait [Kieronczyk et al. 2006]. La prise en compte de la tolérance et/ou de la consommation de l'oxygène est donc un paramètre important pour la compréhension du procédé laitier.

Les études sur le stress acide ainsi que sur l'effet de l'atmosphère gazeuse ont confirmé l'impossibilité pour ND5F d'atteindre une biomasse finale supérieure à $1,27 \text{ g.L}^{-1}$ sur lait écrémé. Avant d'étudier en détail les besoins nutritionnels de la souche, et les bénéfices d'une supplémentation ciblée en éléments nutritifs, nous avons souhaité aborder les conséquences des différents traitements thermiques sur la fermentescibilité d'une même matière première.

I.5 Thermisation différentielle de la matrice lait

Afin d'éliminer la flore endogène du lait, il subit un traitement thermique préventif (pasteurisation, stérilisation UHT, ou reconstitution de poudre de lait préalablement séchée en condition stérile) qui peut différer d'une usine à l'autre ou en fonction du produit à réaliser. Ce paramètre est à considérer dans la définition d'un procédé industriel puisqu'il a déjà été signalé que la thermisation du lait modifie sa composition nutritionnelle et affecte donc sa fermentescibilité [Branger et al. 2007]. En effet, les traitements thermiques présentent l'avantage de libérer des « facteurs de croissance » (acides aminés, bases purines et pyrimidines, vitamines) mais participent également à la synthèse de composés toxiques par l'intermédiaire de la réaction de Maillard. A partir de ces informations, deux nouveaux traitements de stérilisation du lait écrémé ont été développés à côté de la stérilisation dite « classique » (15 min à 115 °C) pour tester un effet éventuel de ce paramètre. Premier cas, le lait dit « autoclavé » subit une thermisation renforcée avec 20 min à 120°C, tandis que dans le second cas, le lait écrémé est reconstitué stérilement sans aucun traitement thermique ni microfiltration. Il est à noter qu'un test préalable de dénombrement de la flore totale a permis de vérifier que la contamination était inférieure à 60 UFC.mL⁻¹ même sans traitement thermique.

Pour analyser l'impact au niveau physiologique de cette thermisation, les profils de croissance et d'acidification sont présentés pour les trois conditions (Figure I. 6A). La croissance de *L. lactis* ND5F à pH libre dans le lait reconstitué stérilement s'arrête dès 6 h de culture et la concentration cellulaire atteinte n'est que de 0,92 g.L⁻¹. La cinétique d'acidification est retardée avec un pH à l'arrêt de croissance de 5,7 soit de près d'une unité supérieure au pH de la culture sur lait classique. De plus, l'absence de thermisation du lait a pour conséquence de rendre inexistante la 1^{ère} phase de croissance habituellement constatée sur le profil de référence. Les fermentations réalisées dans le lait autoclavé présentent un profil de croissance, tout comme une activité acidifiante, non significativement différents du lait classique. Cependant, l'analyse cinétique révèle une augmentation de la durée de la 1^{ère} phase de croissance (3 h au lieu de 2h) avec un taux de croissance légèrement supérieur, de 1,42 h⁻¹, bien que non significatif. Ces observations indiquent que certains facteurs de croissance semblent favorisés par un traitement thermique. L'action notée en particulier sur la 1^{ère} phase de croissance semblerait indiquer une libération augmentée d'acides aminés soit sous forme libre, de di/tri-peptides ou d'oligopeptides après une action de thermisation. Cette hypothèse va être examinée ci-après à partir des dosages des acides aminés libres dans le lait.

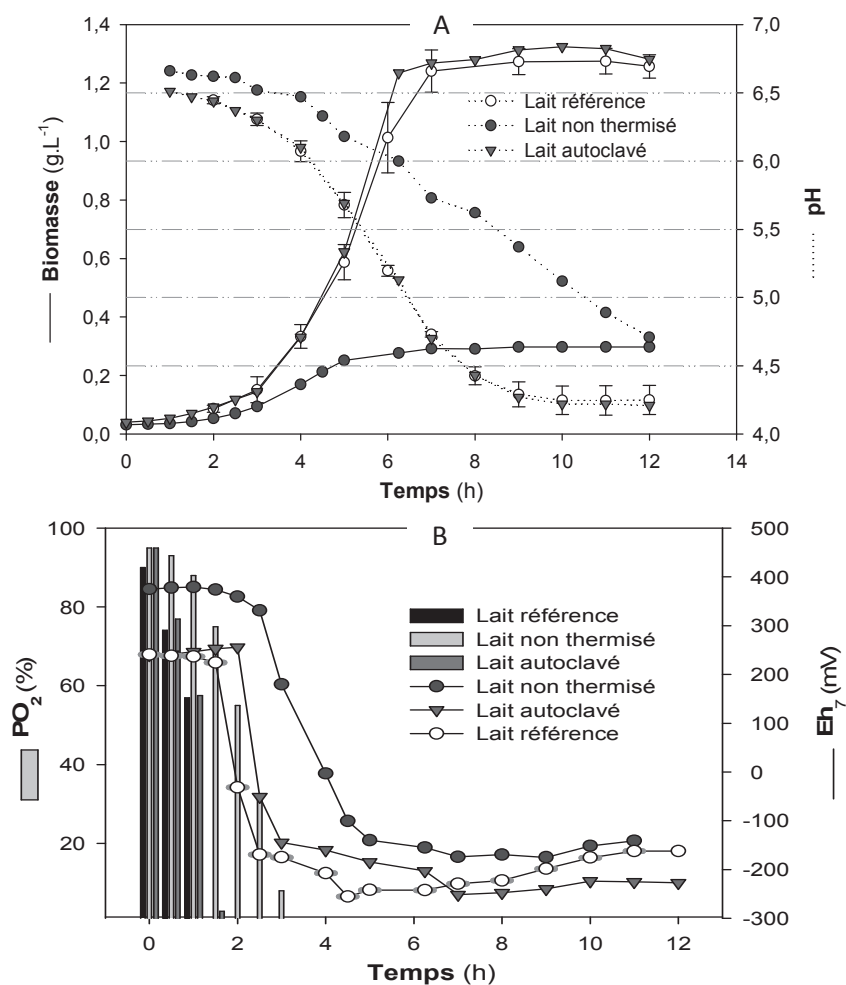


Figure I. 6: Comparatif des cinétiques de croissance et d'acidification (A) ainsi que de consommation d'oxygène et d'activité réductrice (B) en fonction du traitement thermique appliqué sur la matrice lait lors de la culture à pH libre de ND5F. Ecart-types issus de 4 répliquats biologiques pour le lait classique.

Le décalage dans l'activité de biosynthèse de ND5F sur le lait non thermisé est retrouvé sur la cinétique de la PO₂, avec une consommation d'oxygène retardée d'1,5 h (3 h au lieu de 1,5 h) (Figure I. 6.B). Le lait non thermisé présente également un potentiel Eh₇ initial plus oxydant de près de 20 mV en comparaison aux deux laits ayant subi un traitement thermique. Cependant, l'Eh₇ final du lait non thermisé est comparable au lait classique. Les profils de pression partielle en oxygène sont similaires entre le lait référence et le lait autoclavé, mais les cinétiques de réduction divergent à 3 h entre les deux cultures. En effet, la vitesse de réduction de ND5F sur lait autoclavé est deux fois plus lente (-15 mV.h⁻¹) et, une fois la valeur de -225 mV atteinte (à 7 h au lieu de 4,5 h), le potentiel se stabilise et aucune oxydation significative du milieu n'est observée, à la différence du lait classique. Ainsi, l'Eh₇ final entre ces deux laits est plus réducteur de 60 mV en lait autoclavé, avec une valeur stable à 24 h. Les conséquences de cet état réducteur du milieu sur les éventuelles voies

métaboliques de production d'arômes pendant la conservation du produit seraient à analyser pour compléter cette étude.

Pour tenter d'identifier les composés à l'origine de ces différences, le dosage avant inoculation des acides aminés libres dans le lait pour les trois modes opératoires de préparation sont présentés dans le Tableau I. 8. Il apparaît une diminution des concentrations de près de 40 à 50 % des acides aminés précédemment détectés, entre le lait de référence et le lait non thermisé (à l'exception de la lysine). Ces résultats apportent la preuve qu'au minimum une part de la fermentescibilité accrue du lait après thermisation s'explique sur le plan de l'accessibilité en substrat azoté, avec une augmentation des acides aminés libres. Il est également probable que les concentrations en oligopeptides provenant de l'hydrolyse thermique de la caséine soient également plus importantes si des acides aminés sont libérés. Cependant, aucun résultat ne permet de vérifier cette hypothèse. Les dosages des acides aminés dans le lait autoclavé amènent à des observations contrastées, avec d'une part des concentrations en glutamate, asparagine, valine ou même en leucine doublées par rapport au lait de référence, et d'autre part une diminution du même ordre de grandeur pour la lysine, la proline et l'arginine. Ainsi, il apparaît des effets différentiels en fonction de la nature des acides aminés lors d'une thermisation plus intense.

Une recherche bibliographique indique que la lysine, la proline et l'arginine sont identifiées comme faisant partie des acides aminés les plus réactifs pour entreprendre une réaction de Maillard avec la source de carbone présente [Moens et al. 2004; Kwak & Lim 2004]. De plus, la coloration légèrement brunie prise par le lait à la sortie de l'autoclave aurait tendance à confirmer la réalisation de cette réaction chimique et expliquerait que ces acides aminés engagés dans des réactions de condensation avec le lactose soient en concentrations plus faibles. Si une production de composés toxiques pour la cellule a eu lieu en aval à cette réaction de Maillard, aucune observation sur la culture de ND5F dans ce lait ne permet de le vérifier.

- Ainsi, un impact de la thermisation sur la fermentescibilité du lait ainsi que sur l'activité acidifiante de ND5F a pu être mis en évidence en fonction de l'intensité du traitement. En particulier, une libération d'acides aminés libres a été identifiée et semble participer à la concentration cellulaire plus importante observée dans les laits traités thermiquement. Cependant, lorsque le traitement est plus intense, certains acides aminés ne sont plus disponibles à l'état libre du fait d'un probable engagement dans une liaison avec le carbonyle du lactose. Chacune de ces modifications change la composition du lait, mais n'affecte pas de la même façon le comportement macro-cinétique de la souche et nécessite donc un diagnostic global des conséquences de chaque paramètre (acidification,

réduction du milieu, croissance) pour envisager l'impact futur lors d'une production de laits fermentés.

Tableau I. 8: Détermination des concentrations en acides aminés libres dans le lait avant inoculation en fonction du traitement thermique appliqué. D: composé détecté mais non quantifié ($< 1 \text{ mg.L}^{-1}$), ND : Non détecté.

Acide Aminés (mg.L^{-1})	Type de traitement thermique		
	Lait référence	Lait autoclavé	Lait non thermisé
ASP	2,5±0,8	1,5±0,7	ND
GLU	48,7±8	75,8±5,7	12,5±2
ASN	1,1±0,3	3,5±0,8	ND
SER	1,1±0,8	1,2±0,2	D
GLN	ND	D	D
HIS	1,7±0,7	1,1±0,3	2,1±0,1
GLY	6,7±0,8	6,5±0,4	4,1±0,6
THR	3,6±0,5	3,5±0,3	ND
ALA	ND	ND	ND
ARG	4,5±0,4	2,5±0,5	D
TYR	ND	ND	ND
VAL	2,8±0,2	4,6±0,7	D
MET	ND	ND	ND
ILE	ND	ND	ND
PHE	2,6±0,7	2,1±0,7	2,2±0,7
LEU	ND	1,1±0,1	ND
LYS	4,4±1	1,5±1	7,5±1
PRO	6,2±1,9	3,4±0,7	5,1±0,4

▪ En conclusion, chacun des paramètres environnementaux étudiés a démontré son importance sur le comportement macro-cinétique de la souche (croissance, acidification et réduction du milieu) et par conséquent sur les fonctions de ferments de la souche (formation d'un gel lactique, bio-préservation des produits (acidité et concentration en acide lactique non dissociée), flaveur (acidité finale perçue et production de métabolites)). Tous ont mis en évidence l'importance de la connaissance des besoins nutritionnels de ND5F pour prolonger et augmenter la croissance.

Exigences nutritionnelles

Une supplémentation du lait ciblée en éléments nutritionnels est difficilement envisageable dans une démarche d'industrialisation d'un procédé alimentaire, en particulier du fait des problématiques de coût. Cependant, identifier les exigences nutritionnelles spécifiques de ND5F pourrait permettre d'agir sur les leviers et verrous de la production de biomasse. Pour répondre à la problématique de définition des besoins de ND5F, des cultures comparatives sur lait écrémé à pH régulé à 6,5 (indépendance face à l'effet d'auto-acidification), en micro-aérobiose, supplémentées en différents éléments, ont été menées pour identifier les principaux nutriments responsables de l'arrêt de croissance. Trois types de cultures ont été réalisées, via une supplémentation par i) une solution d'acides aminés libres, ii) une solution de vitamines et iii) un mélange d'acides aminés, vitamines et bases. Les compositions et concentrations de ces solutions sont identiques à celles définies pour le milieu MCD [Cocaign-Bousquet et al. 1995] (voir matériels et méthodes). Les paramètres macro-cinétiques de croissance et de production de lactate dans les différentes conditions testées sont résumés Tableau I. 9.

Sur les quatre conditions testées, seuls les milieux avec ajout d'acides aminés libres (seuls ou en combinaison) se singularisent du profil de référence de ND5F à pH régulé à 6,5 sur la majorité des critères cinétiques. En effet, à la différence de l'ajout de vitamines seules, l'addition unique d'acides aminés permet d'augmenter significativement le titre final en biomasse d'un facteur 2,4, avec 3 g.L⁻¹ (Tableau I. 9). De plus, la durée de la 1^{ère} phase exponentielle, tout comme de la période de croissance, est également significativement prolongée (Tableau I. 9). Cependant, aucune différence significative entre les taux de croissances n'a pu être établie. Les acides aminés libres peuvent-être considérés comme les éléments nutritionnels responsables de l'arrêt de croissance de ND5F en lait étant donné la poursuite de la croissance observée.

Cependant, on peut s'interroger sur l'identification exacte des acides aminés responsables d'une telle carence. A la fin de la croissance à pH régulé à 6,5 sur lait écrémé classique, 4 acides aminés étaient absents dans les surnageants (Ala, Arg, Tyr et Val), tout comme sur lait à pH libre. L'absence d'accumulation de ces acides aminés peut éventuellement traduire un état de limitation et désigner ces acides aminés comme responsables de l'arrêt de croissance. Cependant, rien ne permet de mettre en exergue que ce sont ces acides aminés qui étaient réellement limitant, puisque tous les acides aminés ont été ajoutés. En effet, il est possible que la limitation mise en évidence apparaisse avant une concentration nulle des acides aminés et qu'il existe une valeur seuil à partir de laquelle l'efficacité de transport se trouve réduite et ne permet plus de soutenir la croissance. Pour choisir entre ces deux hypothèses, une culture après adjonction spécifique de ces 4 acides aminés devrait

être réalisée pour observer si la croissance est prolongée. L'utilisation des acides aminés comme substrats anaboliques se visualise sur la diminution d'un facteur deux du rendement de production de biomasse à partir du lactose.

Tableau I. 9 : Comparaison de la croissance de ND5F en fonction des différentes catégories nutritionnelles complétant le lait écrémé à pH régulé à 6. Le détail de la légende est présenté Tableau I. 3. Détermination de la significativité des valeurs par t-Test : p-value < à 0,05 (*), < à 0,01 (**) et < à 0,001 (***), la référence est la culture en lait classique.

Période	Paramètres	Lait classique	Lait +acide aminé	Lait +vitamines	Lait +acide aminé +vitamines +bases
Phase de croissance	μ_I (h^{-1})	1,35±0,05	1,44±0,02	1,36±0,03	1,48±0,02
	μ_{II} (h^{-1})	0,73±0,03	0,76±0,03	0,74±0,03	0,77±0,03
	$v_{lactate}$ (mmol.g ⁻¹ de X. h ⁻¹)	28,8±1,5	31,2	27,98	38,91
	$v_{lactose}$ (mmol.g ⁻¹ de X. h ⁻¹)	-6,8±0,2	-9,23	-6,78	-9,57
Arrêt de croissance	Temps (h)	9±0,3	10,5±0,1*	9±0,2	11±0,1*
	Biomasse (g.L ⁻¹)	1,25±0,02	3,00±0,05***	1,27±0,03	3,50±0,05***
	[Lactose consommé] (mM)	22,48±2	98,34	33,22	143,25
	[Lactate total] (mM)	111,79±5	343	115	488
	[Ac.lactique] (mM)	0,26±0,03	0,99	0,32	1,40
	$Y_{X/lactose}$ (g.g ⁻¹)	0,16±0,1	0,09	0,12	0,08
	$Y_{lactate\ total/lactose}$ (g.g ⁻¹)	0,87±0,03	0,61	0,58	0,60

L'arrêt de croissance sur cette culture ne peut s'expliquer ni par une carence en lactose (52 mM restants), ni en aucun acides aminés libres (concentrations résiduelles supérieures à 1 mg.L⁻¹), ni par une toxicité liée à l'activité acidifiante puisque le pH extérieur est régulé et la concentration en acide lactique non dissociée inférieure à 1 mM. Il est supposé qu'à ce niveau de biomasse produite, un autre composé nutritionnel est potentiellement manquant. Pour vérifier cette hypothèse, en supplément à l'ajout en acides aminés, des vitamines et des bases ont été utilisées pour compléter le milieu. La croissance sur ce milieu permet d'atteindre en 11 h de culture une concentration de 3,50 g.L⁻¹ et ainsi de dépasser le niveau final observé lors de l'ajout seul de la solution d'acides aminés. La concentration résiduelle en lactose est de 1,9 g.L⁻¹ pour 488 mM de lactate produit. Une meilleure efficacité de biosynthèse est également observée avec un rendement en biomasse formée diminué à 0,08 g.g⁻¹ lactose.

- Ainsi, ces essais ont permis d'identifier les acides aminés comme les éléments nutritionnels limitant la croissance de ND5F à 1,27 ± 0,05 g.L⁻¹ sur lait écrémé. Cette information peut être

éventuellement utilisée pour augmenter la production de biomasse en projetant *ND5F* dans une utilisation en culture mixte. En effet, certaines associations avec des souches connues pour être particulièrement protéolytiques pourraient être envisagées pour augmenter la quantité d'acides aminés libres et stimuler ainsi la croissance de *ND5F*.

CONCLUSION

Dans l'objectif de caractériser la souche *ND5F*, cette étude a permis de décrypter le comportement global de la souche sur une matrice industrielle dans une condition de référence et d'analyser l'impact de l'environnement sur la réponse physiologique mise en œuvre au cours de la culture. Cette approche comparative s'est appuyée sur un diagnostic multicritères (croissance, acidification, réduction, consommation d'oxygène) afin d'identifier les principaux leviers et verrous de la croissance de *ND5F*. Ainsi, les conséquences de l'auto-acidification ont été traitées en identifiant les responsabilités des deux facteurs inhibiteurs, l'acide lactique non dissocié et la diminution du pH du milieu, sur les fonctions anaboliques et cataboliques. Néanmoins, les cibles cellulaires, pH_i , activités enzymatiques et concentrations de métabolites intracellulaires, resteraient à étudier pour préciser les aspects mécanistiques du mode d'action du stress acide.

Un effet de la présence initiale d'oxygène sur l'activité acidifiante a été identifié, mais également, plus largement, sur le métabolisme, grâce à l'utilisation combinée de milieu semi-synthétique sous différentes atmosphères gazeuses. Les différents modes de préparation du lait, et notamment la thermisation lors de l'étape de stérilisation, sont apparus comme un levier pour augmenter la fermentescibilité de la matière première en jouant sur la composition en acides aminés libres. L'analyse des exigences nutritionnelles de *ND5F* sur lait, a permis d'identifier une limitation de la croissance par les acides aminés que l'ajout exogène permet de rétablir, ouvrant la possibilité d'association avec des souches protéolytiques.

Identifier les éléments de singularité de la souche dans sa fonction de ferments, impose d'élargir cette caractérisation phénotypique dans le temps et d'étudier le devenir du produit lait fermenté après la phase de post-acidification à 14 jours. Mais il est également nécessaire de confronter les caractéristiques de la souche *L. lactis* ssp. *lactis* *ND5F* à la diversité phénotypique des autres souches industrielles pour déterminer les éléments de cette spécificité. Afin de couvrir l'intégralité des propriétés de ferment de la souche, un volet production de composés organiques volatils est à rajouter à la précédente caractérisation réalisée.

CHAPITRE II :

IDENTIFICATION ET CARACTÉRISATION DE LA
SPÉCIFICITÉ DE *L. LACTIS* SSP. *LACTIS ND5F* EN
LAIT PAR UNE DÉMARCHE INTÉGRATIVE
GÉNOTYPE-PHÉNOTYPE.

INTRODUCTION

Pour identifier les éléments de singularité de la souche *L. lactis* ssp. *lactis* ND5F, une première étape de ce travail s'est focalisée sur l'étude du contenu génétique de ND5F afin d'identifier d'éventuels gènes spécifiques et explorer les liens génotype-phénotype. Ce travail vise à rechercher des marqueurs génétiques des propriétés organoleptiques de ND5F. Cette analyse est supportée par une double approche, avec d'une part un positionnement phylogénétique de la souche (MLST) au sein de la diversité des souches de la sous-espèce *lactis*, et d'autre part une caractérisation génomique (électrophorèse en champs pulsé et séquençage du génome). La seconde étape de cette étude vise à préciser les déterminants phénotypiques de la spécificité fonctionnelle de ND5F. Cette problématique impose, là encore, de mener une analyse comparative des propriétés fonctionnelles de la souche face à la diversité d'autres souches industrielles afin d'identifier les éléments de cette spécificité. Pour couvrir l'intégralité des propriétés de ferments laitiers des souches, il est nécessaire de réaliser cette analyse au plus près des conditions industrielles.

Cette analyse a donc été menée sur une matrice en lait non écrémé avec des taux de matière grasse et de protéines équivalents à ceux du procédé. Une autre nécessité est d'inventorier un jeu de variables le plus large possible pour décrire le plus précisément le comportement global des souches. Une diversité d'indicateurs est donc réclamée en mesurant à la fois des paramètres métaboliques (métabolisme carboné mais également azoté et production d'arômes), physiologiques (croissance, viabilité) et technologiques (durée de fermentation, acidification) pendant la fermentation, mais également dans le temps. Une fois la diversité des levains laitiers explorée par une approche combinant génotype et phénotype, le positionnement de ND5F dans cette diversité et la recherche des marqueurs de sa typicité pourront être investigués.

I Analyse du contenu génétique de ND5F

Une étude de génomique fonctionnelle a été menée pour caractériser le génotype de la souche *Lactococcus lactis* ssp. *lactis* ND5F et tenter d'identifier les marqueurs génétiques des propriétés organoleptiques perçues. Trois parties vont ainsi être traitées dans cette section i) le positionnement phylogénétique de la souche (approche MLST), ii) la caractérisation génotypique de la souche (taille du chromosome, nombre et taille des plasmides, macro-profils de restriction) et iii) la caractérisation génomique (séquençage du génome et identification des éventuels gènes spécifiques pour explorer les liens génotype-phénotype de ND5F). Les résultats de ce travail ont été obtenus dans le cadre d'une collaboration avec le Laboratoire de microbiologie et de génétique moléculaire de l'Université Toulouse III (LMGM).

I.1 Positionnement phylogénétique par MLST de ND5F

Dans un premier temps, l'appartenance de la souche ND5F à la sous-espèce *lactis* a été vérifiée par amplification PCR et séquençage de la région variable V1-V4 de l'ADNr 16S de la souche (environ 800 pb). Cette séquence a ensuite été comparée aux séquences de différentes espèces de firmicutes (bases de données interne du LMGM¹, Toulouse). Cette comparaison a confirmé l'appartenance de la souche ND5F à la sous-espèce *L. lactis* ssp. *lactis*.

Le positionnement phylogénétique de la souche a été ensuite réalisé à partir du schéma MLST décrit par Passerini et coll [Passerini et al. 2010]. Il comprend trois gènes de ménage (codant pour des protéines impliquées dans des fonctions cellulaires de base), *glyA* (serine hydroxyméthyltransferase), *pgk* (phosphoglycérate kinase) et *pdp* (pyrimidine-nucléoside phosphorylase), deux gènes cataboliques, *bcaT* (aminotransférase des acides aminés branchés) et *pepXP* (X-propyl-dipeptidyl aminopeptidase), et un gène de régulation du système SOS, *recN* (ATPase impliquée dans la réparation de l'ADN). L'analyse MLST réalisée sur la souche ND5F permet de lui assigner le profil allélique suivant : *bcaT*(1), *glyA*(1), *pdp*(1), *pepXP*(1), *pgk* (5), *recN* (1), définissant les souches appartenant à la séquence type n°15 (ST15). Sur la base de ce typage et des précédents travaux de caractérisation de la phylogénie de la sous-espèce *lactis* [Passerini et al. 2010] ND5F se positionne comme une souche appartenant à une population homogène de faible diversité génétique (constitution d'un seul et même clade séparé en deux sous-groupes, G1 et G2) (Figure II. 1). Cette classe regroupe des souches isolées de produits fermentés ou de levains, et a été nommée par le laboratoire LMGM souches « domestiquées ». Issu d'un événement de diversification récent, ce groupe s'oppose aux souches « environnementales » isolées de plantes, d'animaux et de lait cru. La

¹<http://www.mlst.biotoul.fr/Lactococcuslactissubsp.lactis/16SV1V4firmPCR.fas>

faible diversité génétique des souches « domestiquées » s'illustre notamment par leur regroupement en deux complexes clonaux par analyse e-burst (Based Upon Related Sequences Types). Un complexe clonal rassemble les STs de souches qui partagent cinq des six loci. Il est à noter que le ST15, auquel appartient *ND5F*, est identifié comme le ST ancestral du complexe clonal 1 (CC1), c'est-à-dire que ce ST ne diffère que d'un allèle avec le plus grand nombre de ST du complexe clonal. De plus, le ST15 concentre, à ce jour, le plus grand nombre de souches caractérisées (communication personnelle, P. Le Bourgeois en 2013).

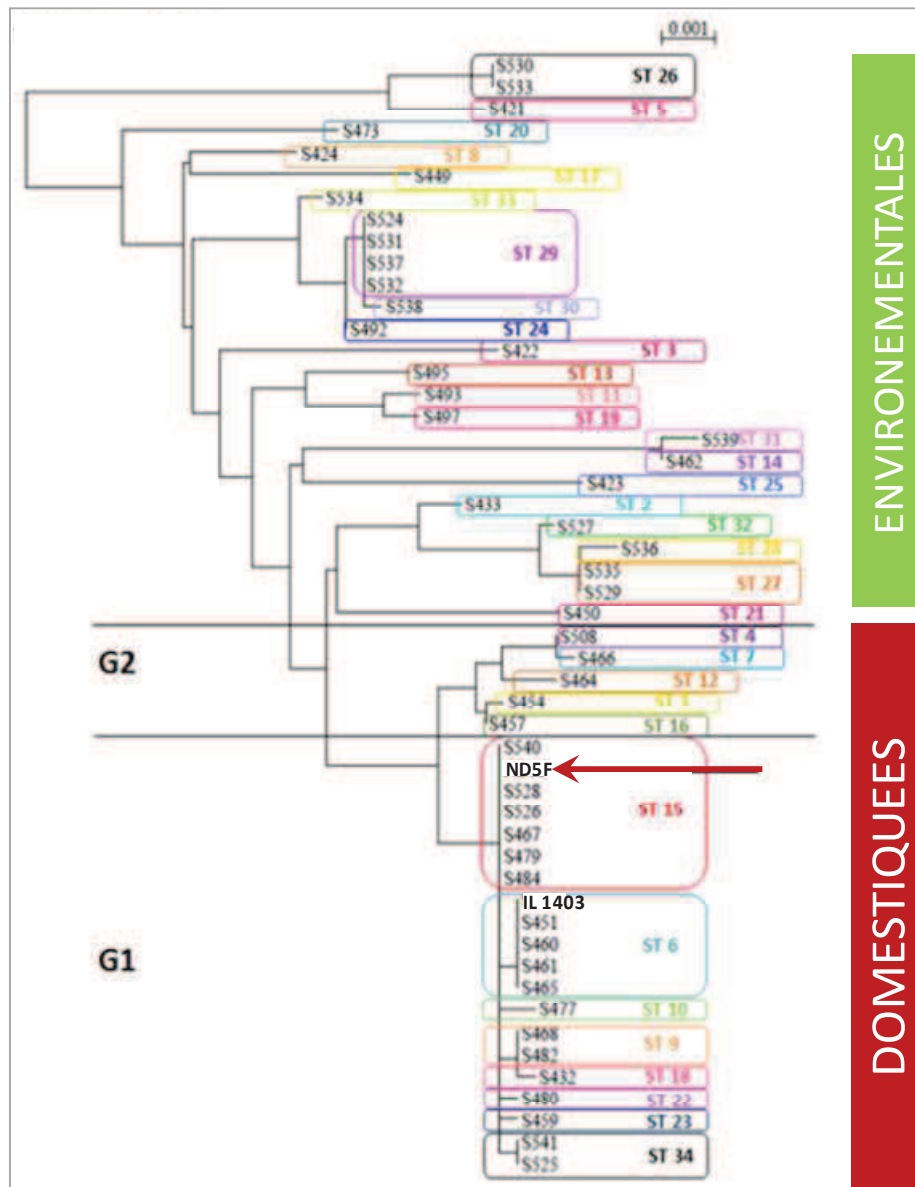


Figure II. 1 : Positionnement de la souche *ND5F* sur l'arbre phylogénétique des souches de *L. lactis* ssp. *lactis* de la collection Genoferment [Passerini et al. 2010]. La flèche noire indique la position de la souche *ND5F*. G1 et G2, les deux groupes déterminés au sein du clade correspondant à la population de souches dites « domestiquées ». Arbre phylogénétique réalisé selon la méthode Neighbor Joining (bootstrap 1000).

I.2 Evaluation génomique de ND5F

La taille du génome de ND5F a été évaluée à partir de l’analyse des profils de restriction par PFGE (Pulsed Field Gel Electrophoresis, ou Electrophorèse en champs pulsés). La PFGE est reconnue comme une méthode de référence pour établir des empreintes génomiques de souches en raison de son pouvoir résolutif. En effet, elle permet de séparer des fragments d’ADN de moins de 1 kb à une dizaine de mégabases. Elle a également été utilisée pour réaliser une comparaison des profils génomiques entre souches. Ainsi, la souche ND5F présente une empreinte génomique singulière par rapport aux souches *L. lactis* ssp. *lactis* de la collection Genoferment précédemment caractérisée [Passerini et al. 2010]. En parallèle, la recherche de marqueurs d’intérêt laitier (dégradation de la caséine, internalisation du lactose et utilisation du citrate) a été conduite par typage moléculaire des régions géniques concernées par PCR.

I.2.1 Taille du génome par analyse des macro-profils de restriction

Après une double digestion de l’ADN de la souche ND5F par les endonucléases *NotI*/*ICeuI* suivie d’une PFGE, la taille du chromosome de la souche ND5F a été estimée à environ 2550 kb (plus ou moins 5%) en effectuant la somme des tailles des fragments de restriction obtenus (Figure II. 2 et Tableau II.1). Il est noté qu’à ce jour, les tailles des chromosomes des souches appartenant au sous-groupe *L. lactis* ssp. *lactis* varient de 2304±41 à 2725±72 kb (source NCBI, Genome Database, 2013).

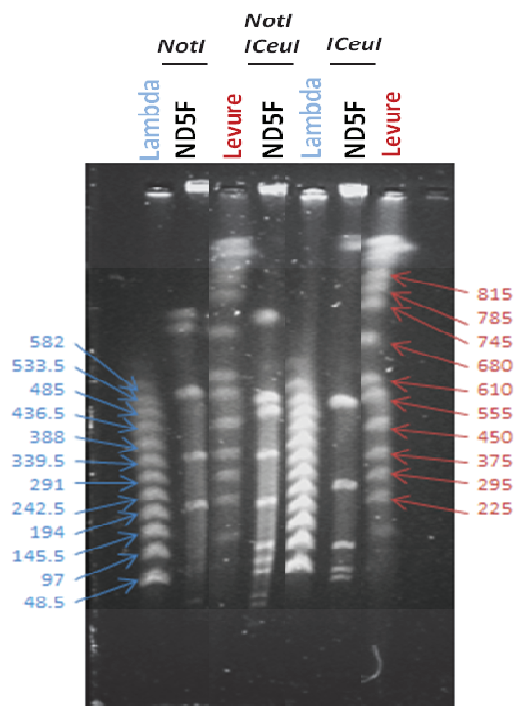


Tableau II. 1 : Estimation de la taille des fragments de ND5F obtenus après digestion par *NotI*-*NotI*/*ICeuI* et *ICeuI*. X* : Taille du fragment indéterminée sur le gel de PFGE. En gras, la taille totale en kb.

<i>NotI</i>	<i>NotI/ICeuI</i>	<i>ICeuI</i>
		X*
710	710	
690		
550	550	550
	490	
360	360	
		260
205	205	
	90	90
	50	
	40	40
		35
	25	
20	20	
	10	

2535	2550	975+X

Figure II. 2 : PFGE de l'ADN de ND5F obtenue après digestion par NotI, NotI/ICeuI et ICeuI. Taille en kb. → dans un premier temps après digestion de l'ADN plasmidique par la Nucléase S1 suivie d'une PFGE. Quatre plasmides de 80, 50, 7-10 et 1,5-3 Kb ont été dénombrés d'après les estimations des gels (résultats non présentés). Ces estimations seront à vérifier sur les données de séquençage.

I.2.2 Recherche de gènes d'intérêt laitier

Même si l'analyse phénotypique de la souche *ND5F* révèle son aptitude à coaguler le lait en moins de 24 h, la présence des deux principaux marqueurs génétiques d'une croissance optimale en culture pure sur lait a été vérifiée puisque toutes les souches domestiquées ne les possèdent pas. Ainsi, le gène *lacE*, codant l'enzyme II spécifique de l'internalisation du lactose au sein du complexe PTS a été ciblée par amplification PCR, tout comme le gène *prtP*, codant la protéase de paroi, responsable de la dégradation de la caséine. Tous deux ont été identifiés sur le génome de *ND5F* (résultats non présentés).

La capacité de *ND5F* à métaboliser le citrate a été évaluée par un test de typage moléculaire de l'opéron chromosomique citrate (région *citM-citG*) codant la citrate lyase et l'oxaloacétate décarboxylase ainsi que du gène plasmidique *citP* codant la citrate perméase. L'absence d'amplification de ces deux cibles pour la souche *ND5F* révèle l'impossibilité de la souche à utiliser le citrate du lait (Figure II. 3).

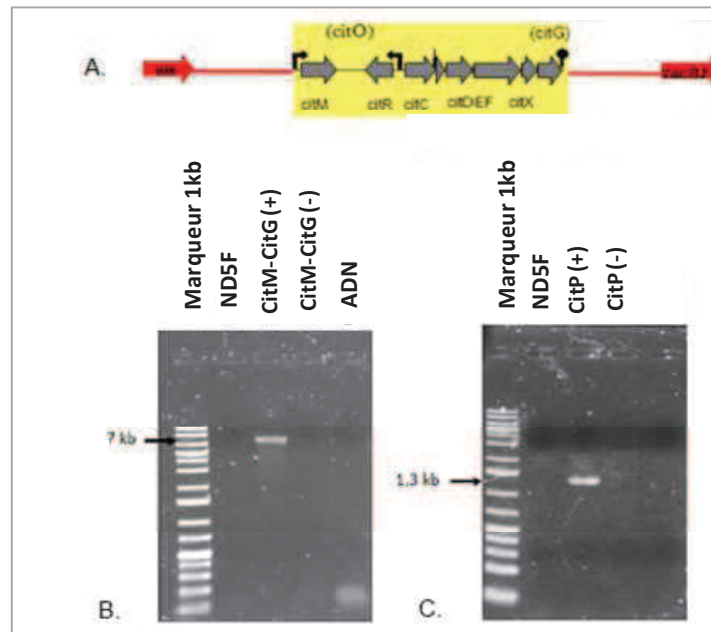


Figure II. 3 : Recherche de l'opéron chromosomique et du plasmide impliqués dans l'utilisation du citrate chez la souche ND5F. A. Schéma des gènes présents dans l'opéron chromosomique. B. Amplification de l'opéron chromosomique (région citM-citG). C. Amplification du gène plasmidique citP.

I.3 Caractérisation génomique de ND5F

Le séquençage, étape préliminaire à la caractérisation génomique, a été réalisé à haut débit sur le génome de la souche ND5F selon deux techniques : « Shotgun » (séquençage aléatoire) et « Long Paired Ends » (détermination de la séquence nucléotidique des extrémités de fragments d'ADN de 3 kb). L'assemblage des fragments et l'annotation structurale des fonctions codantes ont été conduites à partir du logiciel de l'INRA Agmial (Annotation de Génomes Microbiens d'Intérêt Agroalimentaire), et du programme RAST (Rapid Annotation using Subsystem Technology) afin de déterminer les séquences codantes (CDS) et leur orientation. La taille du chromosome de ND5F a pu être affinée à 2 460 576 pb, correspondant à 2649 gènes ou séquences codantes identifiées (CDS) par prédiction bioinformatique. Les quatre plasmides prédits par PFGE ont pu être virtuellement recircularisés après vérification des séquences plasmidiques par PCR, et représentent un total de 165 gènes prédits. Ils sont de respectivement 78 116 pb, 50 057 pb, 7652 pb et 2120 pb.

Une autre étape de la caractérisation génomique de ND5F consiste à réaliser une analyse «différentielle» des gènes de cet organisme avec des génomes modèles voisins permettant l'identification des fonctions absentes ou spécifiques du génome de ND5F. Cette étape inhérente à l'analyse du pan-génome de la sous-espèce *lactis*, permet de déterminer le génome cœur, regroupant les gènes présents dans toutes les souches de la même espèce, et le génome accessoire, qui contient les gènes spécifiques d'une souche ou d'une lignée. Afin de déterminer les gènes

spécifiques de *ND5F*, des comparaisons des contenus génétiques ont été conduites entre cette souche et les autres génomes de *L. lactis* disponibles. De plus, pour chaque gène spécifique de *ND5F*, une annotation fonctionnelle manuelle a été réalisée sur le logiciel BLASTP 2.2.28⁺[Altschul et al. 2005]. L'objectif est d'identifier ainsi sur ces gènes des motifs conservés permettant éventuellement de supposer des fonctions putatives de ces gènes spécifiques.

I.3.1 Comparaison aux souches de la sous-espèce *lactis* appartenant à l'écotype domestiqué

Dans un premier temps, une comparaison des gènes chromosomiques de *ND5F* a été conduite avec le génome de la souche laitière de référence de la sous-espèce *lactis*, *IL1403*. Cette analyse a permis d'attribuer un total de 217 gènes spécifiques, soit 184 kb, à notre souche d'intérêt. Parmi ces gènes, près de la moitié (117) sont des gènes présents sur des prophages ou des ICEs (Integrative Conjugative Elements, ou éléments génétiques mobiles).

Cinquante-sept gènes spécifiques codent des fonctions encore inconnues à ce jour (double annotation par les logiciels RAST et Agmial sans résultat).

Dix-neuf gènes attribués à des fonctions métaboliques, et donc susceptibles d'avoir un intérêt dans notre étude, ont été identifiés. Ils sont décrits en détail dans le Tableau II. 2. Ils appartiennent pour la majorité à des fonctions liées à la synthèse ou à l'hydrolyse de la paroi cellulaire. La présence de ces gènes spécifiques pourrait indiquer un potentiel de lyse cellulaire plus importante de *ND5F* par rapport à la souche modèle *IL 1403*. Si une activité d'autolyse est quantifiée lors de la caractérisation de *ND5F* en lait fermenté, ces gènes peuvent représenter des marqueurs potentiels pour expliquer la spécificité organoleptique de la souche du fait d'une libération du contenu cellulaire dans la matrice.

En parallèle, quinze transporteurs spécifiques ont été identifiés chez *ND5F*. Six gènes (*kdpA*, *kdpB*, *kdpC* dupliqué, *kdpD* et *kdpE*) participent à la transduction du signal d'une limitation en potassium. Le couple de protéines *KdpD* et *KdpE* fonctionne respectivement comme le senseur (canal K⁺ histidine kinase) et le régulateur transcriptionnel (liaison à l'ADN) de ce système à deux composants. Les protéines *KspA-C* forment les différentes sous-unités de l'ATPase en charge de la translocation du potassium [Blattner et al. 1997]. Le gène *nhaP*, également impliqué dans la régulation de la concentration en ions, a été précédemment identifié chez *L. lactis* ssp. *lactis* *KF147* comme un antiport Na⁺/K⁺ et K⁺/H⁺ [Siezen et al. 2010]. Trois gènes, *rgpD*, *ycbk* et un gène non nommé, ont été identifiés comme des transporteurs d'oligosaccharides. Dans le cas du gène *rgpD*, le produit correspond à une protéine en charge de la translocation d'exopolysaccharides synthétisés [Siezen et al. 2010]. Quatre autres gènes ont été également identifiés comme des ABC-transporteurs de sucres par homologie de séquences chez *L. lactis* ssp. *lactis* *CV56* [Gao et al. 2011]. Enfin, sept gènes

spécifiques à *ND5F* codant des « Cold-shock protein » ou « Rep family protein » ont pu être déterminés, tout comme deux gènes impliqués dans la résistance aux phages (*AbiH* et *Abi-like* gènes).

Il est à signaler également que la souche *IL 1403* possède, quant à elle, 142 gènes, soit 116 kb, absents du chromosome de *ND5F*. Ils se partagent principalement en gènes impliqués dans des fonctions liées aux prophages (fonctions souvent souches-dépendantes), au métabolisme du citrate (absent chez *ND5F*), ainsi qu'à des enzymes de restrictions ou à des fonctions inconnues.

Tableau II. 2: Récapitulatif des dix-neuf gènes chromosomiques attribués à des fonctions métaboliques spécifiques à ND5F après comparaison avec la souche IL1403.

- : non déterminé

Taille de la séquence codante(CDS)	Nom de gène si identifié	Annotation fonctionnelle	Fonction métabolique
695	-	Régulateur transcriptionnel HTH-type, famille Rgg/GadR	Régulateur de la transcription
140	-	Glycosyl transférase, similaire à la protéine d'adhésion intercellulaire A (IcaA)	Formation de biofilm
977	-	CDP-glycérol glycérophosphotransférase, TagF domain protéine (EC: 2.7.8.12)	Biosynthèse de la paroi
980	-	Beta-1,3-glycosyltransférase	Biosynthèse de la paroi
962	-	Beta-1,3-glycosyltransférase	Biosynthèse de la paroi
2111	-	N-acétylmuramoyl-L-alanine amidase (EC 3.5.1.28)	Hydrolyse de la paroi cellulaire
785	-	Glycosyl transférase, famille 2	
734	<i>ycbB</i>	Glycosyl transférase, famille 2	Biosynthèse de la paroi
818	<i>LicD</i>	Lipopolysaccharide cholinephosphotransférase LicD1 (EC 2.7.8.-)	Biosynthèse des lipides de la paroi
3020		Protéine de biosynthèse du lipopolysaccharide	Biosynthèse des lipides de la paroi
632		4-hydroxybenzoate polyprényltransférase et prényltransférases	-
710		α/β -hydrolase superfamille protéine	Biosynthèse de la paroi
248	<i>xylS</i>	Maltodextrine glucosidase (EC 3.2.1.20)	-
1250		Arylsulfatase régulateur	-
422	<i>tagD</i>	Glycérol-3-phosphate cytidylyltransférase (EC 2.7.7.39)	-
2072	<i>yccB</i>	N-acétylmuramoyl-L-alanine amidase (EC 3.5.1.28)	Hydrolyse de la paroi, autolyse
836	<i>epaC</i>	Glycosyltransférase, protéine du groupe 2	-
1247		Glycosyltransférase, putative	-
1043		Hydrolyse du peptidoglycane	Hydrolyse de la paroi

Lorsque l'on ajoute à la comparaison les génomes des deux autres souches précédemment identifiées comme domestiquées (*CV56* et *CNCMI-1631*) [Passerini et al. 2010], seuls 70 gènes demeurent spécifiques à *ND5F* (correspondant à 53 kb). Ce nombre représente une diminution de 70 % des gènes précédemment identifiés par comparaison unique avec la souche de référence *IL1403*. Le détail du nombre de gènes spécifiques à *ND5F* après comparaison avec les autres souches domestiquées est présenté Tableau II. 3. Cependant, l'analyse fonctionnelle de ces 70 gènes révèle qu'aucun d'entre eux n'appartient à la catégorie hydrolyse de la paroi, ni même à des fonctions métaboliques ou encore à des transporteurs.

Tableau II. 3 : Nombre de gènes spécifiques à *ND5F* après comparaison avec les autres souches domestiquées.

Catégorie	<i>IL 1403</i>		<i>Souches domestiquées (3 souches)</i>	
	Nombre de gènes	Taille des CDS (pb)	Nombre de gènes	Taille des CDS (pb)
Métabolisme	19	19652	-	
Transporteurs	15	20925	-	
Autres fonctions	6	2457	2	1912
Résistance aux phages	2	1984	1	944
Prophages et ICEs	118	78962	25	19500
Fonction inconnue	57	26724	42	16914
Total	217		70	

Par conséquent, ce résultat compromet l'identification potentielle de marqueurs génétiques spécifiques pour établir un lien entre présence de gène et phénotype pour *ND5F*.

Au contraire, cette analyse confirme à nouveau les précédents résultats obtenus sur la faible diversité génomique des souches domestiquées ou laitières de *L. lactis* ssp. *lactis* qui partagent un large génome cœur.

I.3.1 Comparaison aux autres souches de l'espèce *lactis*

La comparaison du contenu chromosomique de *ND5F* aux deux autres souches de la sous-espèce *lactis* provenant d'isolats environnementaux (*IO-I* et *KF147*) a réduit la taille des régions spécifiques de la souche *ND5F* à 46 668 pb. L'ajout des génomes des souches isolées d'environnement laitier de

la sous-espèce *cremoris* (MG1363, SK11, NZ9000 et A76)², a permis d'identifier une liste de 48 séquences codantes spécifiques à ND5F par rapport à ces neuf souches testées de l'espèce *L. lactis*.

L'intégralité de ces 48 gènes ont été soumis à une analyse fonctionnelle (logiciel BLASTP 2.2.28⁺) [Altschul et al. 2005] définie par la recherche d'homologie de séquences afin d'identifier d'éventuels rapprochements avec des gènes identifiés sur d'autres espèces, ou encore des domaines conservés (seuil de significativité à E-value < 1.10^{-20}). Cette recherche a été conduite préférentiellement sur les séquences protéiques pour s'affranchir des éventuelles pertes d'homologies liées au biais de codon et augmenter les probabilités de prédiction des fonctions métaboliques.

La majorité de ces gènes spécifiques (23) ont été identifiés comme appartenant à des fonctions liées aux prophages ainsi qu'aux séquences d'insertions. Trois gènes peuvent-être potentiellement identifiés comme ayant une fonction putative de résistances à ces phages. En effet, deux de ces séquences présentent des homologies avec le domaine dit AbiH (Bacteriophage abortive infection, E-value < 1.10^{-29}), et la dernière avec le domaine C-Terminal du répresseur XRE des bactériophages chez *L. lactis* (E-value < $2,29.10^{-10}$). Les fonctions codées par trois autres séquences codantes ont été également proposées sur la base de la comparaison de leur séquence protéique avec d'autres espèces au sein du groupe des bactéries lactiques ou d'espèces taxonomiquement proches. Un gène semblerait coder une protéine homologue à la protéine FtsK présente notamment chez *L. garvieae* (100 % de couverture à 97 % d'identité et une E-value de 1.10^{-56}). Chez *E. coli* K-12, des études ont démontré le rôle de cette protéine dans la division cellulaire [G. Liu et al. 1998; Wang & Lutkenhaus 1998] (93 % de couverture à 95 % d'identité et une E-value de 7.10^{-53}). Une homologie d'une séquence codante pour la protéine chaperonne ClpB chez *Lactobacillus casei* (99 % de couverture à 95 % d'identité et une E-value < 1.10^{-167}) a été également identifiée. Cette protéine interagit avec la chaperonne DnaK et est activée lors des réponses au stress osmotique (accumulation de glutamate ou glycine-bétaine) ou thermique [Diamant et al. 2003; Mogk et al. 2003] chez *E. coli* K-12. Le dernier gène spécifique de ND5F semblerait être un régulateur transcriptionnel retrouvé chez *L. garvieae* (100 % de couverture à 98 % d'identité et une E-value de 3.10^{-38}) avec la présence d'un domaine de liaison à l'ADN (E-value de 5.10^{-27}). Les 17 autres séquences codantes demeurent de fonction inconnue en l'absence de domaines conservés identifiés. A partir de ces données génomiques, aucun

² Etude réalisée en Mai 2012

gène ne peut être utilisé comme marqueur génétique pour prédire l'origine de la spécificité organoleptique de *ND5F*. Seuls les gènes putatifs de résistance aux phages pourraient être intéressants s'il est avéré que leur expression confère à *ND5F* un profil de résistance phagique singulier (à valider expérimentalement).

De plus, ces 48 gènes spécifiques se localisent tous au sein de prophages ou d'ICEs (voir Figure II. 4). Il s'agit donc de gènes acquis par des transferts horizontaux de matériel génétique dans ces régions. Il est à noter que ces gènes sont donc potentiellement mobiles. En particulier, il a été identifié chez *ND5F* trois régions ICEs et huit prophages, parmi lesquels cinq seraient, d'après leur taille, prédits comme codants des enzymes effectives suffisantes pour le déclenchement d'une phase lytique.

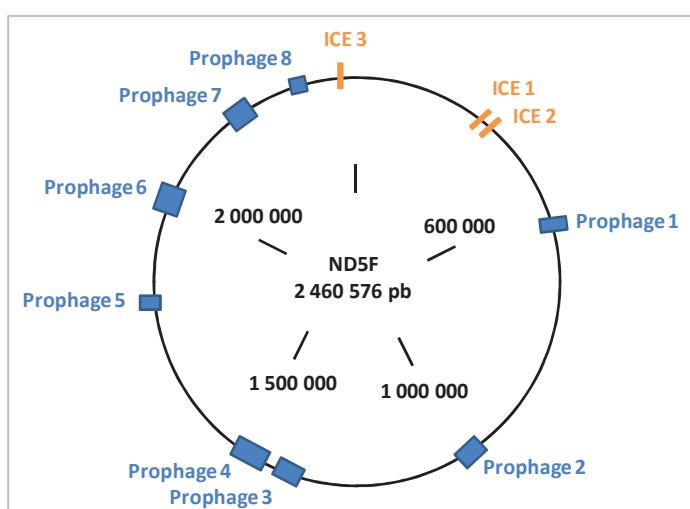


Figure II. 4 : Position des huit prophages et trois ICEs contenant les gènes spécifiques de *ND5F* sur le chromosome.

En l'absence de marqueur génétique prédictif des potentialités organoleptiques de *ND5F*, il apparaît indispensable d'identifier les déterminants technologiques et métaboliques de la singularité organoleptique perçue. Pour répondre à cette problématique, une analyse phénotypique multicritères de la souche dans sa fonction de ferment laitier est nécessaire afin de confronter la signature de *ND5F* à la diversité fonctionnelle de souches dans leur fonction de levains laitiers.

II Etude de la diversité de levains par une approche intégrative génotype-phénotype en lait

Dans cette section, les résultats sont présentés en anglais sous forme d'un article scientifique publié dans le journal *Applied and Environmental Microbiology* (Août 2013, volume 19, numéro 15). Un résumé détaillé des principaux résultats précède la publication.

II.1 Résumé article

Dans cette publication, nous nous sommes intéressés à caractériser la diversité fonctionnelle de levains laitiers de *Lactococcus lactis* ssp. *lactis* et éventuellement prédire leurs potentialités en conduisant une approche intégrative génotype-phénotype. Ce travail a débuté par la construction d'un échantillon de souches représentatif de la diversité génétique de cette sous-population et compatible avec la sélection de souches performantes dans les procédés laitiers. Pour cela, nous avons présélectionné, auprès de collections nationales et industrielles, neuf souches isolées de levains ou de lait cru possédant les marqueurs génétiques *lacE* et *prtP* pour garantir un phénotype laitier (habilité à croître en lait en culture pure et à coaguler efficacement le lait à 24 h). Quatre souches de l'échantillon ont été assignées au biovar. *diacetylactis* sur la base de la présence du gène plasmidique *citP* codant la citrate perméase requise pour l'internalisation du citrate. L'analyse par MultiLocus Sequence Typing (MLST) a vérifié que l'intégralité des souches étaient des souches domestiquées. De plus, huit des neuf souches de cet échantillon ont été identifiées comme appartenant au CC1 (se répartissant en quatre souches au ST15, deux au ST34, une au ST10 et une au ST10) et la dernière au CC2 avec le ST7. A la vue de ces résultats, l'échantillon est considéré comme représentatif de la diversité génétique des souches de l'écotype « domestiquées ». Les macroprofils de restriction obtenus par électrophorèse en champs pulsés (PFGE) ont permis de prouver que les souches du même ST étaient différentes.

La proximité génomique des souches a été évaluée par l'estimation du coefficient de Dice (S_D) pour comparer deux à deux les empreintes génomiques des souches. L'arbre consensus (bootstrap > 0.50) obtenu par le clustering des distances PFGE de chaque souche définit trois principaux groupes de souches dont un uniquement composé des souches du biovar. *diacetylactis*. Cependant, les souches du même ST ou encore du même complexe clonal ne sont pas regroupées ensemble sur ces données génomiques. Il a donc été proposé la première classification génotypique intégrative, basée à la fois sur la diversité génétique (MLST) et génomique (PFGE) des souches, afin d'améliorer le classement des potentialités génotypiques des souches domestiquées. La classification intégrée obtenue est

cohérente avec la discrimination des souches du biovar. *diacetylactis*, reflète la phylogénie en deux complexes clonaux et conserve la singularité génotypique de chaque souche.

La diversité phénotypique de cette sous-population d'intérêt a été évaluée lors de fermentations (6 réplicats par souches) sur un lait enrichi en matière protéique (35 g/kg) et grasse (40 g/kg) afin de mimer les conditions industrielles d'utilisation de ces levains. 82 variables ont été sélectionnées comme des marqueurs importants pour décrire la signature fonctionnelle de ces souches en lait. Ces descripteurs ont été mesurés en dynamique au cours de la maturation du lait ainsi qu'à 14 jours au sein des produits conservés à 4°C (demi-vie du lait fermenté) pour illustrer l'évolution du comportement de chaque souche. Ces critères comprennent des indicateurs physiologiques (taux de croissance, vitesse maximale d'acidification), technologiques (temps avant d'atteindre le pH d'arrêt de 4,60, viabilité) et métaboliques (concentrations exocellulaires en sucres, acides organiques, acides aminés et 47 Composés Organiques Volatiles (COVs), déterminants essentiels de la saveur du produit).

La recherche de corrélations entre les différents jeux de données a permis de tendre vers la reconstitution de voies métaboliques et a démontré la pertinence de la combinaison de ces variables. L'Analyse en Composante Principale (ACP) de ces 82 marqueurs a mis en évidence la forte diversité phénotypique au sein de l'échantillon puisque neuf signatures ont été observées, soit une signature par souche (ellipse à 95 % de confiance). Par conséquent, ce résultat confirme la pertinence de ce jeu de variables pour établir des signatures laitières des souches et prédire leur éventuelle spécificité. De plus, ce haut degré de divergence entre les souches autorise une identification précise des souches en lait fermenté (bootstrap > 0,94). Ainsi, malgré une variabilité génétique limitée d'après les études de phylogénie, une importante diversité phénotypique est présente au sein des souches domestiquées permettant d'envisager des applications spécifiques : i) production intense d'acide lactique pour assurer la bio préservation des produits, ii) synthèse du neurotransmetteur acide γ -aminobutyrique pour un produit à la fonction santé, ou encore iii) biosynthèse de certains COVs à la saveur particulière.

Pour réduire le temps et le coût de ces analyses phénotypiques fastidieuses, une sélection des marqueurs les plus appropriés pour discriminer les neuf souches a été menée (25 % des variables les plus contributives). Vingt variables ont été identifiées, toutes appartenant aux composés organiques volatiles. Ainsi, ces COVs concentrent la variabilité inter-souches plus que les paramètres de croissance, d'acidification ou encore de production de lactate. Ces COVs s'organisent en trois groupes correspondant aux trois voies métaboliques majeures de la production d'arômes : glycolyse (4 variables), protéolyse (5 variables) et lipolyse (11 variables). Alors que les deux premières ont été

étudiées en détail dans le lait [Smit et al. 2005a; Smit et al. 2008; Cheng 2010], la lipolyse a été souvent sous-estimée dans les travaux sur matrice en lait écrémé. Sur ces 20 variables, les neuf souches laitières se structurent en quatre sous-groupes robustes (bootstrap 0,83) de souches aux orientations métaboliques divergentes, déduites de l'interprétation biologique des composantes principales (CP) : CP1 lipolyse, CP2 glycolyse et CP3 protéolyse. Ces 20 COVs combinés aux avancées en GC-MS sur matrice lait [Smit et al. 2004; Bachmann et al. 2009a] ouvrent la voie au criblage haut débit et prédictif des potentialités aromatiques de collections de levains en matrice lait. L'addition de nouvelles souches pourra permettre également de préciser, voire modifier, les groupes de souches obtenus. De plus, ces composés aromatiques peuvent conduire à une réflexion pour sélectionner des marqueurs génétiques prometteurs de ces capacités aromatiques.

Il est à noter qu'une souche, *DIA-A*, constitue à elle seule un sous-groupe, démontrant ainsi son fort degré de typicité aromatique. Autre remarque, les souches du biovar. *diacetylactis* ne sont pas regroupées dans un sous-groupe unique. Cette observation atteste que le niveau de production de diacétyle/acétoïne n'est pas pertinent pour discriminer la signature de ces souches et qu'il existe une diversité métabolique intragroupe plus importante.

L'inadéquation des deux classifications génotypique et phénotypique confirme, comme précédemment démontré, le haut degré de régulation de l'expression du génome ainsi que du protéome [Bachmann et al. 2009b; Tan-a-ram et al. 2011; Kelly et al. 2010]. Pour tenir compte des potentialités génétiques codées au sein du génotype ainsi que de cette caractérisation phénotypique exhaustive, nous avons proposé une classification intégrée génotype-phénotype des souches. Il a été observé que l'addition du génotype conduit à cinq sous-groupes au lieu de quatre, et permet donc une classification plus résolutive de la diversité des souches. En particulier, le précédent sous-groupe de souches non dédiées à une voie métabolique particulière a été profondément modifié avec la prise en compte de ces données génotypiques, et en particulier, la seule souche du CC2 a été singularisée. Ce résultat met en lumière les divergences génétiques de cette souche, et permet de considérer les potentielles modifications évolutives de cette souche phylogénétiquement plus éloignée. Cependant, l'absence de variables génotypiques sélectionnées parmi les variables les plus contributives sur ce nouveau set de données relativise l'importance du génotype par rapport aux larges différences phénotypiques observées sur les souches domestiquées. Cette approche intégrée génotype-phénotype tend donc à proposer aux utilisateurs de ferments une classification prédictive de la fonctionnalité des souches dans des procédés laitiers.

II.2 Publication

III Identification et caractérisation de la signature spécifique de *ND5F*

L'approche intégrative génotype-phénotype développée pour la caractérisation de la diversité fonctionnelle des levains laitiers en laits fermentés a été appliquée au positionnement de *ND5F*, pour évaluer sa spécificité et identifier les déterminants de cette signature. Les termes singularité et typicité ont été introduits pour décrire la notion de spécificité d'une souche. La singularité fait référence à la capacité d'une souche à présenter une signature unique. La typicité d'une souche est définie comme un ensemble de caractéristiques qui confèrent à une souche l'aptitude à former un sous-groupe distinct des autres souches.

III.1 Positionnement génotypique de *ND5F*

Dans un premier temps, la proximité de *ND5F* avec l'échantillon de souches sélectionnées a été évaluée séparément par les deux approches, génétique (MLST) et génomique (PFGE). Logiquement, lors de l'étude de phylogénie, *ND5F* se regroupe en première intention avec les quatre autres souches appartenant au ST15 (*DIA-A*, *S87*, *EIP33A* et *EIP41A*), puis avec celles du Complexe Clonal 1 (résultat non présenté). L'analyse des empreintes génomiques révèle que la souche *ND5F* présente des valeurs de coefficients de similarité inférieures à 0,60 (Dice coefficient, $S_D < 0,60$), une valeur typique de la comparaison de souches sans lien direct de parenté.

L'addition de *ND5F* à la classification génotypique intégrative modifie légèrement la précédente hiérarchisation observée sur les neuf souches domestiquées (publication vs Figure II. 5). En effet, les trois principaux groupes de souches présentent une permutation suite au regroupement de *ND5F* avec la souche *EIP33A*, qui contribue à rapprocher cette dernière du duo de souches *UCMA5712* et *S86*. Cependant, l'appartenance des souches à un complexe clonal est conservée comme le premier élément de structuration de cette diversité génotypique. Ainsi, à la différence de l'analyse de génomique fonctionnelle (partie I), *ND5F* apparaît comme une souche singulière sur cette classification génotypique. Cependant, aucun marqueur de cette spécificité génétique ne peut-être considéré comme prédictif de la fonctionnalité de la souche à partir des données de phylogénie et des profils de macro-restriction.

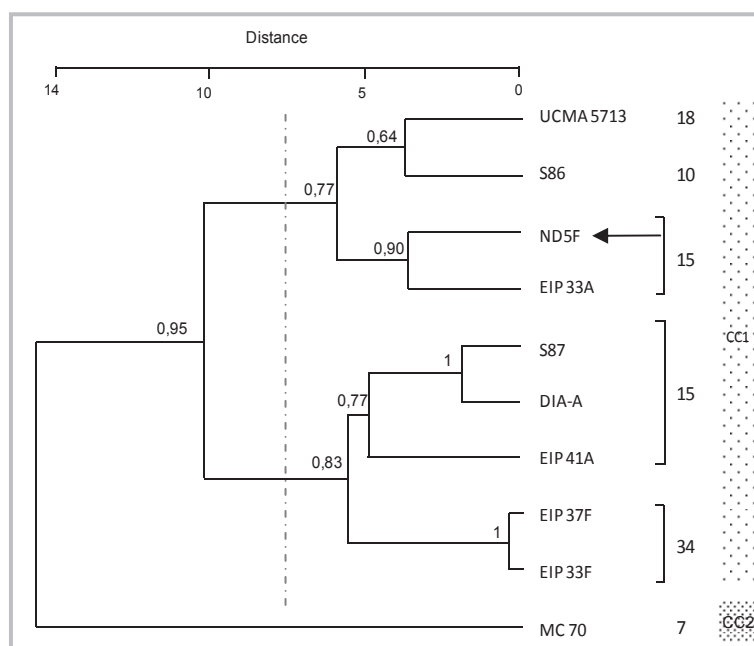


Figure II. 5 : Classification ascendante hiérarchique (CAH) génotypique intégrative de ND5F (flèche noire) et des neuf autres souches domestiquées, à partir des données de phylogénie génétique (MLST) et génomique (PFGE). De gauche à droite : le dendrogramme, le nom des souches, le Sequence Type (ST) et le Complexe Clonal (CC). Les valeurs de bootstrap au dessus de 0,50 sont indiquées pour chaque nœud. La ligne verticale en pointillé matérialise le niveau de coupure de l'arbre pour interprétation.

III.2 Singularité de la signature fonctionnelle de ND5F

Pour caractériser la signature fonctionnelle de ND5F, des essais en lait fermenté sur la même matrice (matière protéique et grasse de 35 et 40 g.kg⁻¹) ont été conduits, et les 82 variables phénotypiques ont été mesurées. Après le traitement des données, une analyse en composante principale (ACP) ainsi qu'une classification ascendante hiérarchique (CAH) ont été réalisées pour positionner la signature de ND5F (Figure II. 6). Les résultats de l'ACP présentés sont limités aux trois premières composantes principales (CP), puisque la variance cumulée atteint 66,9 %, un pourcentage jugé suffisamment haut pour assurer que ces deux graphiques présentent les caractéristiques principales de nos données. Il est à noter que les représentations graphiques d'ACP et de CAH obtenues après l'ajout de ND5F, ainsi que le jeu des 20 variables les plus contributives, sont similaires à ceux de la publication. Par conséquent, ce résultat permet de confirmer la robustesse, face à l'ajout d'autres souches, de la caractérisation fonctionnelle proposée pour les levains laitiers de *L. lactis* ssp. *lactis*.

La présence d'une signature singulière de ND5F est examinée à partir des représentations graphiques d'ACP des 10 souches sur les composantes 1 et 2, ainsi que 1 et 3 (Figure II. 6A et B). Le critère utilisé

pour définir si une souche possède une signature singulière est l'absence de recouvrement de l'intervalle de confiance de la souche avec celui d'une autre souche.

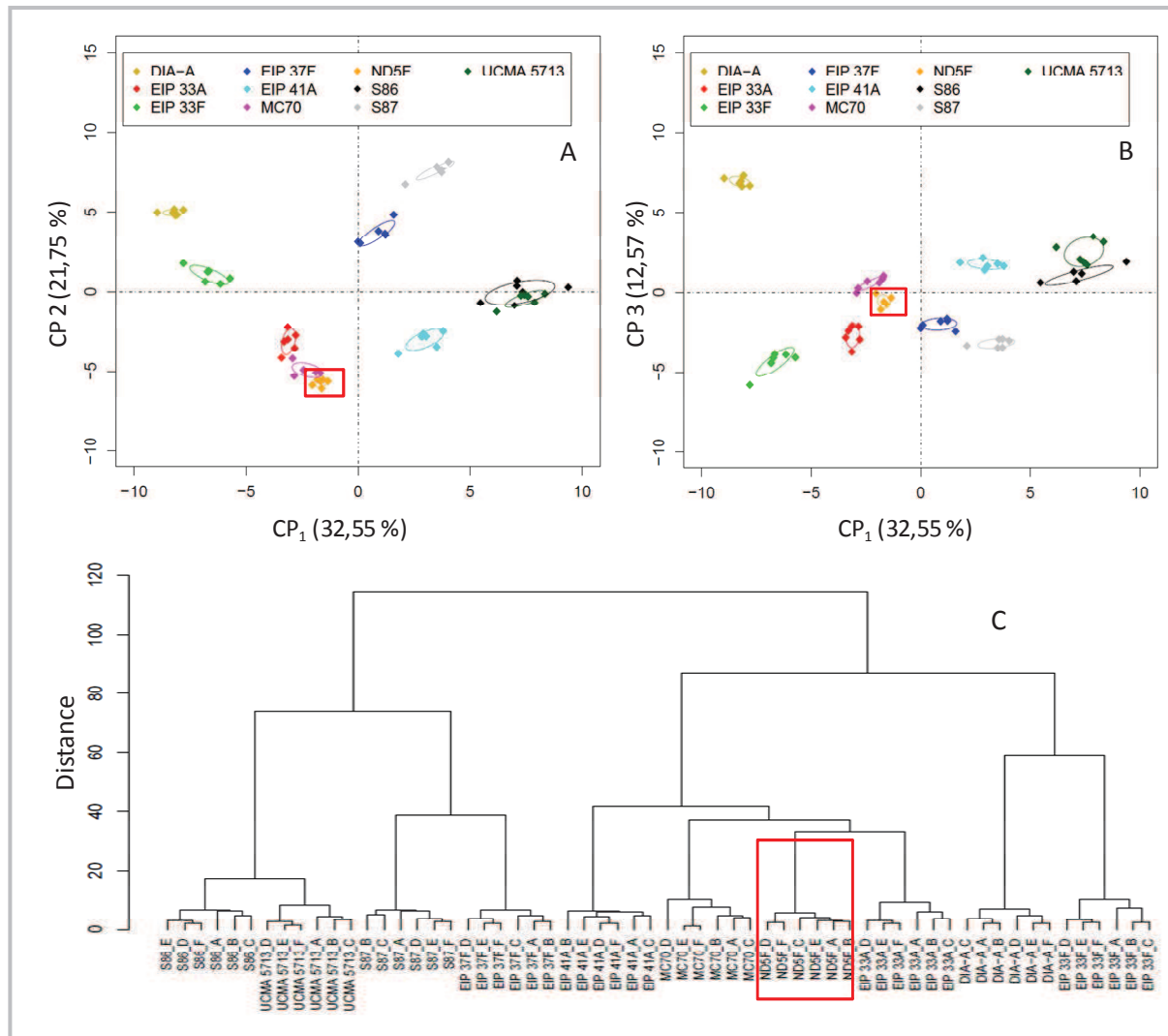


Figure II. 6 : Analyse en composantes principales (ACP) et classification ascendante hiérarchique (CAH) de l'échantillon de neuf souches domestiquées de *L. lactis* ssp. *lactis* ainsi que de la souche ND5F à partir des 82 données phénotypiques. Représentation des 10 souches selon les composantes principales (CP) 1 et 2 (A) et 1 et 3 (B). Chaque losange de la même couleur matérialise un réplicat biologique de la même souche (6 au total par souche). Chaque ellipse colorée représente l'intervalle de confiance à 95 % de la distribution de chaque souche. C) Dendrogramme des souches obtenu par classification ascendante hiérarchique (distance euclidienne et critère de Ward). La localisation spécifique de ND5F est repérée par un carré rouge pour faciliter la visualisation.

Il apparaît sur la projection selon les dimensions 1 et 3 que l'ellipse matérialisant l'intervalle de confiance à 95 % de ND5F est individualisée (Figure II. 6B). Cette observation permet donc d'affirmer que ND5F possède une signature singulière au sein de notre échantillon. De plus, cette signature est confirmée sur la classification, par le groupement des six réplicats expérimentaux de ND5F au sein d'une branche distincte (Figure II. 6C). Cependant, ces résultats sont à relativiser s'il est pris en considération la variance globale de notre jeu de 82 variables. En effet, ND5F n'est individualisée que sur la projection selon les dimensions 1 et 3 de l'ACP qui ne représentent qu'un total de 45 % de la

variance totale et apparaît donc comme un phénomène non majeur au sein de la variabilité du jeu de données (Figure II. 6A). De plus, la signature de ND5F projetée sur les dimensions 1 et 3 est centrale sur l'ACP. Cette localisation révèle que la spécificité de notre souche d'intérêt se définit donc principalement par un positionnement médian sur la majorité des variables mesurées. Puisqu'aucun métabolite n'est surproduit par rapport aux souches de l'échantillon, il est postulé que ce sont les rapports des différents niveaux de métabolites les uns par rapport aux autres et notamment les absences de composés qui supportent la spécificité fonctionnelle de cette signature. Ainsi, l'élément majeur de la singularité de ND5F résulte d'un équilibre de concentrations des composés aromatiques qui confèrent aux produits un terrain organoleptique « balancé » voire « neutre ». Dans ce contexte, certains métabolites aux propriétés aromatiques spécifiques (seuil de détection bas) pourraient s'exprimer facilement et conférer aux produits fermentés par ND5F leur typicité, même si comparativement leur concentration n'est pas maximale au sein de l'échantillon de souches.

Le dendrogramme réalisé sur les 82 descripteurs ainsi que les projections des souches sur les dimensions 1 et 2 (totalisant 55 % de la variance), révèlent que ND5F présente une proximité importante avec les deux souches MC70 et EIP33A. Cette proximité est confirmée par la classification à partir des variables les plus contributives du jeu de données (20 variables identiques aux 20 COVs de la publication) (Figure II. 7). En effet, les souches de l'échantillon s'organisent au sein des quatre mêmes groupes, et ND5F est positionnée au sein du sous-groupe IV avec les souches EIP33F, EIP41A, EIP33A et MC70 partageant le même comportement global. L'analyse des coefficients de corrélations de ces 20 COVs avec les CPs permet de déterminer les orientations métaboliques majeures de chaque axe et d'identifier les principaux traits des sous-groupes de souches. Ces résultats sont, une fois de plus, similaires à ceux sans ND5F. Ainsi, la 1^{ère} dimension représente l'activité de lipolyse, la 2^{ème} la glycolyse et la 3^{ème} majoritairement la protéolyse. Il peut-être donc déduit que le sous-groupe IV regroupe des souches à la signature globalement équilibrée sans voie métabolique prédominante. En particulier, ND5F, MC70 et EIP33A présentent des activités lipolytiques faibles ainsi qu'une production d'arômes à partir des voies en aval du pyruvate (diacétyle, acétoïne, acétaldéhyde et butane-2,3-diol) réduite.

Ce résultat confirme l'hypothèse d'un terrain organoleptique neutre pour ND5F avec l'absence de nombreuses molécules aromatiques.

Ainsi, ce jeu de 20 composés organiques volatiles, sélectionné pour maximiser la variabilité des souches, n'est pas opportun pour identifier des marqueurs de spécificité de ND5F. D'autres variables doivent nécessairement intervenir dans la spécificité aromatique de ND5F et permettre d'expliquer les différences organoleptiques perçues entre les souches.

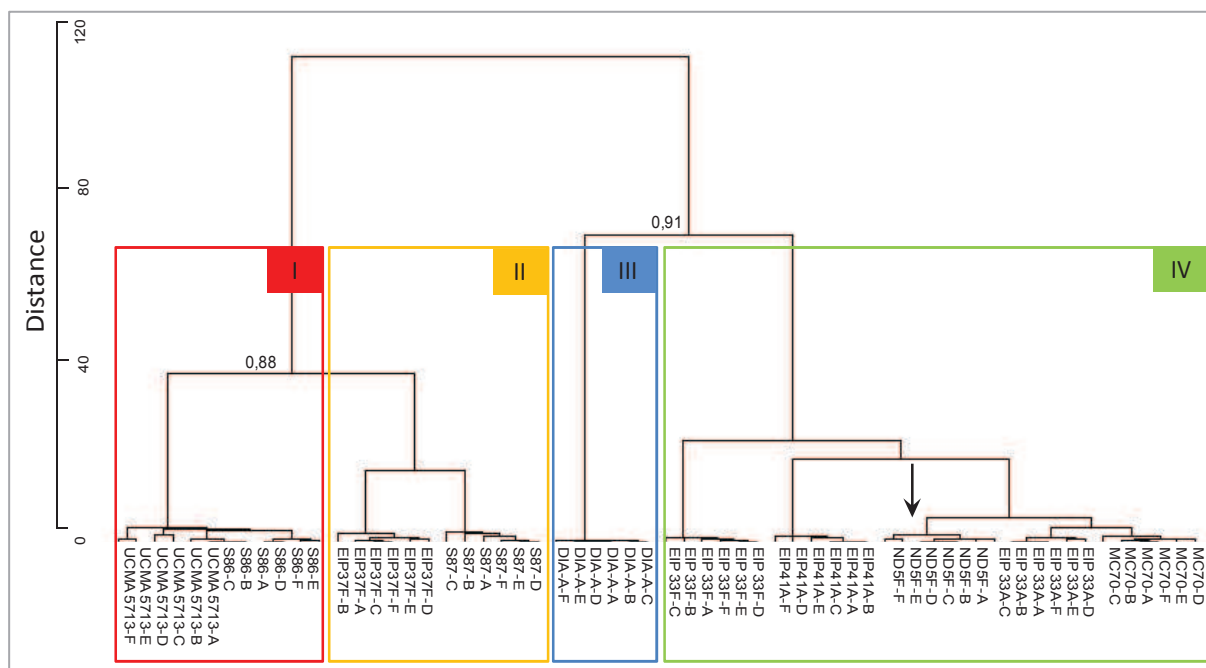
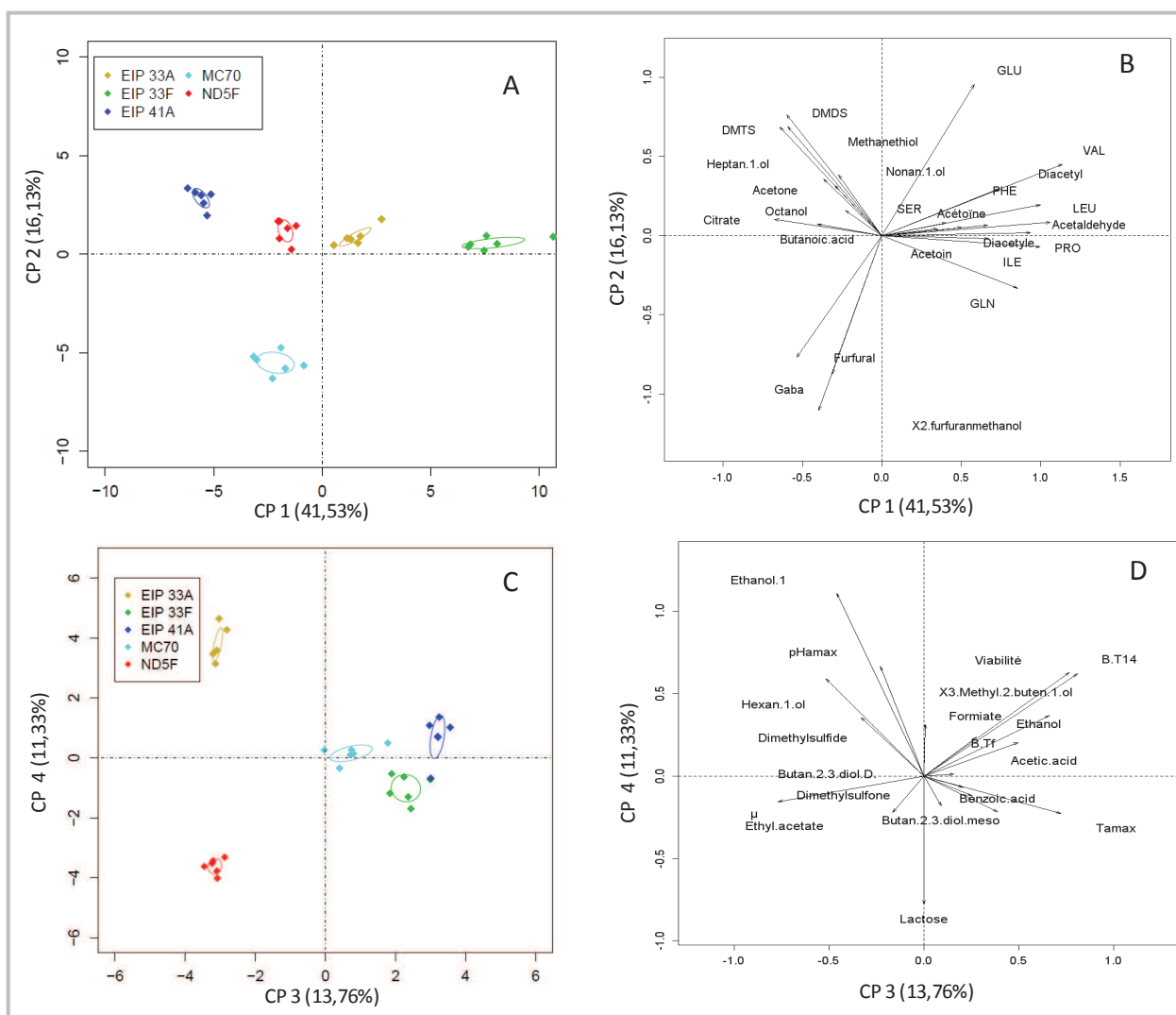


Figure II. 7 : Dendrogramme obtenu après Classification Ascendante Hiérarchique (CAH) de l'échantillon de neuf souches domestiquées de *L. lactis* ssp. *lactis* ainsi que de la souche *ND5F* (flèche noire) à partir des 20 variables les plus contributives, les COVs. Représentation des 4 principaux groupes identifiés avec notification de la valeur de bootstrap calculée pour chaque sous-groupe.

Pour confirmer cette hypothèse, des ACPs restreintes aux souches du sous-groupe IV, les plus proches de *ND5F*, ont été conduites sur l'intégralité des 82 variables (Figure II. 8). Il est observé une signature singulière pour chacune de ces cinq souches (Figure II. 8 A) et l'analyse conjointe des variables permet d'identifier un jeu de variables spécifiques pour leur signature (Figure II. 8 B). Ainsi, *EIP41A* est singularisée par des concentrations élevées en sulfure (di/triméthylsulfure), *EIP33F* par d'importantes concentrations résiduelles en acides aminés dans le surnageant, et *MC70* par la présence de GABA, de furanméthanol et de furfural. Sur cette projection, les souches *EIP33A* et *ND5F* sont proches, avec une position centrale, ce qui rend difficile toute attribution de marqueurs spécifiques à ces deux souches et confirme l'hypothèse de l'absence de marqueur individuel majeur de la singularité de *ND5F*.

De nouvelles représentations des souches sur des plans plus éloignés (3 et 4) ont été étudiées (Figure II. 8 C et D). Même si la variance totale est faible, seulement 25 % sur CP 3 et 4, il est possible d'attribuer la spécificité de *EIP33A* à une production élevée d'éthanol dans les laits fermentés. *ND5F* se démarque d'un point de vue métabolique par la production d'éthyle acétate, une faible présence d'acides organiques (formiate, acétate, lactate) et une concentration résiduelle de lactose à 14 J élevée. Physiologiquement, le comportement de *ND5F* apparaît comme contrasté, avec un taux de

croissance élevé pendant la maturation et une biomasse à 14 jours (B.T14) faible, tout comme la viabilité mesurée. Ces différentes variables sont autant de déterminants physiologiques spécifiques à ND5F qui confirme une vraie singularité de la souche par rapport à ces quatre souches, même si sur



les autres variables, le comportement de ces cinq souches est semblable.

Figure II. 8 : Analyse en composantes principales (ACP) des cinq souches du cluster IV dont ND5F à partir des 82 données phénotypiques. Représentations respectives des souches sur les dimensions 1 et 2 (A), et 3 et 4 (C). Chaque losange de la même couleur matérialise un réplicat biologique de la même souche. Représentations des variables les plus contributives (seuil à 75%) pour expliquer cette distribution de souches sur les dimensions 1 et 2 (B) et 3 et 4 (D). Abbréviation : μ : taux de croissance, DMS : Diméthylsulfide, DMDS : Diméthyltrisulfide et DMTS : Diméthyltrisulfide.

Cependant, ces variables seules ne permettent pas d'expliquer la typicité « crème » perçue lors de la dégustation de ND5F. De plus, elles ne sont que difficilement généralisables aux neuf autres souches alors que la souche ND5F est réellement reconnue lors d'une dégustation. Cette recherche de déterminants est d'autant plus difficile que les deux phénomènes majoritaires de la signature de ND5F mis en évidence sont, d'une part la faible activité lipolytique et glycolytique de la souche, et

d'autre part un comportement médian sur la majorité des autres variables mesurées. Ainsi, l'identification des déterminants de la signature de *ND5F* contraint à se focaliser sur une approche cherchant non plus à maximiser la variabilité des souches les unes par rapport aux autres (méthodes ACP-CAH), mais à isoler les variables qui singularisent *ND5F* face à un groupe composé des autres souches. Pour répondre à cette problématique, une analyse cette fois-ci supervisée est nécessaire.

III.3 Caractérisation métabolique de la typicité organoleptique de *ND5F*

Pour tenter de mettre en évidence la typicité de la souche, une nouvelle approche a donc été développée en utilisant une analyse supervisée discriminante dite de régression catégorielle des moindres carrés (abrégée en PLS-DA). Dans ce cas, la « réponse à expliquer » est une variable catégorielle prédéfinie, choisie afin de mettre en évidence spécifiquement les marqueurs responsables des différentes catégories présupposées. La classification examinée dans ce travail s'appuie sur l'opposition de la signature de *ND5F* à l'ensemble des autres souches. Cette partition repose sur l'hypothèse d'une discrimination systématique des laits fermentés de *ND5F* face à ceux d'autres souches. La réalisation de tests de dégustation par un panel d'experts a permis de valider ce pré-requis avec un taux d'erreur nul.

III.3.1 Identification des marqueurs discriminants de la typicité

Dans un premier temps, les 82 données phénotypiques ont été testées en tant que variables prédictives du partitionnement de l'échantillon en deux classes, la première constituée uniquement par les six réplicats de *ND5F*, groupe « *ND5F* » et la seconde composée des laits fermentés par les neuf autres souches de l'échantillon (soit 54 expériences), groupe « *Autres souches* ». Il est à noter que dans le cas d'une simulation de répartition aléatoire des expériences dans ces deux classes, le taux d'erreur de prédiction de classement attendu est de 29,89 %.

Dès la 1^{ère} dimension, le taux d'erreur de classement obtenu ici est nul, ce qui témoigne de la pertinence de la partition de ces signatures phénotypiques au sein des deux groupes choisis (Figure II. 9 A). La représentation graphique des individus confirme, par un positionnement spatial isolé et hautement spécifique de *ND5F* dans le quart gauche supérieur (Figure II. 9.B), la typicité de la souche. Les signatures des autres souches apparaissent comme peu marquées et forment un groupe homogène, ce qui renforce la spécificité de *ND5F*.

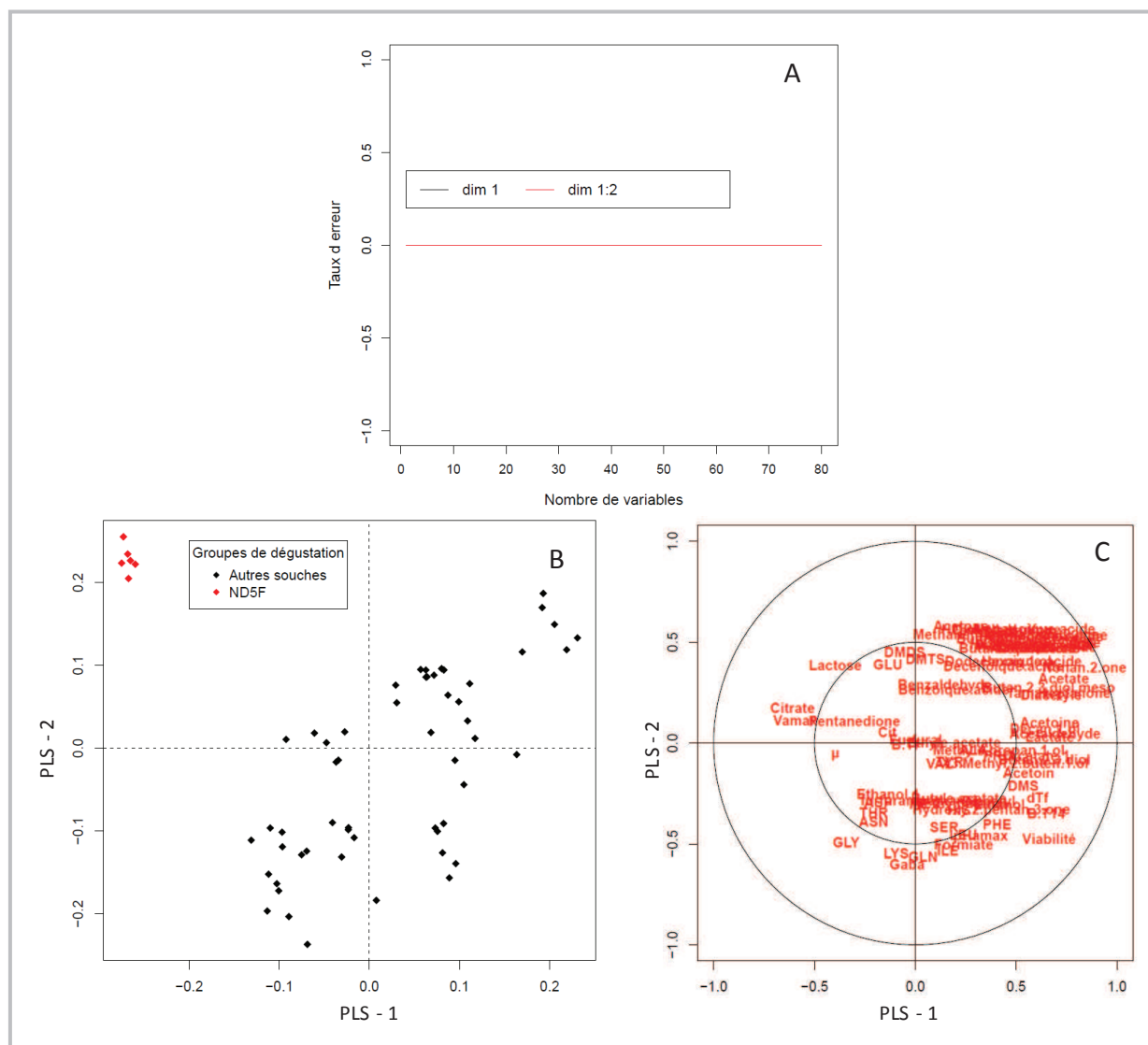


Figure II. 9 : Analyse PLS-DA de la classification en deux groupes « ND5F » et « Autres souches » à partir des 82 variables phénotypiques. A : Evolution du taux d'erreur de classement des souches en fonction du nombre de métabolites sélectionnés pour les deux premières dimensions prises en compte. Représentations des individus (B) et des variables (C) de la PLS-DA sur les composantes 1 et 2 de la PLS. Le cercle interne sur la représentation des variables matérialise un coefficient de corrélation de 0,50 avec les composantes PLS.

Les marqueurs phénotypiques responsables de cette discrimination sont déduits de la représentation des variables. Ainsi, la typicité de ND5F qui se visualise sur les dimensions 1 et 2 est liée à des concentrations résiduelles en lactose et citrate élevées, ainsi qu'à la présence de pentanedione dans les laits fermentés à 14 jours, et à des vitesses maximales d'acidification ($V_{\max}^a$) et de croissance (μ) importantes. Il est nécessaire de signaler que la singularité de ND5F est également liée aux faibles concentrations rencontrées en acides organiques (lactate, acétate, formiate), ainsi que pour la majorité des COVs (acétoïne/diacétyl, éthanol, acétaldéhyde, acides gras, lactones, méthyl cétones, alcools, et diméthylsulfide). Une fois de plus, ce résultat confirme l'hypothèse de travail développée au cours de cette étude identifiant la spécificité des laits fermentés par ND5F comme des produits au

terrain organoleptique neutre et sans amertume ou acidité (vraisemblablement liée aux taux faibles de lactate et d'acétate).

C'est dans ce contexte qu'il a été identifié la pentane-2,3-dione, molécule aux notes « crème », « beurre » et « gras » [Acree & Arn 2012] comme l'un des déterminants métaboliques de la typicité de ND5F. Il existe donc une parfaite adéquation entre les notes aromatiques de ce composé et le descripteur global utilisé pour identifier les laits fermentés par ND5F. La 2,3-pentanedione fait partie de la famille des dicétones dites vicinales dont le représentant le plus connu est la butane-2,3-dione également appelée diacétyl [Ott et al. 1997; Ott et al. 1999; Ott et al. 2000]. Ce composé est issu de la décarboxylation oxydative du 2-acétohydroxybutanoate, intermédiaire de la voie de biosynthèses des acides aminés branchés (voir Figure 12bibliographie) [Ott et al. 2000]. Ce précurseur est formé par l'acétolactate synthase anabolique (AHAS) qui condense une molécule de pyruvate provenant de la glycolyse avec de l'oxobutanoate issu de la dégradation de la thréonine. Il est constaté également que ND5F est inversement corrélée avec l'indicateur viabilité, la durée de fermentation (dT_f) et la majorité des acides aminés. Si cette diminution de viabilité est corrélée à un phénomène de lyse cellulaire et donc de libération du contenu cellulaire, cet indicateur deviendrait un déterminant majeur impactant la production aromatique également.

L'indicateur de viabilité est estimé ici par le rapport du dénombrement à 14 J sur le dénombrement à 24 h, et représente donc une estimation de la proportion de cellules à être métaboliquement actives à 14 J. Pour vérifier si les cellules de ND5F sont effectivement lysées à 14 J, et non juste viables mais non métaboliquement actives, une estimation de la viabilité a été réalisée à partir de marquages fluorescents. Ainsi, un double marquage avec le couple d'agents fluorescents DAPI/Iodure de propidium (IP) a été effectué pour déterminer la proportion de cellules lysées sur le total de cellules (voir Figure II. 10). Après observation d'une dizaine de champs indépendants sur trois réplicats biologiques, une proportion moyenne de 50 ± 10 % de cellules lysées a été observée. Ce résultat permet donc de valider la présence d'un phénomène de lyse (la moitié des cellules) au sein des échantillons de laits fermentés de ND5F à 14 J.

Les variables identifiées lors de cette analyse constituent donc les déterminants phénotypiques de la typicité fonctionnelle de la souche ND5F en tant que levain laitier. Il est à noter que certains de ces marqueurs avaient été précédemment identifiés lors de l'analyse par ACP de la diversité des souches à partir des 20 COVs (absence des COVs produits par lipolyse ou glycolyse), ou encore par l'ACP restreinte sur le cluster IV (concentration résiduelle en lactose, et faible production d'acides organiques).

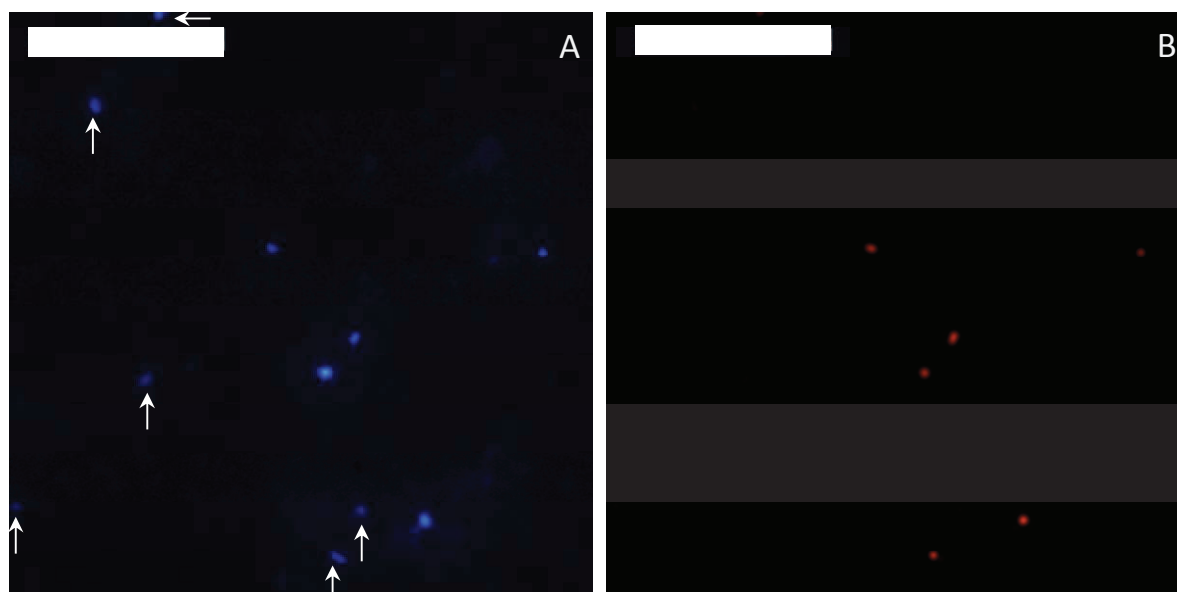


Figure II. 10 : Observation au microscope à fluorescence de la proportion de cellules lysées de *L. lactis* ssp. *lactis* ND5F sur un échantillon de lait fermenté à 14 J. Observation à l'objectif x100 d'un état frais après double coloration DAPI/Iodure de propidium (IP). A : Coloration des cellules totales en bleue par la fluorescence émise par le DAPI après fixation sur les acides nucléiques. B : Coloration des cellules lysées en rouge par la fluorescence émise par l'IP. Après comparaison des deux images, les cellules non lysées sont indiquées sur la figure A par des flèches blanches. La barre blanche indique l'échelle d'observation à 20 µm.

III.3.2 Sélection des marqueurs et évaluation du degré de typicité

L'objectif de cette section est de limiter le nombre de marqueurs phénotypiques à mesurer dans un souci de simplification de la problématique de caractérisation ultérieure de la signature de ND5F. Pour cela, une sélection de variables a été proposée à partir de la somme des connaissances acquises tout au long de cette étude : ACP restreinte aux souches appartenant au même cluster que ND5F (Figure II. 8), analyse des données de classification de variables, mais également ACPs réalisées sur des sous-groupes de variables organisées par famille métabolique ainsi que sélection de variables par la technique de sparse-PLS-DA. L'analyse a conduit à retenir 14 variables choisies pour discriminer le plus ND5F ; elles sont présentées par type d'indicateur dans le Tableau II. 4.

La qualité prédictive de ces variables est évaluée par une analyse PLS-DA d'une classification en deux groupes « ND5F » et « Autres souches ». Le taux d'erreur de prédiction de ce classement est nul dès la prise en compte des deux premières composantes, ce qui confirme la connaissance acquise sur les principaux marqueurs discriminants de la typicité de ND5F. Le positionnement du sous-groupe « ND5F » est singulier, avec un isolement au sein du quart inférieur de la Figure II. 11 A.

Tableau II. 4: Présentation des variables sélectionnées comme discriminantes pour ND5F face aux autres souches.

Variable sélectionnée	Famille de marqueur
Lactose Citrate Lactate Formiate Acétate	Source de carbone et métabolites en aval du pyruvate
Glycine	Acide Aminé
Viabilité cellulaire à 14 Jours μ , taux de croissance dTf, durée de fermentation V_{max}^a , Vitesse maximale d'acidification	Indicateurs physiologiques et technologiques
Méthanthiol DMDS, Dimethyldisulfide 2,3-pentanedione 2-furfuranmethanol	COVs

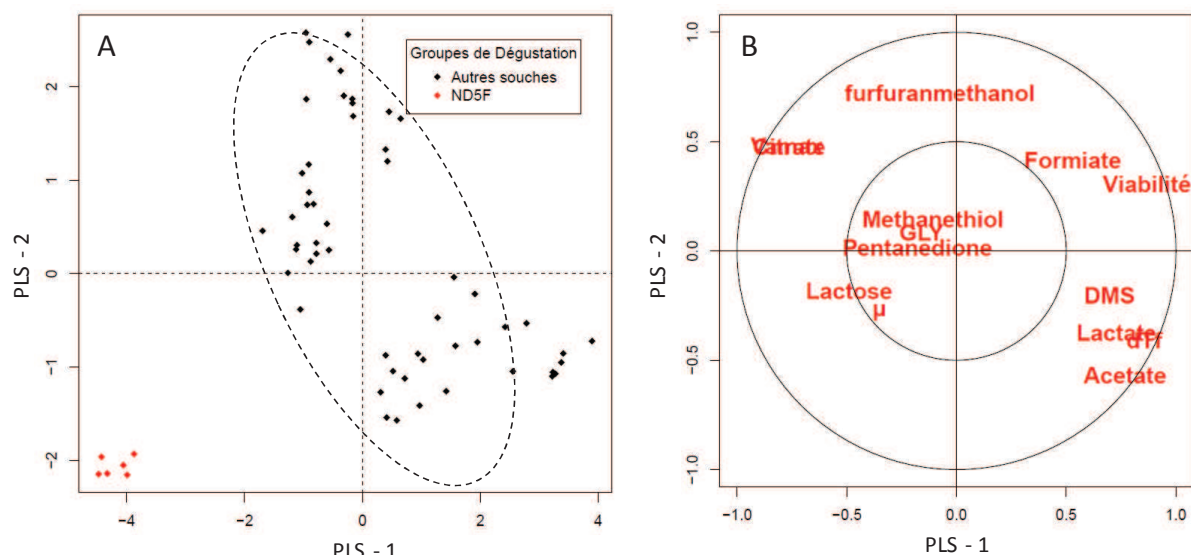


Figure II. 11 : Analyse PLS-DA de la classification en deux groupes « ND5F » et les « Autres souches » à partir des 14 variables sélectionnées comme les plus discriminantes. Représentations respectives des individus (A) et des variables (B) sur les composantes 1 et 2. L'ellipse en pointillés matérialise la zone de répartition des souches du groupe « Autre souches ». DMS : Diméthylsulfide.

En toute logique, il est retrouvé dans cette singularité les marqueurs phénotypiques déjà précités : concentration résiduelle en lactose et citrate élevée, fort taux de croissance, importante V_{max}^a pour

une faible viabilité et de faibles taux en formiate et lactate. Les très faibles taux de 2-furfuranméthanol et diméthylsulfide mesurés dans *ND5F*, viennent exacerber la différence entre les deux sous-groupes sur la dimension 2.

Ce résultat est également synthétisé sous la forme du Tableau II. 5 qui récapitule les principales relations fonctionnelles de chaque marqueur avec la typicité de *ND5F*.

Tableau II. 5 : Caractéristiques des déterminants phénotypiques sélectionnés par rapport à la signature de *ND5F*. + : corrélation positive, - : corrélation négative.

Variable sélectionnée	Relation à la signature de <i>ND5F</i> : type de corrélation
Lactose	+
Citrate	+
Lactate	-
Formiate	-
Acétate	-
GLY, Glycine	-
Viabilité cellulaire à 14 Jours	-
μ , taux de croissance	+
dTf, durée de fermentation	-
V^a_{max} , Vitesse maximale d'acidification	+
Méthanthiol	-
DMDS, Dimethyldisulfide	-
2,3-pentanedione	+
2-furfuranmethanol	-

La seconde partie de cette analyse consiste à contrôler si, à partir des 14 variables identifiées comme discriminantes pour *ND5F*, d'autres souches de l'échantillon parviennent à être ou non typiques. Pour ce faire, des analyses de PLS-DA ont été réalisées en singularisant une à une toutes les souches de l'échantillon pour observer si une typicité peut être mise en évidence (voir Tableau II. 6).

Seules les souches *S87* et *ND5F* parviennent à être correctement discriminées des autres souches avec un taux d'erreur nul à partir de ce jeu de 14 données. Cependant, l'analyse des graphiques d'individus et de variables, Figure II. 12A et .B, permet d'affirmer que la typicité de *S87*, bien que mise en évidence à partir des mêmes variables, est différente de celle de *ND5F*. En effet, le positionnement de *S87* se singularise principalement par des concentrations élevées en lactate-acétate, méthanthiol, diméthylsulfide et une forte viabilité, à l'inverse de *ND5F* (voir Tableau II. 5 et Figure II. 10).

Tableau II. 6 : Etude comparative des résultats de PLS-DA sur la typicité relative des 10 souches de l'échantillon à partir des 14 variables.

Souche testée	Taux et pourcentage relatif d'erreur maximale	Séparation en PLS-DA	
		PLS1-PLS2	PLS1-PLS3
DIA-A	0,02 soit 7 %	OK	NOK
EIP 33A	0,10 soit 33 %	NOK	NOK
EIP 33F	0,07 soit 23 %	NOK	NOK
EIP 41A	0,04 soit 13 %	OK/NOK	NOK
MC70	0,016 soit 6 %	OK	NOK
ND5F	0	OK	OK
S86	0,10 soit 33%	NOK	NOK
S87	0	OK	OK
UCMA5713	0,10 soit 33%	NOK	NOK

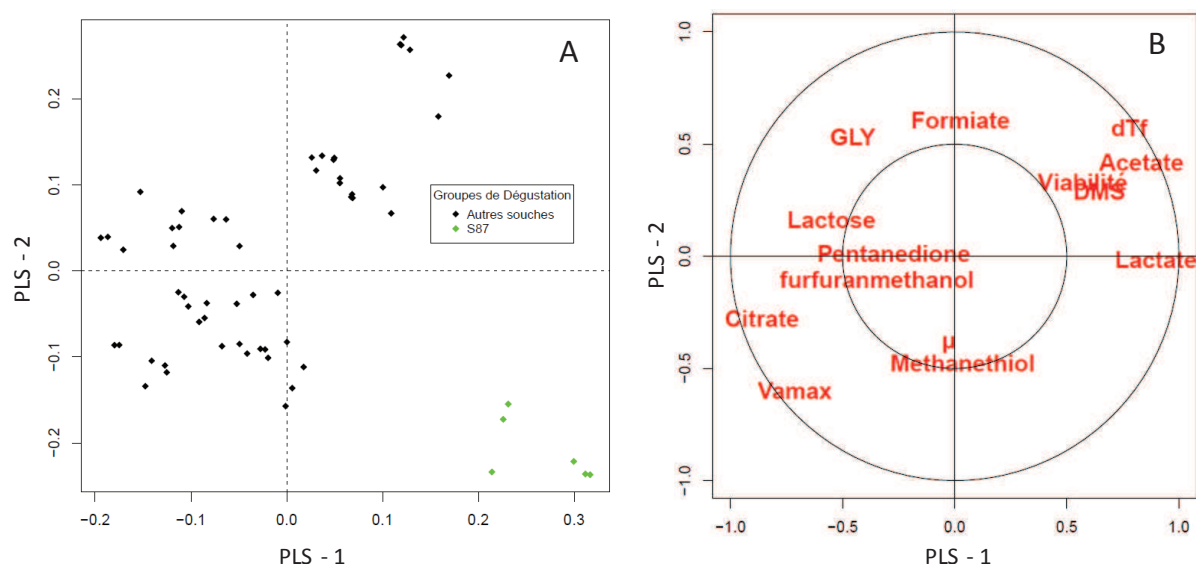


Figure II. 12 : Analyse PLS-DA de la typicité de la souche S87 à partir des données des 14 variables phénotypiques sélectionnées sur la base d'une classification en deux groupes « S87 » et « Autres souches ». Représentations respectives des individus (A) et des variables (B) sur les composantes 1 et 2. DMS : Diméthylsulfide.

En conclusion, ces résultats permettent d'alléguer qu'à partir du jeu de données testées, aucune souche de l'échantillon testé ne présente une typicité similaire à ND5F. Ce dernier résultat met en lumière également l'importance d'associer à ND5F non seulement les 14 variables mais également les relations (corrélations positives ou négatives) qui la définisse.

III.4 Approche intégrative génotype-phénotype pour affiner la classification fonctionnelle de ND5F

L'approche intégrative de caractérisation des levains laitiers mise au point à partir de données de diversité phénotypique et génotypique a été appliquée pour positionner la signature de ND5F dans le modèle de diversité génotypique-phénotypique des levains laitiers en laits fermentés. L'ACP révèle que la souche ND5F présente une signature phénotypique et génétique spécifique au sein de l'échantillon de souches (Figure II. 13). En effet, l'ellipse d'intervalle de confiance à 99% de ND5F ne se recoupe avec aucune autre.

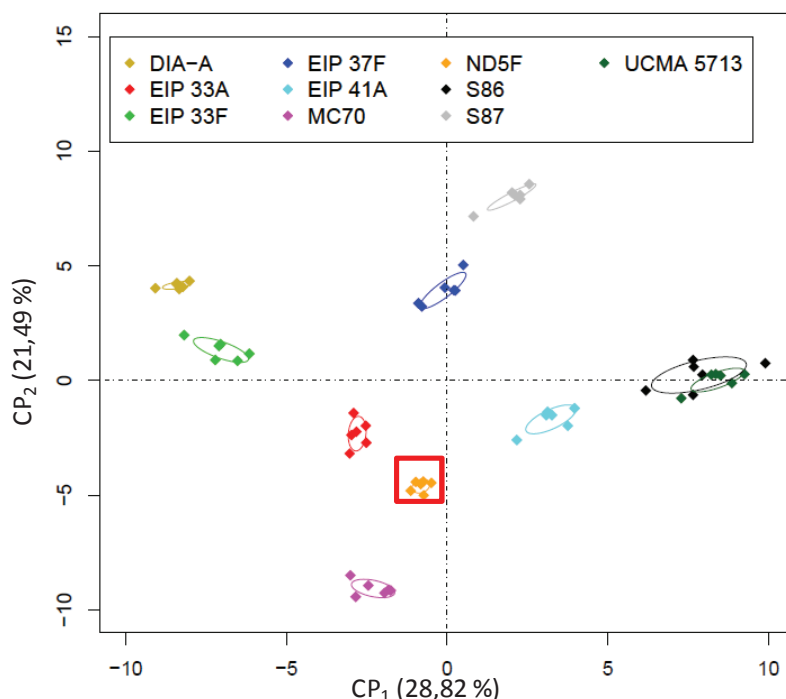


Figure II. 13 : Analyse en composantes principales (ACP) de l'échantillon de neuf souches domestiquées de *L. lactis* ssp. *lactis* ainsi que de la souche ND5F à partir des données génotypiques et phénotypiques. Représentation des 10 souches selon les composantes principales 1 et 2. Chaque losange de la même couleur matérialise un réplicat biologique de la même souche (6 au total par souche). Chaque ellipse colorée représente l'intervalle de confiance à 99% de la distribution de chaque souche. La localisation spécifique de ND5F est repérée par un carré rouge pour faciliter la visualisation.

Cette spécificité est donc le résultat d'une combinaison unique des 102 variables mesurées (niveaux de concentrations des métabolites ainsi que les distances génétiques à chaque souche). Notons que l'utilisation des variables génétiques est indispensable pour séparer ND5F de la souche MC70 sur les dimensions 1 et 2 (Résultat déjà présenté III.2) mettant ainsi en exergue la contribution des variables génétiques dans cette classification. L'analyse des variables met, une fois de plus, en exergue la participation des distances phylogénétiques, dont celle à MC70 du CC2, à cette séparation sur la composante 2.

Afin d'identifier quelles sont les variables les plus discriminantes de cette distribution, et quantifier leur contribution, une sélection des variables les plus contributives (seuil à 75 %) a été réalisée. Les 25 variables ainsi sélectionnées comprennent 22 indicateurs identiques au jeu obtenu en l'absence de *ND5F*, avec une majorité là encore de COVs (21/25), auxquels s'ajoutent la concentration en citrate et deux variables génotypiques. Ces deux variables représentent les distances calculées à partir du positionnement phylogénétique des deux souches du ST34, *EIP33F* et *EIP37F*. La classification des souches à partir de ces 25 marqueurs conduit à définir cinq sous-groupes principaux comme en l'absence de *ND5F* (Figure II. 14). Cependant, la composition des sous-groupes est sensiblement différente. En effet, au sous-groupe III comprenant la souche *DIA-A* est désormais ajoutée la souche *EIP33F*, résultat vraisemblable du fait de la proximité de ces deux souches sur la classification génotypique (voir Figure II. 5). Une seconde différence résulte du regroupement des souches *EIP33A* et *EIP41A* avec *ND5F* dans un seul et même cluster.

Cette classification, directement dépendante de l'ajout de *ND5F*, s'explique par la proximité génotypique de *ND5F* avec *EIP33A* qui a permis de former un sous-groupe central avec lequel *EIP44A* partage désormais le plus d'homologie. Ce résultat renforce nos précédents résultats démontrant l'importance de prendre en considération l'aspect génotypique pour développer une classification fonctionnelle qui tienne compte d'un éventuel rapprochement des souches lié à l'expression d'un potentiel commun dans un autre contexte laitier. De plus, l'ajout de *ND5F* permet de tester la robustesse de l'analyse présentée.

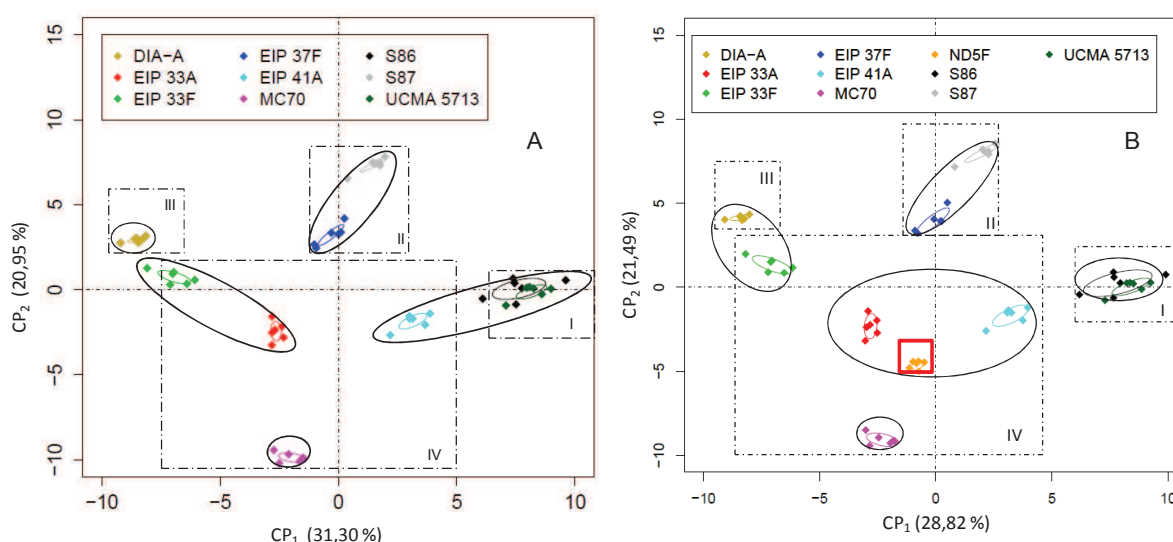


Figure II. 14 :Analyse en Composantes Principales (ACP) à partir des données génotypiques et phénotypiques avec apposition des résultats de classification sur les 25 variables les plus contributives (seuil à 75 %). Analyses sur les neuf souches de l'échantillon (A) ou avec l'ajout de *ND5F* (B).

CONCLUSION

La caractérisation génotypique de *ND5F* a permis de classer et de positionner phylogénétiquement la souche et d'attester de sa singularité génomique (empreintes PFGE). Ainsi, il est établi que *ND5F* appartient à la sous espèce *L. lactis* ssp. *lactis*, possède un typage allélique ST15 à partir du schéma MLST et fait partie des souches dites domestiquées de faible variabilité génétique. De plus, le séquençage de la souche *ND5F* a identifié 2 814 gènes (chromosomiques et plasmidiques) sur le génome de *ND5F*, dont seuls 48 sont spécifiques par comparaison aux génomes séquencés de l'espèce *L. lactis*. La majorité de ces gènes codent des activités liées aux prophages ou à des fonctions inconnues. En particulier, aucun gène impliqué dans un métabolisme spécifique n'a été identifié, empêchant toute prédiction et orientation génétique de la spécificité aromatique de la souche.

La seconde partie de cette étude s'est focalisée sur le développement d'une méthode de caractérisation de la diversité fonctionnelle de levains laitiers domestiqués en lait. Cette analyse a mis en évidence, à partir d'un échantillon représentatif de la diversité génétique des souches domestiquées, une importante diversité phénotypique de ces souches, et ceci malgré leur faible diversité génétique. Parmi les 82 descripteurs phénotypiques de l'activité des souches, 20 COVs ont été identifiés comme les principaux déterminants de cette variabilité définissant quatre sous-groupes de souches (de I à IV) à l'orientation métabolique divergente : lipolyse, glycolyse, protéolyse et non exprimée. L'apport du génotype (phylogénie par MLST et diversité génomique par PFGE) a ajusté la classification fonctionnelle des ferments en fonction du potentiel génotypique de chaque souche.

Le positionnement de *ND5F* au sein de cet échantillon de souches a révélé que les deux phénomènes majoritaires de la singularité de cette souche sont la faible activité lipolytique et glycolytique de la souche, et un comportement médian sur la majorité des autres variables mesurées. C'est dans ce contexte organoleptique neutre que la pentane-2,3-dione, molécule aux notes « crème », « beurre » et « gras » [Acree & Arn 2012] a été identifiée comme l'un des déterminants métaboliques de la typicité organoleptique de *ND5F*. En effet, il existe une parfaite adéquation entre les notes aromatiques de ce composé et le descripteur global utilisé pour identifier les laits fermentés par *ND5F*. Sur la base des connaissances phénotypiques de la singularité de *ND5F*, une liste de 14 marqueurs discriminants a pu être établie (4 COVs, 4 indicateurs physiologiques et technologiques, 5 sucres et métabolites en aval du pyruvate, ainsi qu'un acide aminé). L'identification de ces 14 déterminants et des relations fonctionnelles avec la signature de *ND5F* simplifie la problématique de caractérisation de la typicité de *ND5F*.

CHAPITRE III :

APPORT DE LA DIVERSITÉ PROTÉIQUE DE
LEVAINS DE *L. LACTIS* SSP. *LACTIS* A LA
CLASSIFICATION INTÉGRATIVE GÉNOTYPE-
PHÉNOTYPE

INTRODUCTION

Afin d'étudier l'origine de la diversité fonctionnelle mise en évidence sur les levains laitiers à l'échelle phénotypique (Chapitre II), des études comparatives de la sous-espèce *lactis* ont été menées au niveau protéique. L'approche protéomique a été retenue par rapport aux études transcriptomiques pour expliquer les singularités aromatiques de ces souches. Les niveaux de protéines, produits finaux de l'expression génique, semblent être un reflet plus précis, rendant compte des modifications post-transcriptionnelles et post-traductionnelles. En effet, les protéines sont responsables de la plupart des réactions cellulaires alors que les ARNm ne sont que des intermédiaires de synthèse. De plus, les régulations post-transcriptionnelles viennent perturber la corrélation entre quantité d'ARNm et de protéines [Dressaire et al. 2010; Picard et al. 2013].

Dans un premier temps, une approche de protéomique comparative a été menée sur les neuf souches de *L. lactis* ssp. *lactis* précédemment évaluées, afin d'identifier de potentiels mécanismes à l'origine de ces signatures phénotypiques. Une fois la diversité protéique des levains laitiers explorée, la classification obtenue a été comparée aux précédentes classifications phénotypique et génotypique déjà présentées. Afin de compléter l'étude de la diversité fonctionnelle des levains, les données génotypiques, protéiques et phénotypiques ont été intégrées pour proposer une caractérisation des potentialités des souches en environnement laitier.

Dans un second temps, une analyse de la singularité de l'expression protéique de la souche *L. lactis* ssp. *lactis* ND5F a été menée sur le contenu protéique, à partir du positionnement de ND5F dans cette diversité. En particulier, des marqueurs de sa typicité ont été recherchés par analyse différentielle et approche multi-variables pour décrire cette spécificité. L'analyse fonctionnelle de ces marqueurs vise à expliciter les propriétés phénotypiques observées.

I Exploration de la diversité protéique de levains de *L. lactis* ssp. *lactis* en lait

I.1 Exploration fonctionnelle des protéomes

Le protéome des neuf souches a été réalisé sur des extraits cellulaires de cultures en lait enrichi en matière protéique (35 g.kg⁻¹) et matière grasse (40 g.kg⁻¹), prélevés en phase stationnaire à pH 4,60. Après lyse cellulaire, extraction et digestion des protéines, les peptides obtenus ont été analysés par LC-MS/MS (voir matériels et méthodes). A partir des 54 échantillons analysés, correspondant à trois réplicats biologiques par souche et deux réplicats techniques par extrait, un total de 1062 protéines a été identifié et a fait l'objet d'une quantification relative. Cette identification a été réalisée par comparaison à la base de données de séquences protéiques non redondantes de NCBI nr constituée des séquences de *Lactococcus lactis* ssp. *cremoris* et ssp. *lactis* disponibles en date de novembre 2012 (identification des protéines avec un taux de faux-positif de 1% sur au moins un protéome).

Il est à noter que les séquences de la trypsine (enzyme de digestion) et des caséines ont été détectées au sein des échantillons, mais ont été exclues du jeu de données après normalisation des données pour la suite de l'analyse. Ainsi, pour chacune de ces 9 souches, nous disposons de 1057 protéines quantifiées.

I.1.1 Etude comparative par rapport aux données bibliographiques

Dans un premier temps, la diversité de ces données protéiques recueillies a été évaluée par comparaison avec les précédents protéomes complets publiés. Il n'existe à l'heure actuelle que des protéomes obtenus par électrophorèse bidimensionnelle à différents gradients de pHs pour déterminer l'expression des protéines de *L. lactis* [Guillot et al. 2003; Beyer et al. 2003; Palmfeldt et al. 2004; Drews et al. 2004; Gitton et al. 2005; Budin-Verneuil et al. 2005; Budin-Verneuil et al. 2007; Cesselin et al. 2009; Zhang et al. 2010; Mazzoli et al. 2010; Dressaire et al. 2011; Rochat et al. 2012]. En moyenne, la carte protéomique de *L. lactis* ssp. *lactis* IL1403 comprend 330 protéines identifiées, soit environ un quart des protéines cytoplasmiques à pHi acide, et moins de 13 % du protéome synthétique de la souche [Guillot et al. 2003; Gitton et al. 2005; Kilstrup 2005]. En effet, le protéome théorique de la souche IL1403 contient un nombre potentiel de 2321 protéines (N° Accession NCBI NC_002662.1). Sur les 1057 protéines de *L. lactis* détectées au sein de notre analyse, 1043 protéines sont communes au protéome synthétique de la souche IL1403 et représentent 49 % du contenu

protéique potentiel (Figure III. 1A). Ces neuf protéomes sont donc, d'après nos connaissances, le jeu de protéines identifiées et quantifiées le plus complet pour examiner l'expression différentielle de protéines sur des souches de *L. lactis*. Ce résultat s'explique d'une part par la méthode d'analyse en LC-MS/MS label free, plus sensible que les analyses 2DE-MS/MS traditionnelles, qui permet de détecter des protéines à des seuils de quantification plus bas [Shevchenko et al. 1996; Gygi et al. 2000]. D'autre part, le fait que notre jeu de données ait été identifié grâce à l'utilisation conjointe de 9 souches différentes a permis de mutualiser les caractéristiques des protéines identifiées. Ces 1057 protéines correspondent donc à un pan-protéome établi sur 9 souches.

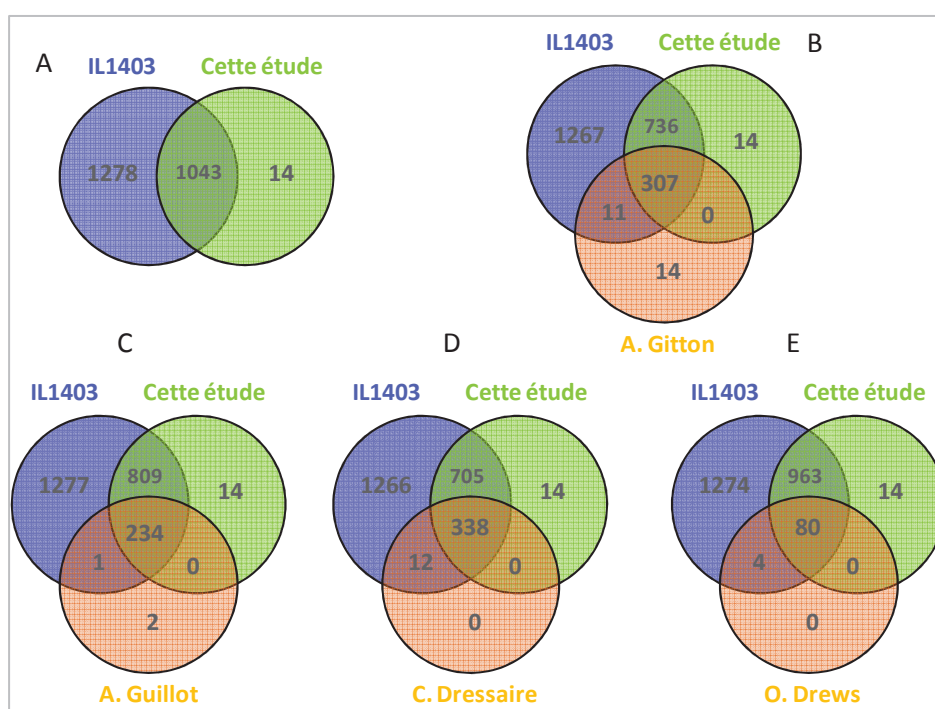


Figure III. 1 : Diagrammes de Venn comparatifs entre les neuf protéomes de *L. lactis* ssp. *lactis* de cette étude, le protéome synthétique de la souche *L. lactis* ssp. *lactis* IL1403 et les autres protéomes complets de *L. lactis* publiés. Comparaisons avec A) le protéome synthétique de *L. lactis* ssp. *lactis* IL1403 (N Accession NCBI NC_002662.1), B) le protéome acide (pH4-7) de *L. lactis* ssp. *cremoris* NCDO763 en lait [Gitton et al. 2005], C) le protéome acide (pH 4-7) de *L. lactis* ssp. *lactis* IL1403 sur M17-lactose et M17-glucose [Guillot et al. 2003], D) le protéome acide (pH 4.5-5.5 et 5-6) de *L. lactis* ssp. *lactis* IL1403 sur MCD en carence d'isoleucine [Dressaire et al. 2011] et E) le protéome basique (pH 9-12) de *L. lactis* ssp. *lactis* IL1403 sur milieu chimiquement défini SA [Drews et al. 2004].

Les quatorze protéines singulières qui sont absentes du protéome synthétique de la souche IL1403 présentent des correspondances avec les séquences protéiques des souches *L. lactis* ssp. *lactis* CV56 ou ssp. *cremoris* NZ9000. Deux de ces quatorze séquences correspondent à une ubiquinone/ménaquinone méthyltransférase et une isopropylmalate synthase, les douze autres restent des fonctions encore indéterminées. Le protéome en lait écrémé de *L. lactis* ssp. *cremoris* NCDO763 a été inclus dans cette étude pour une comparaison au plus près de nos conditions opératoires [Gitton et al. 2005]. Sur les 332 protéines identifiées par Gitton et coll., seulement 14

protéines présentes chez la souche *IL1403* n'ont pas été détectées dans notre jeu. Elles regroupent principalement des protéines impliquées dans la voie de biosynthèse des acides aminés (notamment l'histidine) et des fonctions putatives (Figure III. 1 B). Les quatorze protéines absentes de *IL1403* et spécifiques à *NCDO763*, se partagent entre des séquences putatives identifiées chez *L. lactis* ssp. *cremoris* NZ9000, et des protéines responsables de l'assimilation et du catabolisme du lactose/galactose. Ces protéines codées par des gènes de l'opéron plasmidique *lac* sont donc logiquement absentes du protéome théorique de *IL1403* (souche curée de plasmide). L'absence de détection de ces protéines sur notre jeu peut éventuellement s'expliquer par l'étude d'un protéome en phase stationnaire. En effet, il a été précédemment démontré que les enzymes contrôlant le transport des glucides voient leur expression réduite en cours de culture [N. Larsen et al. 2006].

Les trois autres protéomes complets publiés ont été réalisés sur des extraits cellulaires de la souche *IL1403*, en phase exponentielle après culture sur M17-lactose ou M17-glucose [Guillot et al. 2003] et sur milieu chimiquement défini [Dressaire et al. 2011]. Là encore, la majorité des protéines détectées est incluse dans notre jeu de 1043 protéines (Figure III. 1 C et D). Les deux protéines apparemment spécifiques au protéome de Guillot ont été annotées comme une cyclopropane acide gras synthase (gènes *cfa*) et une protéine putative (gène *yphL*). Ces dernières ne sont plus recensées dans le protéome synthétique de la souche *IL1403*, suggérant de potentielles erreurs d'annotation. Le protéome basique (pH 9-12) de *IL1403* sur milieu chimiquement défini SA a permis d'identifier, par gel 2D, 84 nouvelles protéines par rapport aux précédents protéomes acides [Drews et al. 2004] (Figure III. 1 E). Quarante-deux de ces protéines sont également présentes au sein des protéomes réalisés dans cette étude, attestant que la méthode d'extraction et d'identification mise au point permet d'obtenir un échantillon de protéines cytoplasmiques aussi bien acides que basiques. Enfin, le grand nombre de protéines communes avec les études de Dressaire et coll (384), suggère une similarité dans les réponses observées avec la carence en isoleucine, en accord avec les très fréquentes limitations en acides aminés rencontrées dans les environnements laitiers nécessitant la mise en place d'une nutrition azotée.

I.1.2 Analyse fonctionnelle des protéines identifiées

A partir des 1043 protéines communes au protéome théorique de la souche *IL1403*, la représentativité fonctionnelle de cet échantillon de protéines identifiées a été examinée. En effet, cette étude protéomique ne pouvant se faire que sur une fraction du potentiel protéique des souches, il est nécessaire d'étudier si des catégories fonctionnelles de protéines théoriquement attendues sont absentes ou inversement enrichies pour détecter d'éventuels biais. Pour répondre à cette problématique, une analyse fonctionnelle sur la base des catégories définies par Bolotin a été

conduite afin d'identifier si un sous-groupe fonctionnel de protéines serait éventuellement sur ou sous représenté dans cet échantillon par rapport au protéome synthétique de *IL1403* [Bolotin et al. 2001]. Ainsi, sur la base d'une distribution hypergéométrique, la significativité des probabilités calculées pour chaque catégorie et sous-catégorie fonctionnelle a été examinée (seuil de 1 %) (voir Tableau III. 1). Il apparaît que deux catégories fonctionnelles sont largement sous-représentées par rapport au protéome théorique : i) les fonctions dites « autres » associées principalement aux phages et transposons ($p\text{-value} \leq 3,6.10^{-18}$) et ii) les fonctions inconnues ($p\text{-value}$ de 6.10^{-20}). L'absence d'expression de la quasi intégralité des protéines phagiques était attendue puisqu'aucun cycle lytique n'est survenu au cours des différentes phases de cultures. L'important appauvrissement constaté en protéines de fonctions inconnues traduit vraisemblablement le caractère non essentiel de ces protéines pour répondre aux besoins cellulaires de la bactérie dans cet environnement. A l'opposé, la majorité des protéines des autres catégories sont sur représentées par rapport à la distribution du protéome théorique. En particulier, les protéines impliquées dans le métabolisme énergétique, les processus de réplication et de traduction ainsi que le métabolisme des purines et pyrimidines, présentent les probabilités d'enrichissement les plus importantes ($p\text{-value} \leq 2,5.10^{-5}$). Cette observation transcrit l'état cellulièrement actif des cultures au moment du prélèvement, bien qu'en phase stationnaire.

I.1.3 Réponse adaptative à la croissance en lait

L'abondance des protéines a été examinée sur ces neuf protéomes afin d'identifier les protéines invariantes qui constituent le cœur de la réponse protéique adaptative en lait lors de la phase stationnaire à pH acide. Pour cela, nous avons déterminé la fraction de protéines les plus abondantes communes sur chaque protéome. A partir des 25 % de protéines les plus abondantes pour chaque protéome, 138 protéines communes ont été sélectionnées sur les 264 potentielles correspondant au quart du protéome. Il est à noter que ce sous-ensemble de protéines communes ne comprend que 52 % des protéines les plus fortement exprimées de chacun des protéomes. Ainsi, ce résultat souligne la présence de protéines abondantes et différenciellement exprimées au sein de ces neuf souches, preuve d'une singularité dans l'expression protéique des levains.

L'analyse fonctionnelle des 138 protéines fortement exprimées et communes à toutes les souches révèle que trois catégories sont sur exprimées au sein des protéines les plus concentrées. La première catégorie concerne les protéines impliquées dans les étapes de traduction qui représentent à elles seules un tiers de ce pool de protéines, témoignage d'une machinerie de production de protéines hautement concentrée dans les cellules pourtant en phase stationnaire. Ainsi, 14 des 21

ARNt-amino acyl synthases de la bactérie sont identifiées, tout comme 23 protéines ribosomales sur les 62 détectées (p -values de respectivement $4,4.10^{-9}$ et $5,8.10^{-7}$).

Tableau III. 1 : Examen des catégories fonctionnelles et sous-catégories significativement sur ou sous-représentées entre les protéines identifiées et le protéome synthétique de *L. lactis* ssp. *lactis* IL1403. Seules les sous-catégories différentiellement exprimées sont présentées. Code p-value : * < 0,01, ** < 0,001 et *** < 1.10^{-4} .
- : Non significatif.

Catégories fonctionnelles, et sous catégories	Nombre de protéines au sein des 1043 protéines	Nombre de protéines dans le protéome synthétique de IL1403	Enrichissement (E) ou appauvrissement (A)	p-value
ACIDES AMINES, AMI	48	74	E	** (5,7.10⁻⁴)
Acides aminés de la famille aromatique	15	17	E	** (2,7.10 ⁻⁴)
Acides aminés de la famille aspartate	13	17	E	* (8,3.10 ⁻³)
PROCESSUS CELLULAIRES, CET	25	46	-	-
BIOSYNTHÈSE DE COFACTEURS, COF	28	44	E	* (9,1.10⁻³)
ENVELOPPE, ENV	54	85	E	* (3,4.10⁻³)
Murine sacculus et peptidoglycan	26	29		*** (5,4.10 ⁻⁷)
ACIDES GRAS ET MÉTABOLISME DES PHOSPHOLIPIDES, FAT	21	26	E	* (2,4.10⁻³)
MÉTABOLISME CENTRAL INTERMÉDIAIRE : POLYSACCHARIDES, SUCES AMINES... INT	8	27	-	-
MÉTABOLISME ÉNERGÉTIQUE, NRJ	120	177	E	*** (1,5.10⁻¹⁰)
Glycolyse	11	33	E	* (4,0.10 ⁻³)
Aérobie	26	37	E	* (1,5.10 ⁻³)
AUTRE CATÉGORIES, OTH	40	314	A	* (3,4.10⁻²)
Adaptation et conditions atypiques	18	21	E	** (1,3.10 ⁻⁴)
Prophages et fonctions connexes aux phages	18	220	A	*** (5,7.10 ⁻³⁷)
Transposon et fonctions connexes	1	66	A	*** (3,6.10 ⁻¹⁸)
PURINES, PYRIMIDINES, NUCLÉOTIDES ET NUCLÉOSIDES, PUR	58	61	E	*** (<3,3.10⁻¹⁷)
Voie de sauvetage et d'interconversion	19	19	E	*** (2,3.10 ⁻⁷)
Biosynthèse de ribonucléotides pyrimidiques	13	13	E	*** (2,9.10 ⁻⁵)
Biosynthèse de ribonucléotides puriques	16	16	E	*** (2,6.10 ⁻⁶)

Tableau III. 1.Suite :Examen des catégories fonctionnelles et sous-catégories significativement sur ou sous-représentées entre les protéines identifiées et le protéome synthétique de *L. lactis* ssp. *lactis* IL1403. Seules les sous-catégories différentiellement exprimées sont présentées. Code *p*-value : *<0,01, **<0,001 et ***< 1.10⁻⁴.

- : Non significatif.

Catégories fonctionnelles, et sous catégories	Nombre de protéines au sein des 1043 protéines	Nombre de protéines dans le protéome synthétique de la IL1403	Enrichissement (E) ou appauvrissement (A)	<i>p</i> -value
FONCTIONS RÉGULATRICES, REG	58	137	-	-
Protéines de liaisons au GTP	10	10	E	** (3,3.10 ⁻⁴)
RÉPLICATION, REP	41	65	E	* (2,1.10⁻³)
Réplication de l'ADN, restriction, modification, recombinaison et réparation	33	53	E	*** (7,8.10 ⁻³)
TRADUCTION, TRD	122	138	E	*** (<2,2.10⁻¹⁶)
Amino acyl-ARNt synthase	20	21	E	*** (1,2.10 ⁻⁶)
Facteurs de traduction	13	13	E	*** (2,9.10 ⁻⁵)
Protéines ribosomales : synthèse et modifications	62	75	E	*** (9,2.10 ⁻¹²)
Dégradation des protéines, peptides et glycopeptides	19	19	E	*** (2,3.10 ⁻⁷)
TRANSCRIPTION, TRP	26	32	E	** (2,5.10⁻⁵)
Synthèse d'ARN, modifications et transcription d'ADN	18	23	A	* (1,1.10 ⁻³)
TRANSPORT ET PROTÉINES DE LIAISON, TSP	92	223	-	-
Sucres, acides organiques et alcools	7	28	A	* (8,4.10 ⁻³)
FONCTIONS INCONNUES, UNK	287	870	A	*** (6,0.10⁻²⁰)

Egalement comprises dans cette catégorie, les enzymes de dégradation des peptides et des protéines, avec notamment les peptidases intracellulaires (PepN, PepXP, PepC, PepF, PepT et PepO), sont enrichies (p -value de $2,8.10^{-2}$). Cette présence s'explique par l'implication de ces enzymes dans la phase de protéolyse pour répondre aux besoins en acides aminés, et confirme la similarité de réponse avec la carence en acide aminé supposée précédemment. Dans la deuxième catégorie sur représentée, se trouvent les enzymes appartenant au métabolisme énergétique, avec l'intégralité des enzymes de la glycolyse (PGI, FFK, FBA, TPI, GAPDH, PGK, PMG, ENO et PYK) ainsi que la lactate déshydrogénase et la pyruvate carboxylase (*pycA*), 1^{ère} réaction du cycle tricarboxylique (p -value de $1,8.10^{-8}$). Ce résultat confirme l'abondance des enzymes impliquées dans le flux catabolique et l'importance dans la production d'énergie de ces voies. Le dernier sous-ensemble concerne les protéines impliquées dans le métabolisme des nucléotides (p -value de $1,2.10^{-2}$). En particulier, il est retrouvé des gènes de la synthèse *de novo* d'inosine monophosphate (IMP) (*purD*, *purL*, *purB*, *purH*) ainsi que ceux de la transformation de l'IMP en bases puriques (*guaB* et *guaC*).

La même approche a été transposée aux 25 % de protéines les plus faiblement concentrées (seuil 0,25). Cent-vingt-et-une protéines communes ont pu être sélectionnées, prouvant, là encore, que près de 64 % des protéines les plus faiblement concentrées sont spécifiques de certains des neuf protéomes. Les enzymes impliquées dans le transport et les réactions de synthèse membranaires (lipoprotéines, porines) sont les deux seules catégories enrichies (p -values de respectivement $5,4.10^{-3}$ et $2,3.10^{-2}$). Ce résultat révèle donc des niveaux de production basaux faibles et globalement invariants entre les souches, pour ces deux types de famille de protéines. Néanmoins, nous ne pouvons pas exclure que ces niveaux faibles d'expression communs à toutes les souches révèlent un biais expérimental des techniques de protéomique, comme par exemple la difficulté à extraire les protéines membranaires et pariétales. Toutefois, l'analyse fonctionnelle des 1057 protéines ne révélait aucun appauvrissement de ces protéines par rapport au protéome synthétique de la souche IL1403.

I.2 Sélection des protéines discriminantes

Au lieu de mener une analyse de la singularité des protéomes basée sur la comparaison deux à deux des protéomes et l'identification des protéines différentiellement exprimées, il a été choisi de conduire une analyse multi-variables. En effet, ces méthodes permettent d'appréhender la diversité protéique de cette sous-population de levains dans sa globalité en considérant l'interdépendance des protéines entre elles. De plus, elle renseigne sur l'organisation de la diversité de ces neuf protéomes. L'analyse des données à partir des méthodes exploratoires non supervisées telles que l'Analyse en Composantes en Principales (ACP) ou la Classification Ascendante Hiérarchique (CAH), n'a pas permis

de regrouper ensemble les différents réplicats biologiques d'une même souche. L'une des principales causes avancées pour expliquer ce phénomène est le large déséquilibre entre le nombre important de protéines (1057) face au faible nombre d'individus (27 échantillons). De plus, il est vraisemblable que seule une partie de ces variables est réellement biologiquement informative au regard de la problématique d'identification de signatures protéiques. Il est à noter que ce phénomène est récurrent sur les jeux de données à l'échelle « -omique » et nécessite de réduire le nombre de variables en ne sélectionnant que les plus discriminantes (Dash & Liu 1997; Lê Cao et al. 2007; Saeys et al. 2007). Pour répondre à ce double objectif de classification supervisée selon le facteur souche, et d'identification des protéines responsables de cette classification, il a été développé une approche basée sur les algorithmes des forêts aléatoires, ou random forest (RF). Cette méthode a notamment déjà été utilisée avec succès dans de nombreux cas pour identifier des marqueurs discriminants au sein de larges jeux de données transcriptomique, métabolomique ou encore protéomique [Shi et al. 2005; Stiglic et al. 2011; Acharjee et al. 2011; Meding et al. 2012; Chen et al. 2013]. Cependant, le nombre de classes est généralement plus faible que dans le cas présent, autour de trois ou quatre.

Dans un premier temps, une analyse par RF du jeu sans sélection de variables a été conduite sur les 1057 protéines afin d'estimer le taux d'erreur de prédiction médian par rapport à la classification souhaitée avec cette méthode. Après 1000 simulations, ce taux était de 48 %. Ce résultat est à relativiser face au taux d'erreur de prédiction estimé à près de 90 % si le classement des souches était réalisé aléatoirement. Ainsi, l'utilisation de la RF permet de diviser par deux le taux d'erreur de prédiction, en sélectionnant à chaque nœud de décisions le sous-set de protéines le plus pertinent pour regrouper les réplicats biologiques. La technique de sélection de variables mise au point pour cette étude utilise une hiérarchisation de l'importance des variables pour déterminer et identifier un nombre restreint de variables. L'importance d'une variable est calculée à partir de l'erreur de classement après brassage des variables et attribution aléatoire aux individus. Par conséquent, plus une variable est importante, plus l'erreur de classement est grande. L'objectif est de ne sélectionner que les protéines qui concentrent la majorité de l'information pour tendre vers la classification souhaitée. Pour cela, nous avons choisi de ne retenir, sur les 100 protéines les plus importantes, que celles communes à 70 % sur les 1000 simulations menées.

Cette procédure nous permet de définir les variables importantes et robustes de notre jeu de données. Il a été ainsi sélectionné 47 protéines qui se répartissent en onze catégories fonctionnelles (Figure III. 2). La catégorie majoritaire est celle des protéines inconnues, ce qui réclame d'approfondir la fonctionnalité de ces protéines divergentes entre les souches. De plus, onze protéines ont été identifiées comme appartenant au métabolisme énergétique (enzymes en aval du pyruvate, des

voies des pentoses phosphates, de génération d'énergie par utilisation des acides aminés, et de transport d'électrons). Cette catégorie s'est d'ailleurs révélée être enrichie dans le set des quarante-sept protéines sélectionnées par rapport au jeu de 1057 protéines (p -value : 0,012).

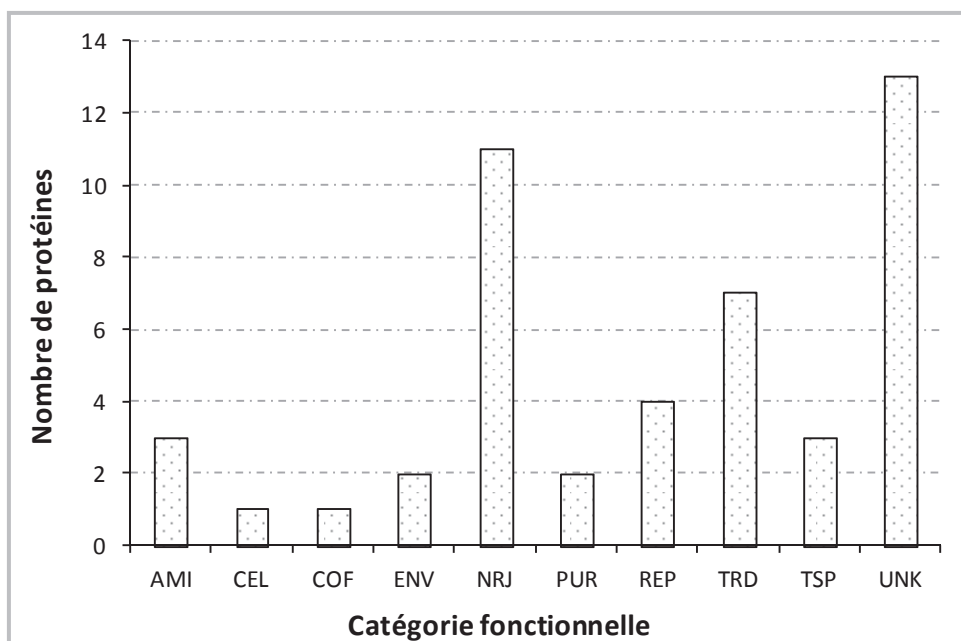


Figure III. 2 : Distribution des quarante-sept protéines discriminantes sélectionnées en fonction de leurs catégories fonctionnelles. Voir Erreur ! Source du renvoi introuvable. pour le détail des abréviations des catégories fonctionnelles. P-value d'enrichissement calculée pour chaque catégorie fonctionnelle. Code p -value : * $<0,05$, ** $<0,01$ et *** $<1.10^{-3}$.

Le détail de ces 47 protéines est présenté Tableau III. 2. Elles ont été organisées en quatre classes (Protéo-I à Protéo-IV) en fonction de leur profil d'expression sur les neuf souches (p -value de bootstrap de 0,82). La classe Proteo-I dénombre 29 protéines, avec un enrichissement spécifique en protéines impliquées dans le métabolisme énergétique (p -value : 0,011). Aucune amplification significative n'a été détectée sur les classes Proteo-II, Proteo-III et Proteo-IV, comprenant respectivement 9, 5 et 4 protéines. Pour déterminer si ces 47 protéines discriminantes regroupent des protéines au profil d'expression globalement faiblement ou fortement exprimé, chaque protéine a été classée par ordre croissant en fonction de son abondance. Pour faciliter l'analyse de ce classement, les rangs des protéines ont été normalisés par le nombre total de protéines par protéome (1057), pour présenter un rang correspondant à l'abondance relative de chaque protéine dans le protéome étudié. La concentration de près de 68 % des protéines est au-dessus de 50 % (voir rang médian Tableau III. 2), preuve que ces protéines sélectionnées concernent des protéines en général fortement abondantes au sein des cellules.

Tableau III. 2 : Analyse fonctionnelle des quarante-sept protéines discriminantes sélectionnées pour les neuf protéomes. Organisation des protéines en quatre classes selon les profils d'expression après classification. Il est indiqué pour chaque protéine le rang relatif minimal, médian et maximal au sein des neuf protéomes.

Classe	Gène	Catégorie	Protéine	Rang (%) : min- médiane-max
Protéo I	<i>pdhA</i>	NRJ	Pyruvate déshydrogénase E1 component alpha sous-unité (EC 1.2.4.1)	63-81-94
	<i>ybjB</i>	UNK	Protéine inconnue L196216	13-23-33
	<i>ycgG</i>	NRJ	NADH Oxidoréductase	67-76-93
	<i>optA</i>	TSP	Oligopeptide ABC transporteur, protéine de liaison au substrat	74-88-98
	<i>recA</i>	REP	RecA, Recombinase A	52-75-81
	<i>nifJ</i>	NRJ	Pyruvate-flavodoxine oxidoréductase	23-42-65
	<i>yfbB</i>	UNK	Protéine inconnue L111351	9-30-76
	<i>uvrB</i>	REP	Exonuclease ABC sous-unité B	44-65-82
	<i>glyA</i>	AMI	Serine hydroxyméthyltransferase (EC 2.1.2.1) (Serine méthylase) (SHMT)	3-28-77
	<i>ygiK (rnmV)</i>	UNK	Protéine inconnue L88730, protéine connexe à la primase	73-89-93
	<i>tagZ</i>	ENV	Protéine de biosynthèse de l'acide téichoïque	27-38-60
	<i>als</i>	NRJ	α -acétolactate synthase	74-85-97
	<i>citC</i>	NRJ	Acétate-SH-citrate lyase ligase (EC 6.2.1.22)	30-53-77
	<i>ymjF</i>	ENV	UDP-N-acétylglucosamine 2-épimérase	37-56-70
	<i>rplJ</i>	TRD	50S protéine ribosomale L10	64-90-96
	<i>yuiC</i>	UNK	Protéine inconnue L68401, YitL protéine homologie	35-67-89
	<i>ywiD</i>	UNK	Protéine inconnue L64833	25-47-69
	<i>ywjA</i>	UNK	Protéine inconnue L75803	8-18-38
	<i>rplM</i>	TRD	50S protéine ribosomale L13	80-94-99
	<i>gadB</i>	NRJ	Glutamate décarboxylase, sous-unité B (EC 4.1.1.15) (GAD)	73-89-95
	<i>gnd</i>	NRJ	6-phosphogluconate déshydrogénase (EC 1.1.1.44)	81-94-98
	<i>pepT</i>	TRD	Peptidase T (EC 3.4.11.14) (Tripeptide aminopeptidase)	80-91-96
	<i>pepC</i>	TRD	Aminopeptidase C (EC 3.4.22.40)	58-81-88
	<i>purK</i>	PUR	Phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21)	3-35-70
	<i>citD</i>	NRJ	Citrate lyase acyl-carrier protéine (EC 4.1.3.6)	53-75-87
	<i>metB2 (metC)</i>	AMI	Cystathionine β -lyase (EC 4.4.1.8) (CBL) (β -cystathionase) (Cystéine lyase)	85-95-97
	<i>sodA</i>	CEL	Superoxide dismutase [Mn] (EC 1.15.1.1)	77-88-94
	<i>rexA (addA)</i>	REP	Exonucléase RexA	38-63-81
	<i>rplP</i>	TRD	50S protéine ribosomale L16	64-88-95

Tableau III. 2 suite : Analyse fonctionnelle des quarante-sept protéines discriminantes sélectionnées pour les neuf protéomes. Organisation des protéines en quatre classes selon les profils d'expression après classification. Il est indiqué pour chaque protéine le rang relatif minimal, médian et maximal au sein des neuf protéomes.

Classe	Gène	Catégorie	Protéine	Rang (%) : min- médiane-max
Protéo-II	<i>ycjB</i>	UNK	Protéine inconnue L91807, putative glycoprotéine endopeptidase	42-61-84
	<i>yfgQ</i>	TSP	ATPase impliquée dans le transport de cation	7-21-40
	<i>yhjB</i>	UNK	Protéine inconnue L193291, présence d'un domaine rhodanèse	5-15-61
	<i>mae</i>	NRJ	Malate oxidoréductase (EC 1.1.1.38)	72-86-98
	<i>citF</i>	NRJ	Citrate lyase alpha chaîne (EC 4.1.3.6)	84-92-99
	<i>yneC</i>	UNK	Protéine inconnue L142410	51-66-94
	<i>trxA</i>	COF	Thiorédoxine	58-83-92
	<i>LLNZ-00265</i>	UNK	Protéine inconnue identifiée chez <i>L. lactis</i> ssp. <i>cremoris</i> NZ900	75-93-98
	<i>rplK</i>	TRD	50S protéine ribosomale L11	33-91-52
Protéo-III	<i>yjbC</i>	UNK	Protéine hypothétique L112952, Peptidyl-prolyl cis-trans isomérase	2-8-18
	<i>dnaC</i>	REP	Réplivative DNA hélicase	16-24-85
	<i>purH</i>	PUR	Protéine de biosynthèse des purines, PurH	88-98-100
	<i>ileS</i>	TRD	Isoleucyl-ARNtsynthase (EC 6.1.1.5) (IleRS)	6-11-34
	<i>ydgB</i>	TSP	Perméase acide aminé	0-7-33
Protéo-IV	<i>gltB</i>	AMI	Glutamate synthétase (grande sous-unité)	68-75-97
	<i>ytdB</i>	UNK	Protéine d'initiation de la division cellulaire, DivIVA	71-89-98
	<i>butA</i>	NRJ	Acétoïne réductase (EC 1.1.1.5)	25-45-90
	<i>CVCAS-1336</i>	UNK	Protéine inconnue identifiée chez <i>L. lactis</i> ssp. <i>lactis</i> CV56	3-10-30

Huit protéines présentent une abondance faible, en deçà du quart, mais une seule protéine (gène *yjbC*, du groupe Protéo III correspondant à une Peptidyl-cis-trans isomérase, impliquée dans le repliement des protéines transmembranaires) présente pour chacune des souches une concentration maximale inférieure à 25 %.

Au sein de ces 47 protéines discriminantes, il peut être souligné la présence de trois enzymes en aval du pyruvate et impliquées dans la production d'acétolactate, diacétyle/acétoïne et butanediol. Ainsi, l'abondance en acétolactate synthase (*als*), pyruvate déshydrogénase (*pdhA*) et acétoïne réductase (*butA*) diverge selon les souches, pouvant aller d'un rang minimal de 25 % à plus de 90 %. Il pourra être exploré le lien entre niveaux d'expression protéique et différences de production de ces arômes. Trois des enzymes de dégradation du citrate sont également présentes : l'acétate SH-citrate lyase (*citC*) et deux des sous-unités de la citrate lyase (*citD* et *citF*). Ces protéines associées aux souches du biovar *diacetylactis* préfigurent d'une séparation potentielle de ces souches par rapport à la sous-population de levains. Il peut être également noté qu'il a été identifié l'aminotransférase cysthionine- β -lyase (*metB2*), engagée dans les réactions d'élimination de la chaîne latérale de la méthionine pour produire du méthaneithiol. Là encore, il pourra être examiné les corrélations entre les concentrations de cette enzyme et le dosage de méthaneithiol entre les souches. Pour finir, la présence de la superoxyde dismutase (*sodA*) témoigne d'une réponse au stress oxydatif au sein des différentes souches (rang minimal de 77 %) et cela, même si le procédé de production n'inclue pas une phase d'aération proprement dite. Le niveau d'expression de la protéine reste néanmoins un facteur de discrimination entre les souches.

I.3 Identification des signatures protéiques

A partir de ces 47 protéines, une classification ascendante hiérarchique permet de positionner les relations fonctionnelles entre les neufs protéomes (Figure III. 3). Il est observé que chaque souche se singularise autour de ses trois réplicats, avant de se regrouper avec d'autres souches (p -value de bootstrap $\geq 0,86$). Ce résultat démontre la qualité de la sélection de ces 47 protéines pour illustrer la diversité protéique de ces neuf souches. Ce résultat confirme par ailleurs la très faible erreur de classification à 0,6 % obtenue lors de la classification supervisée en RF des neuf souches sur ce sous-ensemble de protéines. Cette diversité protéique est en accord avec le haut degré de singularité phénotypique précédemment observé pour chaque levain en lait. Cependant, il est également possible d'identifier une organisation de cette sous-population en trois groupes principaux (Groupes A, B et C) (p -value de bootstrap $\geq 0,83$). Le groupe A rassemble trois des quatre souches appartenant au biovar *diacetylactis* (*EIP33F*, *EIP37F* et *DIA-A*). Seule la souche *S87*, possédant également le gène

Pour interpréter au mieux la significativité biologique de ces groupes de souches et identifier les protéines discriminantes pour chacune des souches, une ACP a été réalisée à partir de ces 47 protéines. Cette dernière résume près de 75% de la variabilité sur les trois premières composantes (Figure III. 4). En adéquation avec les données de classification, les ellipses de confiance ne se superposent pas, preuve de la singularité de chaque protéome. Pour faciliter l'interprétation, les trois sous-groupes de souches A, B et C, sont représentés sur les graphiques d'individus (Figure III. 4.A et B). Ces trois groupes sont principalement séparés le long de l'axe 1 qui représente près de 55 % de la variance expliquée. L'analyse des variables permet de décrypter biologiquement la construction de cet axe (Figure III. 4.C et D). Ainsi, les variations des protéines de la classe Protéo-III (PurH, DnaC et IleRS) négativement corrélées à la première dimension s'opposent aux trois autres classes de protéines Protéo-I, II et IV.

Au sein du sous-groupe B, la proximité des profils d'expression sur les protéines de la classe Protéo-III, et les abondances plus faibles dans la majorité des autres protéines (protéines Protéo-I, II et IV), expliquent le regroupement des souches *EIP33A* et *MC70*. La position centrale du sous-groupe A témoigne de concentrations équilibrées pour les protéines des classes Protéo-III et I. Par conséquent, ce n'est pas la présence de concentrations les plus élevées en citrate lyase (CitD et CitF) qui explique le fondement de ce regroupement mais une proximité plus globale des profils protéiques. Cette donnée est également cohérente avec l'absence de regroupement de la dernière souche *diacetylactis*, *S87*.

La position individualisée des trois souches du biovar *diacetylactis* s'explique par des divergences sur la deuxième composante. Ainsi, les souches *EIP33F* et *EIP37F* corrélaient positivement avec la perméase d'acide aminé codée par le gène *ydgB*, mais également avec les protéines de la classe Protéo-II. La souche *DIA-A* se singularise par une présence importante en acétoïne réductase (AR, gène *butA*), enzyme responsable de la conversion de l'acétoïne en butane-2,3-diol. Cet élément corréla parfaitement avec la description phénotypique de cette souche qui produisait les quantités relativement élevées en butane-2,3-diol. D'autres protéines ont également été identifiées comme marqueurs de la singularité protéique pour cette souche : glutamate synthétase (gène *gltB*), protéine de division cellulaire DivIVA (gène *ytdB*), des protéines de biosynthèse de l'acide teichoïque (gène *tagZ*) et enfin la protéine inconnue (gène *CVCAA-1336* de la souche *CV56*).

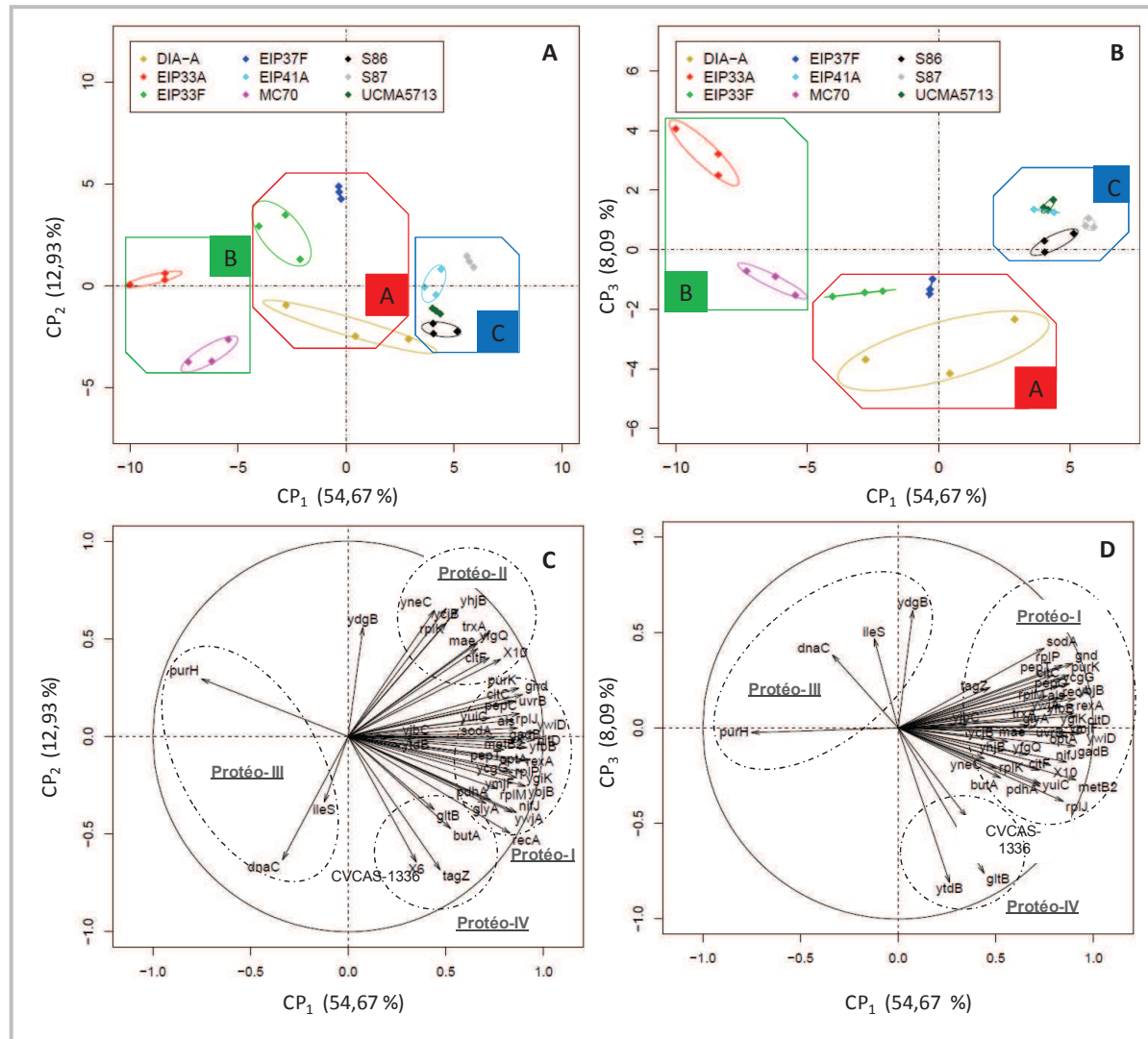


Figure III. 4 : Analyse en composantes principales de la diversité des neuf protéomes de *L. lactis* ssp. *lactis* après croissance en lait et prélèvement à pH 4,60. Représentation des individus sur les dimensions 1 et 2 (A) et 1 et 3 (B), avec mention des trois sous-groupes de souches A-B-C. Les 47 protéines sont représentées par leur nom de gènes selon les dimensions 1 et 2 (C) et 1 et 3 (D), et des ellipses en pointillé matérialisent les différentes classes de variables Protéio I, II-III-IV. Les ellipses de couleur désignent les intervalles de confiance à 95 % autour des différents réplicats des souches.

Les quatre souches du sous-groupe C se caractérisent par des concentrations importantes en protéines hautement corrélées entre elles et appartenant à la classe I. Ainsi, il est retrouvé des enzymes liées à l'utilisation du pyruvate et à l'assimilation du citrate (ALS, PDH, CitC et CitD), à la voie des pentoses phosphates (6-phosphogluconate déshydrogénase), ou encore participant aux étapes de nutrition azotée, avec des peptidases (PepT et PepC) et le transporteur d'oligopeptides Opp (sous-unité OppA) (Figure III. 4.C et D, Tableau III. 2). Un second ensemble de cinq protéines peut-être distingué en lien avec la réponse aux différents stress environnementaux (oxydatif et acide). Dans le cadre de la lutte contre le stress oxydatif, il a été identifié au sein de cette classe la superoxyde dismutase (gène *sodA*), qui élimine l'anion toxique O_2^- , et une NADH oxydase (gène *ycgC*) en charge de la régénération alternative du NADH [Sanders et al. 1999]. Deux autres enzymes de ce processus

de réponse sont également présentes dans ce groupe protéique : la recombinaise A, enzyme clef dans la réponse SOS engagée dans la réparation des dommages de l'ADN [Duwat et al. 1999], et l'exonucléase RexA également impliquée dans la recombinaison et la réparation de l'ADN.

Ainsi, l'une des singularités de ces quatre souches réside dans les concentrations plus élevées pour ces enzymes abondantes (rang médian au-delà de 70 %), ce qui pourrait permettre à ces souches de lutter plus efficacement contre la toxicité de l'oxygène. En réponse à l'autoacidification des souches, l'expression de la glutamate décarboxylase (gène *gadB*) assure l'élimination d'un proton cytoplasmique pour neutraliser le pH interne et permet, par l'action conjointe de l'antiport glutamate/ γ -aminobutyrate, l'export d'acide γ -aminobutyrique [Sanders et al. 1998]. Là encore, cette expression au sein de ces souches pourrait conférer une résistance plus importante de ces souches à l'acidification du lait, prolongeant ainsi leur activité métabolique.

I.4 Vers une classification intégrée de la diversité des levains

Bien qu'il ait pu être observé une singularité des souches sur chacun des niveaux étudiés, les classifications présentées ici (au niveau d'expression des protéines) et précédemment (au niveau du phénotype ou du génotype ou bien en intégrant les deux) pour décrire la diversité de cette sous-population de levains divergent fortement. Ces différences de positionnement des souches retranscrivent le haut degré de régulation contrôlant l'expression du génome pour déterminer le phénotype laitier d'une souche. Ces observations imposent donc d'aborder cette problématique de diversité de levains de *L. lactis* ssp *lactis* par une démarche intégrée de ces différents niveaux pour refléter la prise en compte d'un système global d'expression. Deux approches complémentaires ont été menées afin de tenter d'établir cette classification intégrative modèle en fonction du questionnement biologique.

I.4.1 Evaluation des marqueurs prédictifs de la diversité des souches

Dans un premier temps, il a été décidé de restreindre cette analyse aux principaux marqueurs de la diversité phénotypique (les 20 COVs) et protéique (les 47 protéines) afin d'évaluer la capacité prédictive de ce nombre réduit de variables à décrire la singularité des souches. Il a été aussi intégré à cette analyse les 18 distances génotypiques pour maximiser l'apport de l'information génétique.

La classification intégrative des souches révèle que chacune de ces souches est parfaitement singularisée à partir des 87 variables sélectionnées (p -value de bootstrap $\geq 0,95$) (Figure III. 5) démontrant ainsi la capacité de ces marqueurs discriminants à prédire une typicité par souche.

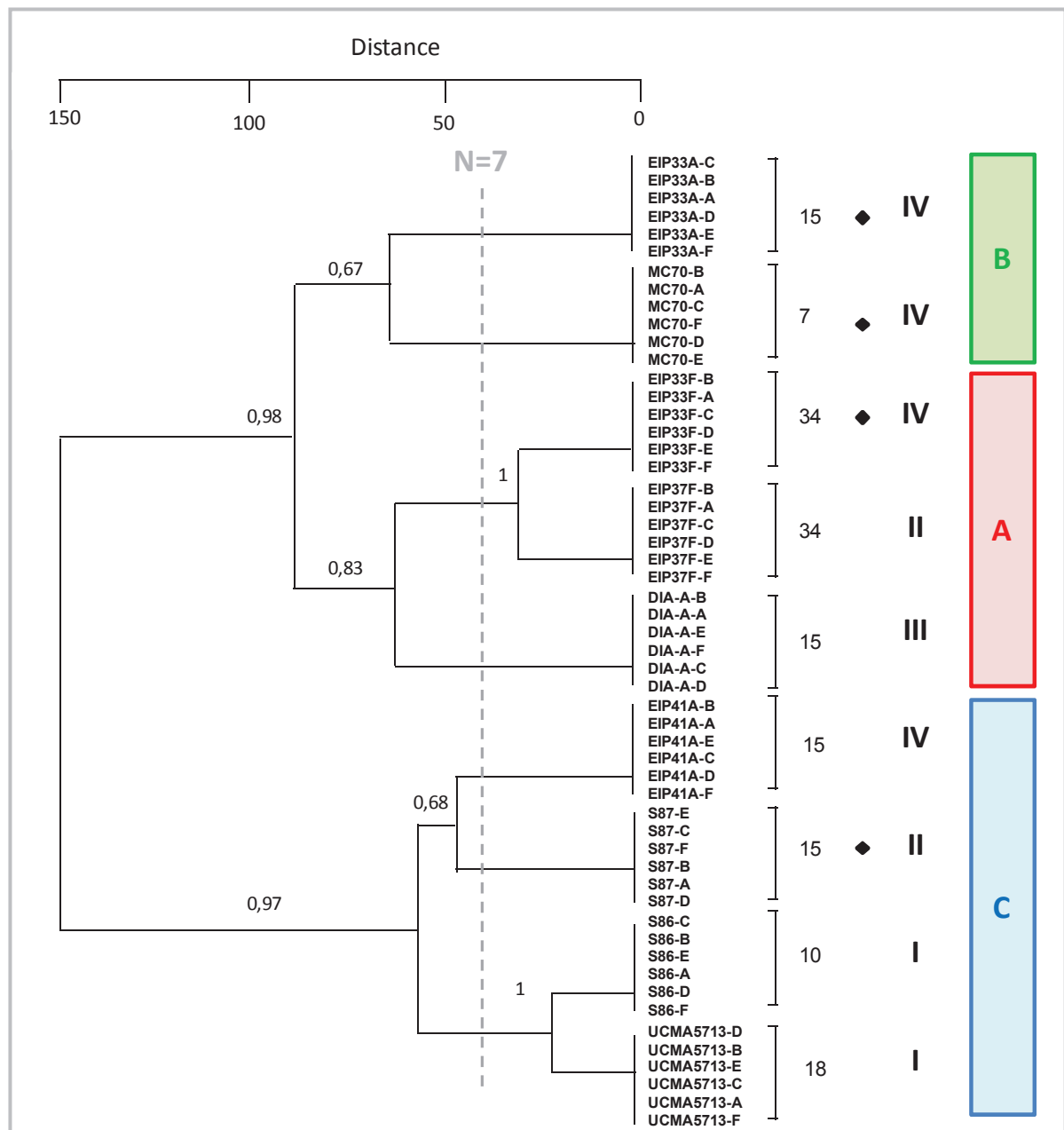


Figure III. 5 : Classification intégrative du génotype au phénotype de la singularité des souches de *L. lactis* ssp. *lactis* après croissance en lait à partir des marqueurs prédictifs identifiés. La distance euclidienne et le critère de Ward ont été utilisés pour cette classification sur la base des 18 distances génétiques (9 distances de phylogénie par MLST et 9 distances génomiques par empreintes génomiques), 20 COVs et 47 protéines pré sélectionnés. De gauche à droite, le dendrogramme des souches, le nom des 54 expériences (neuf souches avec six réplicats de A à F, données protéiques complétées par la moyenne), STs de la MLST, présence du gène *citP* (losange noir), et numéro des quatre groupes phénotypiques (I à IV) et des trois protéiques (A, B et C). Les valeurs de bootstrap sont indiquées pour chaque nœud (n=1000). La ligne verticale en pointillé décrit le niveau de coupe du dendrogramme pour interprétation.

L'analyse détaillée du dendrogramme et de l'ébouilisé de la variance (non présenté) établit que cette diversité peut se subdiviser en 7 sous-groupes dont l'organisation reprend les caractéristiques de la diversité protéique précédemment explorée (groupes A, B et C). Ainsi, il semblerait que les 47 protéines représentent un profil d'expression d'une singularité plus importante pour distinguer les

souches que les 20 COVs. Cependant, l'apport des variables génotypiques et phénotypiques est responsable, quant à lui, de la subdivision plus fine en sept sous-groupes.

Alors que certaines variables fortement corrélées entre les jeux peuvent apparaître comme redondantes pour l'identification des souches, d'autres, au contraire, permettent de mettre en évidence des potentialités fonctionnelles qui n'avaient pas été testées lors de l'analyse phénotypique. Ainsi, des essais pourront-être menés pour vérifier si les divergences d'expression de protéines impliquées notamment dans la lutte contre les stress (acide ou oxydatif) retranscrivent des capacités de résistance particulière des souches.

Dans le prolongement de cette analyse, il a été décidé d'examiner la diversité des souches selon une approche intégrative dite « globale », se basant sur une analyse des 1157 indicateurs génotypiques, protéiques et phénotypiques. Cette démarche a pour objectif de maximiser l'exploration des différentes potentialités fonctionnelles des souches, et également de tester la pertinence des marqueurs prédictifs de la diversité des souches décrits ci-dessus. Pour faciliter cette étude, une sélection supervisée de variables par random forest a été conduite comme précédemment, pour isoler les variables discriminantes de la singularité des souches. La RF sur ces 1157 variables présente une erreur de prédiction d'à peine 1,85 %, ce qui confirme les capacités de ces différents jeux de données à discriminer les neuf souches.

Après sélection, 73 variables ont été identifiées, se répartissant en 14 distances génotypiques, 21 protéines et 38 indicateurs phénotypiques. La classification des souches se subdivise en 9 sous-groupes qui individualisent chacune des souches (résultats non présentés). Néanmoins, l'organisation de l'arbre en 4 groupes principaux reprend la structure de la diversité protéique (groupes A, B et C), à l'exception de la souche S87 qui forme un groupe unique. Ce résultat est cohérent avec la classification des marqueurs prédictifs, validant ainsi la pertinence de ce sous-ensemble d'indicateurs. De plus, l'importance des protéines pour structurer la diversité des souches est réaffirmée, prouvant que la vingtaine de protéines sélectionnées concentre les principaux éléments divergeants des souches. Cependant, l'apport des variables génotypiques et phénotypiques est responsable de la singularité des souches.

I.4.2 Classification intégrée de la typicité aromatique

Le dernier axe de recherche, plus applicatif, concerne la recherche à l'échelle intégrative de marqueurs prédictifs de la diversité fonctionnelle des souches dans le modèle de production des laits fermentés. En effet, il a été précédemment observé lors de l'analyse phénotypique que les souches se regroupent en 4 sous-groupes en fonction des orientations aromatiques principales de leur métabolisme : glycolyse, lipolyse, protéolyse ou équilibrée (voir chapitre II). Cette classification phénotypique correspond à un résultat directement interprétable pour un producteur de ferment dont le but est de positionner la typicité aromatique de ses souches. Etre capable d'isoler des variables (génétiques, protéines ou d'autres indicateurs phénotypiques que les COVs) plus facilement et/ou plus rapidement mesurables pour prédire cette même diversité est un enjeu pour le criblage de demain. Or les deux précédents examens de la singularité des souches ont démontré que l'expression protéique structurait la classification intégrative proposée et que cette dernière ne correspond pas à la classification à partir des variables phénotypiques.

Nous avons donc procédé à une sélection de variables par RF selon la classification supervisée des quatre groupes phénotypiques de souches. Le taux d'erreur de prédiction avant sélection est de 1,8 % (correspondant à un réplicat de la souche *DIA-A* incorrectement classé). Nous avons, dans cette approche, utilisé la totalité des données phénotypiques, génotypiques et protéomiques, soit 1157 variables pour procéder à la sélection.

Quatre-vingt-une variables ont été sélectionnées, un nombre similaire à celui obtenu pour la classification discriminante des souches avec le même protocole de sélection (voir ci-dessus). Cependant, il est observé un rééquilibrage à part équivalente entre la proportion de variables génotypiques et phénotypiques. En effet, ces 81 variables se décomposent en 8 distances génotypiques (une seule distance MLST liée au complexe clonal 2, et 7 distances génomiques à l'exclusion de celle de *MC70*), 39 indicateurs phénotypiques (Tableau III. 3) et 34 protéines (Tableau III. 4), soit respectivement une proportion pour chaque niveau de 42 %, 46 % et 3 %. Au sein des 39 variables phénotypiques, la grande majorité concerne les COVs, avec 34 variables identifiées réaffirmant le poids prépondérant de ces variables sur la diversité fonctionnelles des souches. En plus du lactate, la vitesse maximale d'acidification et la durée de fermentation ont été retenues pour compléter ce sous-ensemble de variables discriminantes des 4 orientations métaboliques majeures des souches.

Tableau III. 3 : Liste des trente-neuf variables phénotypiques sélectionnées lors de la recherche des marqueurs prédictifs de la diversité fonctionnelle à l'échelle intégrative. En gras et en italique, les variables déjà sélectionnées lors de l'analyse phénotypique seule.

Composés Organiques Volatiles		Acides organiques	Indicateurs technologiques
Acétaldéhyde	Acétone	Lactate	Durée de fermentation
Ethyl acétate	<i>Ethanol</i>	Acétate (HPLC-GC-MS)	Vitesse d'acidification, $V_{\max}^a$
Diacétyl	2-Heptanone		
2-Méthylpropanol	<i>Acétoïne</i>		
2-hydroxypentan-2-one	2-Nonanone		
Benzaldéhyde	<i>Butane-2,3-diol-D</i>		
Butane-2,3-diol-L	<i>Octanol</i>		
<i>Undecan-2-one</i>	<i>Acide butanoïque</i>		
Nonanol	Décanol		
2-Tridécan-2-one	<i>Acide hexanoïque</i>		
Diméthylsulfone	<i>Acide heptanoïque</i>		
Octalactone	2-Pentadecanone		
<i>Acide octanoïque</i>	Hexadécanal		
<i>Acide nonanoïque</i>	Décalactone		
<i>Acide décanoïque</i>	Acide décenoïque		
<i>δ-Dodécalactone</i>	Acide benzoïque		
Acide dodecanoïque	Heptanol		

Seules 9 protéines sont communes aux précédentes analyses de la diversité protéique des souches (YcgG, YcbH, YcjB, OptA, TagZ, Mae, CitF, YuiC et CitD), et aucun enrichissement catégoriel particulier n'a été détecté sur les 34 protéines (Tableau III. 4). A noter cependant, la présence de protéines du complexe de transport d'oligopeptides Opt (OptS, OptA et OptD) qui viennent renforcer la proportion d'enzymes liées à la nutrition azotée, avec la peptidase PepA, ou encore de l'acétyl-CoA carboxyltransférase, enzyme clef de la biosynthèse des acides gras. L'UDP-N-acétylglucosamine enolpyruvoyl transférase (MurA), impliquée dans la 1^{ère} étape de l'assemblage du peptidoglycane, a été également sélectionnée. Cette liste d'enzymes devra être étudiée plus en détails pour tenter d'établir des relations fonctionnelles entre les variables phénotypiques mesurées.

Tableau III. 4: Liste des trente-quatre protéines sélectionnées lors de la recherche des marqueurs prédictifs de la diversité fonctionnelle à l'échelle intégrative. En gras et en italique, les protéines déjà sélectionnées lors de l'analyse protéique de la diversité.

Gène	Catégorie	Protéine
<i>accD</i>	FAT	Acétyl-CoA carboxylase carboxyl transférase sous-unité β (EC 6.4.1.2)
<i>atpB</i>	NRJ	H ⁺ -ATPase, protéine de la sous-unité F0 (EC 3.6.3.14)
<i>mae</i>	NRJ	Malate oxidoréductase (EC 1.1.1.38)
<i>ycgG</i>	NRJ	Oxidoréductase
<i>citF</i>	NRJ	Citrate lyase alpha chaîne (EC 4.1.3.6)
<i>citD</i>	NRJ	Citrate lyase acyl-carrier protéine (EC 4.1.3.6)
<i>galM</i>	NRJ	Aldose 1-épimérase (EC 5.1.3.3)
<i>gltD</i>	AMI	Glutamate synthétase petite sous-unité (EC 1.4.1.13)
<i>ycjM</i>	PUR	Phosphatase
<i>dcdA</i>	PUR	Déaminase DCMF
<i>pdp</i>	PUR	Pyrimidine-nucléoside phosphorylase (EC 2.4.2.2)
<i>optA</i>	TSP	Oligopeptide ABC transporteur - protéine de liaison au substrat
<i>optS</i>	TSP	Oligopeptide ABC transporteur - protéine de liaison au substrat
<i>optD</i>	TSP	Oligopeptide ABC transporteur – protéine de liaison à l'ATP
<i>pepA</i>	TRD	Glutamyl aminopeptidase
<i>infC</i>	TRD	Facteur d'initiation de la traduction IF-3
<i>frr</i>	TRD	Facteur de recyclage des ribosomes (RRF)
<i>thiD2</i>	COF	Phosphométhylpyrimidine kinase (EC 2.7.4.7)
<i>ycbH</i>	ENV	Transférase de sucre
<i>murA2 (murZ)</i>	ENV	UDP-N-acétylglucosamine 1-carboxyvinyltransférase 2 (EC 2.5.1.7)
<i>tagZ</i>	ENV	Protéine de biosynthèse de l'acide téichoïque
<i>pgsA</i>	ENV	CDP-diacylglycérol-phosphate phosphatidyltransférase (EC 2.7.8.5)
<i>sbcC</i>	REP	Nucléase sbcCD sous-unité C
<i>yrfE</i>	REG	Régulateur transcriptionnel
<i>yjfl</i>	UNK	Protéine hypothétique L158463
<i>ykiE</i>	UNK	Protéine hypothétique L85764
<i>ysjD</i>	UNK	Protéine hypothétique L76848
<i>ycjB</i>	UNK	Protéine hypothétique L91807, putative glycoprotéine endopeptidase
<i>yqjE</i>	UNK	Putative N6-adenine-specific DNA méthylase
<i>yuiC</i>	UNK	Protéine hypothétique L68401
<i>ywfG</i>	UNK	Protéine hypothétique L39650

Cette sélection orientée de variables permet de retrouver l'intégralité des 4 sous-groupes phénotypiques, conformément à ce qui avait été choisi (sous-groupes I à IV) (Figure III. 6). En effet, le groupe I (*UCMA5713*, *S86*), le groupe II (*S87*, *EIP37F*), et la souche *DIA-A* du groupe III, forment des sous-groupes identiques à ceux obtenus à l'échelle phénotypique. Seule la souche *EIP41A* du groupe IV est regroupée avec le groupe II en lieu et place d'un positionnement avec les trois autres souches du groupe IV (*EIP33F*, *MC70* et *EIP33A*). Cependant, ce regroupement est tout juste significatif (p -value de bootstrap de 0,51), ce qui indique que la souche *EIP41A* aurait plutôt tendance à être singularisée dans un sous-groupe unique. Ce partitionnement des souches démontre que l'intégration des données protéiques et génomiques tend à rapprocher la souche *EIP41A* du groupe II, et que seules des différences de régulations dans l'activité de ces protéines à concentrations équivalentes peuvent expliquer les différences phénotypiques observées. Cette classification intégrative de la typicité aromatique a donc démontré son efficacité pour correspondre aux orientations aromatiques principales du métabolisme des souches : glycolyse, lipolyse, protéolyse ou équilibrée.

Ainsi, ces différentes classifications intégratives proposées soulignent la complexité à établir une unique classification fonctionnelle des levains. Chaque organisation est dépendante de la problématique biologique sous tendue. Cependant, la proximité des deux classifications « évaluation des marqueurs prédictifs de la diversité des souches » et classification « globale » permet de confirmer le potentiel prédictif des marqueurs, aussi bien phénotypiques que protéiques étudiés, pour décrire les singularités de chaque souche en lieu et place d'une étude de toutes les variables. La classification intégrée selon la typicité aromatique ouvre la voie à une analyse ciblée des potentialités aromatiques des souches.

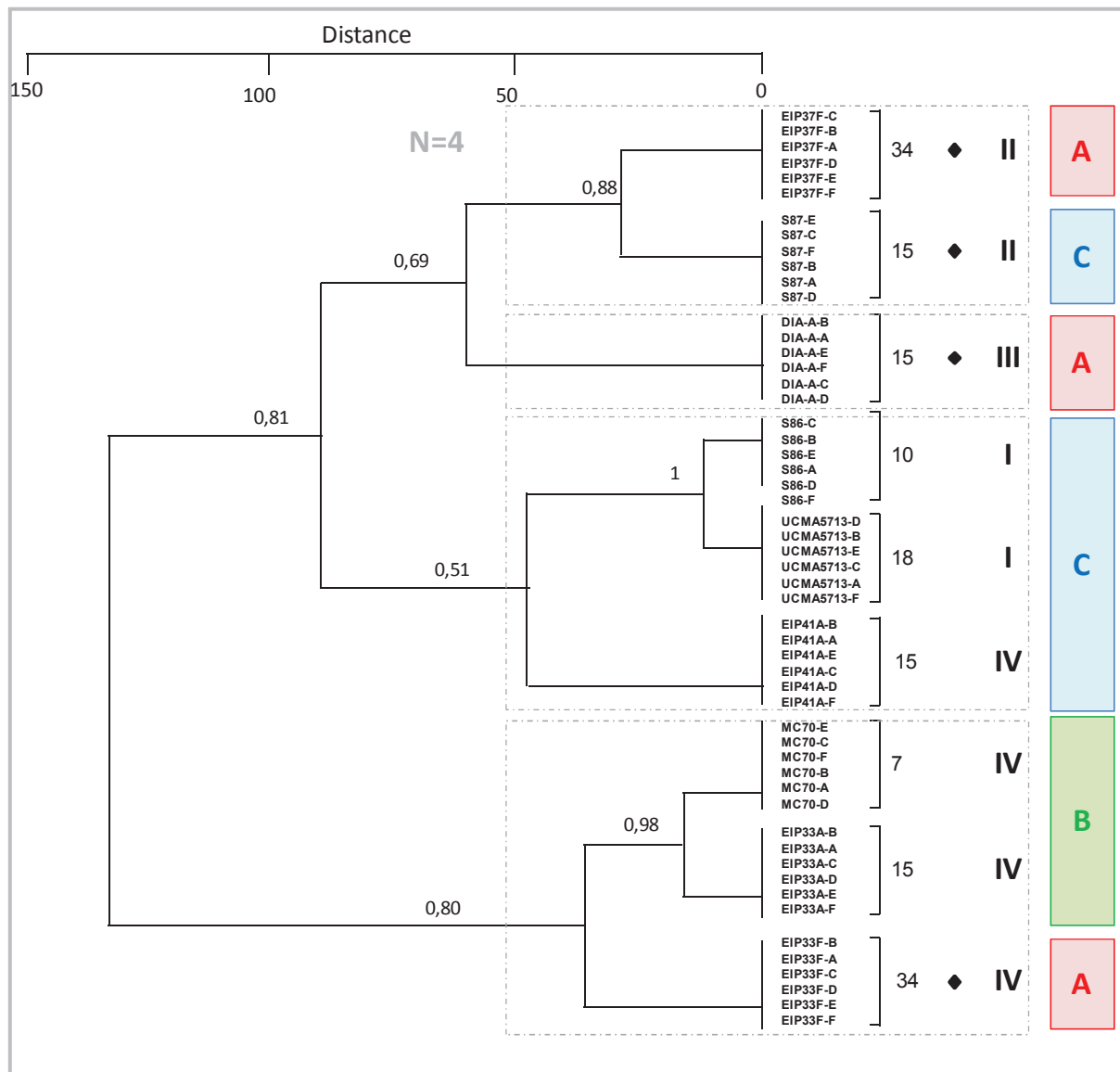


Figure III. 6 : Classification intégrative du génotype au phénotype des souches de *L. lactis* ssp. *lactis* après croissance en lait d'après les quatre groupes phénotypiques observés. La distance euclidienne et le critère de Ward ont été utilisés pour cette classification sur la base des 8 distances génétiques, 39 indicateurs phénotypiques et 34 protéines pré sélectionnés. De gauche à droite, le dendrogramme des souches, le nom des 54 expériences (neuf souches avec six réplicats de A à F, données protéiques complétées par la moyenne), STs de la MLST, présence du gène *citP* (losange noir), et numéro des quatre groupes phénotypiques (I à IV) et des trois groupes protéiques (A, B et C). Les valeurs de bootstrap sont indiquées pour chaque nœud (n=1000). La ligne verticale en pointillé décrit le niveau de coupe du dendrogramme pour interprétation.

II Positionnement protéique de la signature de *ND5F*

Ce second axe vise à affiner la description de la spécificité fonctionnelle de *ND5F* au sein de la diversité des autres souches. En particulier, l'étude de l'expression protéique de *ND5F* pourrait mettre en évidence des mécanismes explicatifs de sa singularité. De plus, l'ajout d'une nouvelle souche permettra de conforter l'analyse exploratoire de la diversité des levains et de vérifier notamment la robustesse des protéines discriminantes étudiées et d'identifier les protéines singulières à l'ajout de *ND5F*. Dans cette analyse, deux biovariants de *ND5F* de provenance différente, désignés *ND5F-Lab* et *ND5F-Danisco*, ont été analysés.

II.1 Identification des protéines spécifiques à *ND5F*

Pour identifier les protéines spécifiques à *ND5F*, une analyse fonctionnelle des protéines différenciellement exprimée entre le protéome de *ND5F* et les neuf autres souches a été menée. Cette approche a été également complétée par une sélection des protéines importantes selon une classification supervisée de type RF, comme présenté précédemment pour les autres souches.

II.1.1 Recherche des protéines différenciellement exprimées

En première intention, il a été ainsi recherché par analyse différentielle, les protéines discriminantes de *ND5F*. La souche *ND5F-Danisco* a été choisie comme référence. Un critère de sélection sur le ratio d'expression ou *Fold change*, *FC*, a été appliqué ($> 1,5$ ou $< -1,5$), suivi de l'ajustement des p-values selon la méthode BH [Benjamini & Hochberg 1995]. La limite de significativité a été imposée à FDR 10 %. De plus, afin d'éviter de commenter des modifications d'expression de protéines très faiblement abondantes et potentiellement proches du bruit de fond, il a été décidé d'éliminer des comparaisons deux à deux les protéines dont la quantification était inférieure au quantile 0,1 dans les deux cas. Pour faciliter l'interprétation des résultats, chaque protéine différenciellement exprimée a été replacée dans sa catégorie fonctionnelle, et un test d'enrichissement a été réalisé (Tableau III. 5). Dans un premier temps, il a été comparé les deux biovariants de *ND5F* entre eux. Une seule protéine, la préphénate déshydratase (*pheA*), impliquée dans la biosynthèse des acides aminés aromatiques, a été identifiée comme différenciellement sous-exprimée dans la version *Lab* de la souche. A l'exception de cette enzyme, il peut être considéré que les protéomes de ces deux biovariants sont similaires dans les conditions d'expression étudiées. Ce constat confirme, d'une part la qualité de l'analyse, d'autre part la proximité phénotypique également observée entre ces deux biovariants (résultats non présentés).

Tableau III. 5 : Examen des catégories fonctionnelles différentiellement sur ou sous-représentées pour les comparaisons entre le protéome de la souche *ND5F-Danisco* et les 10 autres souches de l'étude. Pour chaque comparaison, le nombre de protéines différentiellement (sur/sous)-exprimées pour la souche considérée dans chaque catégorie fonctionnelle est mentionné (FDR $\leq 10\%$) (voir Tableau III.1, détails des catégories). Les catégories significativement enrichies sont désignées en fonction de la p -value d'enrichissement selon le code suivant : * $<0,1$ ** $<0,05$ et *** $<1.10^{-2}$.

Catégorie fonctionnelle	<i>ND5F-Lab</i>		<i>UCMA5713</i>		<i>S86</i>		<i>S87</i>		<i>EIP37F</i>		<i>EIP41A</i>		<i>EIP33F</i>		<i>MC70</i>		<i>DIA-A</i>		<i>EIP33A</i>	
	Sous	Sur	Sous	Sur	Sous	Sur	Sous	Sur	Sous	Sur	Sous	Sur	Sous	Sur	Sous	Sur	Sous	Sur	Sous	Sur
Protéines différentielles	1	0	58	117	66	163	49	133	195	64	62	148	273	32	551	32	136	80	672	40
AMI	1*	0	5	4	5	5	4	4	8	0	6*	7	10	0	24	2	5	7*	32	40
CEL	0	0	0	0	0	2	0	0	2	3	0	2	6	1	17*	1	0	2	17	1
COF	0	0	0	3	1	4	1	3	7	1	1	5	7	1	16	1	1	4	17	0
ENV	0	0	4	5	5	6	5*	7	12	2	6*	6	15	1	28	1	10	7	37	0
FAT	0	0	0	3	2	7**	0	3	4	1	0	4	5	0	13	1	3	1	17*	0
INT	0	0	0	2	1	2	0	3*	3	1	0	3*	3	1	5	1	1	1	6	1
NRJ	0	0	3	15	2	15	2	20	20	4	1	16	31	5	67	2	16	10	80	4
OTH	0	0	4	3	1	5	1	1	7	5	4	7	10	2	20	2	7	2	23	1
PUR	0	0	3	9	2	12	2	11	10	3	4	15***	21**	3	38**	2	11	7	37	6*
REG	0	0	3	8	5	6	2	4	12	3	3	7	11	2	33	2	4	3	31	3
REP	0	0	4	5	4	5	3	5	10	5*	4	5	14	1	18	1	6	1	28	1
TRD	0	0	6	20**	7	38***	5	22**	21	10	6	22	34	7**	81***	3	18	9	92**	6
TRS	0	0	1	2	1	4	1	2	5	1	7	2	9	1	15	1	2	2	16	2
TSP	0	0	5	6	6	12	4	9	17	7	7	11	26	2	38	3	10	3	46	6
UNK	0	0	20	32	24	40	9	39	7	18	19	36	71	5	138	9	42	21	193	9

Entre 175 (*UCMA5713*) et 790 (*EIP 33A*) protéines différentiellement exprimées ont été identifiées par rapport à *ND5F-Danisco* lors des comparaisons des différents protéomes, ce qui démontre l'importance de la singularité protéique de la souche *ND5F* (Tableau III. 5). Cependant, à partir de cette étude, il est difficile d'identifier une catégorie de protéines systématiquement sur ou sous-exprimées par rapport au protéome de *ND5F*. Toutefois, les protéines impliquées dans la traduction et dans le métabolisme des purines sont significativement enrichies lors de la comparaison à respectivement six et quatre protéomes. Mais ces résultats révèlent des tendances hétérogènes en fonction des souches analysées, avec une sous-expression des protéines de la traduction dans *ND5F-Danisco* par rapport aux souches *UCMA5713*, *S86*, *S87* et *EIP33F* alors que c'est le phénomène inverse qui est observé lors de la comparaison à *MC70* et *EIP33A*.

Par conséquent, pour dégager des marqueurs protéiques de la singularité de *ND5F*, nous avons recherché les protéines communément sur et sous-exprimées entre toutes les souches (à l'exception de *ND5F-Lab*). Alors qu'aucune protéine commune sous-exprimée n'a pu être identifiée, une liste de huit protéines significativement surexprimées sur le protéome de *ND5F-Danisco* par rapport aux autres souches a été déterminée (Tableau III. 6).

Tableau III. 6 : Protéines significativement sur-exprimées dans le protéome de *ND5F-Danisco* par rapport aux autres souches.

Gène	Catégorie	Protéine	Ratio de surexpression : min-médian-max
<i>qor</i>	NRJ	Quinone oxidoréductase	2,2-4,5-20,5
<i>aroB</i>	AMI	3-déhydroquinate synthétase (EC 4.2.3.4)	2,3-4,2-8,1
<i>rplT</i>	TRD	Protéine ribosomale 50S L20	2,5-5,5-16,6
<i>lirE</i>	REG	Régulateur du système à deux composants, <i>lirE</i>	11,6-77,2-304,7
<i>rpmIB</i>	PUR	DTDP-glucose 4,6-déshydratase	3-4,8-12,6
<i>copB</i>	TSP	Transporteur ATPase du cuivre et potassium	10-16,4-71,9
<i>yjiB</i>	UNK	Protéine inconnue, L81441, homologie avec l'hémolysine	2,1-69,9-19,4
<i>ychG</i>	UNK	Protéine hypothétique L78434	5,1-7,9-29,4

Parmi ces protéines, il a été identifié le régulateur transcriptionnel *lirE* de l'un des systèmes à deux composants de la famille *OmpR*, impliqué dans la régulation de l'activité phosphate [O'Connell-Motherway et al. 2000]. Dernièrement, l'expression de ce gène a été identifiée dans la réponse cœur au stress de *L. lactis*, et notamment à la carence en isoleucine, avec une expression négativement contrôlée par le taux de croissance dans différentes conditions [Dressaire et al. 2008]. Le régulon de ce régulateur reste cependant à déterminer. Il est également présent au sein de cette liste, la 3-déhydroquinate synthase (DHQ synthase), une métallo enzyme impliquée dans la biosynthèse de chorismate qui conduit à la production d'acides aminés aromatiques. La fonctionnalité de ces huit

protéines nécessite d'être étudiée plus en détails pour appréhender les conséquences physiologiques de cette surexpression sur le phénotype observé.

II.1.2 Sélection des variables discriminantes

En complément de l'approche différentielle, une classification supervisée en 11 classes (deux classes pour *ND5F* en fonction du biovariant) a été réalisée à partir des 1057 protéines par RF. L'analyse sans a priori de la position des deux souches lors de l'analyse des signatures nous permettra de vérifier la proximité des deux biovariants. A partir de cette analyse, 54 protéines ont pu être identifiées lors de la sélection des variables les plus importantes. La comparaison de ce jeu de protéines à celui précédemment identifié sans *ND5F* (47 protéines, Tableau III. 2), révèle un cœur de 30 protéines discriminantes communes à ces deux classifications (Figure III. 7 A). Ainsi, ces protéines apparaissent comme robustes pour décrire la diversité d'expression protéique des levains de *L. lactis* en lait. Dans ce sous-ensemble commun, il est dénombré un tiers de protéines appartenant au métabolisme énergétique (enrichissement avec une p -value de $1,21.10^{-3}$). Ces protéines sont listées ci-après : acétoïne réductase (*butA*), malate oxydoréductase (*mae*) et pyruvate déshydrogénase (*pdhA*) du métabolisme fermentaire, trois des sous unités de la citrate lyase (*citCDF*), la phosphogluconate déshydrogénase (*gnd*) de la voie des pentoses phosphate, la glutamate décarboxylase (*gadB*), les oxydoréductases pyruvate-flavodoxine (*nifJ*) et YcgG. Le second groupe le plus abondant au sein de ces 30 protéines concerne les protéines impliquées dans la traduction, avec trois protéines ribosomales 50S L10-L11-L16 (*rplJ*, *rplK*, *rplP*), l'isoleucyl ARNt synthase (*ileS*) et l'aminopeptidase PepC (*pepC*). La sérine méthylase (*glyA*), la cystathionine- β -lyase, CBL (*metB2*), tout comme les deux exonucléases RexA et UvrB, et la protéine PurK, ont été également identifiées comme appartenant à cette liste. La sous-unité A du transporteur d'oligopeptides Opt ainsi que l'ATPase échangeuse de cations YfgQ ou encore la protéine d'initiation de la division cellulaire DivIVA, viennent s'ajouter aux sept gènes de protéines de fonctions inconnues (*ybjB*, *ycjB*, *yfbB*, *ygiK*, *yhjB*, *yneC*, *yuiC*) pour compléter ce sous-ensemble. Ces protéines apparaissent donc comme des éléments structurants de la diversité protéique des levains et semblent relativement robustes à l'ajout de deux nouvelles classes.

En étudiant les corrélations qu'elles partagent avec les 47 protéines discriminantes lors de l'analyse de la diversité protéique des 9 souches, les 24 protéines nouvellement sélectionnées suite à l'addition des deux biovariants de *ND5F* peuvent-être scindées en deux sous-groupes (Figure III. 7 B et C). En effet, il a été observé que 10 de ces 24 protéines partagent des corrélations avec les autres protéines spécifiques de la sélection sans *ND5F* (coefficients de corrélation entre 0,70 et 0,86, p -value < 0,05). Ainsi, bien que les protéines soient différentes, notamment pour certaines en termes de fonction, leurs profils d'expression sur les neuf souches précédentes apparaissent similaires, ce

qui sous-entend un potentiel comparable de discrimination des 9 souches. Cependant, s'il est étudié le positionnement spécifique des réplicats de *ND5F*, il est remarqué que les deux biovariants se singularisent plus sur ces dix protéines spécifiques à *ND5F* que sur les précédentes protéines sélectionnées sans *ND5F* (voir exemple entre *YwjA* et *GmK*) (Figure III. 7 B). Ainsi, le plus fort degré de discrimination de *ND5F* des autres souches sur ces 10 protéines explique vraisemblablement le choix de ces protéines lors de la procédure de sélection en lieu et place aux 17 autres protéines. Le second groupe correspond à 14 protéines qui ne présentent pas de corrélations avec les 47 autres protéines du jeu sans *ND5F*, et sur lesquelles les protéomes de *ND5F* se distinguent par une abondance maximale ou minimale (voir exemple de la protéine *AroB*, Figure III. 7 C). Par conséquent, la sélection de ces protéines participe activement à singulariser la signature protéique de *ND5F*. Dans le cadre de cette étude, leur fonctionnalité va être étudiée en détail (Tableau III. 7).

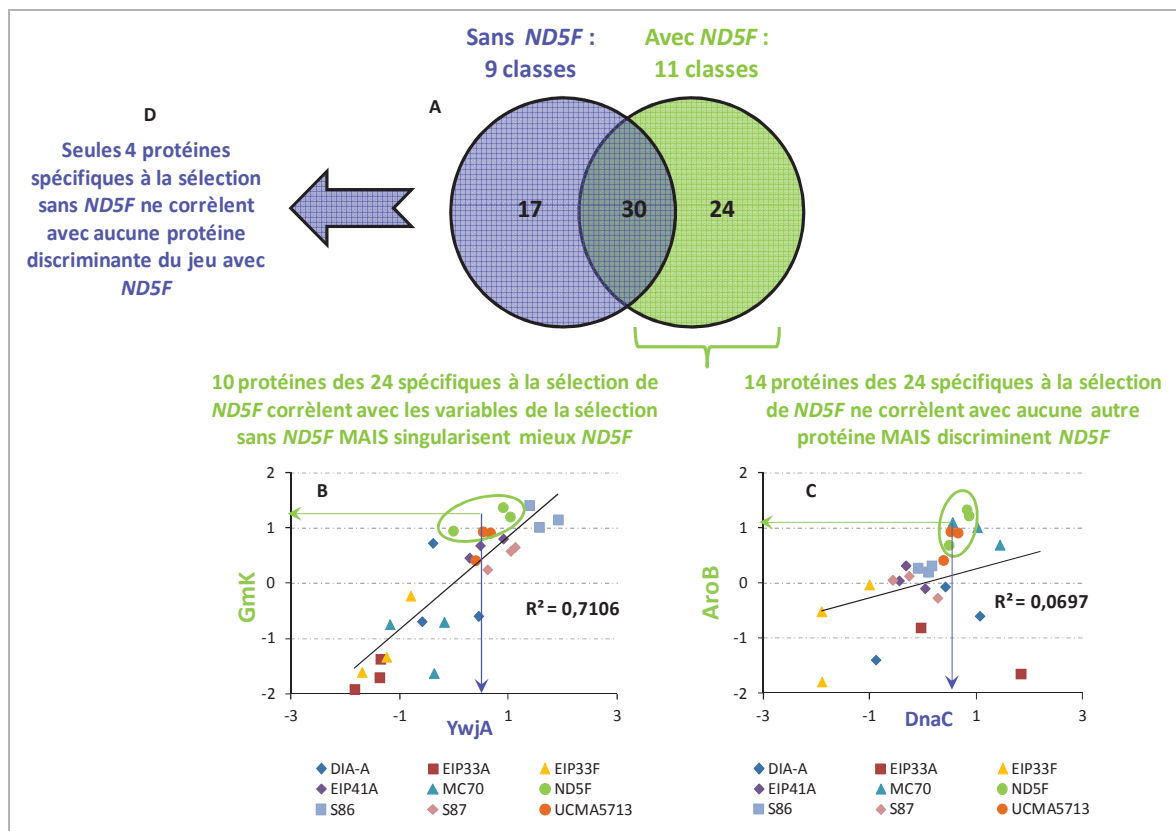


Figure III. 7 : Comparaison des différents jeux de protéines discriminantes sélectionnées en fonction du nombre de souches à classer. A) Diagramme de Venn pour comparer les protéines discriminantes identifiées sans et avec ajout de *ND5F*. B) Recherche des corrélations entre les protéines spécifiques à la sélection de *ND5F* et celles sans *ND5F* : exemple de corrélation entre deux protéines (*YwjA* sans *ND5F* et *GmK* avec *ND5F*). C) Analyse du positionnement de *ND5F* sur les protéines spécifiques de la classification avec *ND5F* : exemple de deux protéines avec et sans *ND5F*. D) Analyse des corrélations entre les protéines spécifiques à la sélection sans *ND5F*. Note : les quantités relatives ont été centrées et réduites pour s'affranchir de l'abondance des protéines et comparer les profils d'expression.

Tableau III. 7 : Analyse fonctionnelle des vingt-quatre protéines spécifiques de la classification supervisée des onze protéomes dont *ND5F*. Organisation des protéines en quatre classes selon les profils d'expression après classification ascendante hiérarchique. Il est indiqué pour chaque protéine le rang relatif minimal, médian et maximal au sein des 11 protéomes, et le rang médian spécifique pour *ND5F* (moyenne des biovariants *ND5F-Danisco* et *Lab*). Les protéines différentielles surexprimées sont surlignées par une bande grise.

Classe	Gène	Catégorie	Protéine	Rang (%) :min-médiane-max	Rang médian pour <i>ND5F</i> (%)
Protéo-NDI	<i>ptsI</i>	TSP	Phosphoénolpyruvate phosphotransférase enzyme E1 (EC 2.7.3.9)	87-93-96	96
	<i>serC</i>	AMI	Phosphoserine aminotransférase (EC 2.6.1.52) (PSAT)	40-49-65	65
	<i>qor</i>	NRJ	Quinone oxidoréductase	46-59-77	77
	<i>yjaB</i>	UNK	Protéine universelle de stress	17-29-59	40
	<i>purQ</i>	PUR	Phosphoribosylformylglycinamide synthase I	58-76-84	68
	<i>gmk</i>	PUR	Guanylate kinase	27-36-47	42
	<i>frr</i>	TRD	Facteur de recyclage des ribosomes (RRF)	57-75-81	81
	<i>oppA</i>	TSP	Protéine de liaison au substrat de l'oligopeptide Opp (OppA)	90-95-99	99
Protéo-ND II	<i>osmC</i>	OTH	Protéine induite par le stress osmotique ou oxydatif	51-63-79	71
	<i>rheA</i>	TRS	ARN hélicase ATP dépendante	60-77-84	82
	<i>llrE</i>	REG	Régulateur du système à deux composants, IlrE	1-2-21	21
	<i>yngE</i>	TSP	Transporteur ABC des sucres-protéine de liaison à l'ATP	47-60-75	70
	<i>aroB</i>	AMI	3-déhydroquinate synthase (EC 4.2.3.4)	38-51-66	66
	<i>rplT</i>	TRD	Protéine ribosomale 50S L20	79-87-97	97
	<i>rpsK</i>	TRD	Protéine ribosomale 30S S11	59-70-89	81
	<i>rplB</i>	TRD	Protéine ribosomale 50S L2	75-91-97	96
	<i>rplD</i>	TRD	Protéine ribosomale 50S L4	80-82-99	99
	<i>cbiO2</i>	TSP	Transporteur ABC - protéine de liaison à l'ATP	23-30-72	46
Protéo-ND III	<i>tagD1</i>	ENV	Glycérol-3-phosphate cytidiltransferase	12-25-57	37
	<i>xseA</i>	REP	Probable exodésoxyribonucléase VII grande sous-unité (EC 3.1.11.6)	54-63-94	89
	<i>yjiB</i>	UNK	Protéine inconnue, L81441, homologie avec des domaines d'hémolysine	30-45-82	82
	<i>adk</i>	PUR	Adénylate kinase (EC 2.7.4.3)	88-92-100	98
Protéo-ND IV	<i>ysaC</i>	TSP	Transporteur ABC -protéine de liaison à l'ATP	4-9-44	4
	<i>guaC</i>	PUR	GMP réductase (EC 1.6.6.8)	84-91-93	84

L'analyse globale des 54 protéines (30 communes et 24 protéines spécifiques à l'ajout de *ND5F*) confirme un enrichissement en protéines du métabolisme énergétique (p -value $3,4.10^{-2}$). Elles s'organisent en quatre classes principales à partir de leur profil d'expression (Protéo-ND I à Protéo-ND IV). Les 24 protéines spécifiques à l'ajout de *ND5F* ainsi que leur organisation dans les quatre classes principales sont présentées (Tableau III. 7). Parmi elles, nous retrouvons cinq des huit protéines différentiellement surexprimées précédemment identifiées chez *ND5F* par rapport aux neuf protéomes, preuve du recoupement et de la complémentarité des approches statistiques menées.

Sur les 54 protéines importantes identifiées, les classes Protéo-ND I et Protéo-ND II concentrent les protéines pour lesquelles l'expression est maximale au sein des protéomes de *ND5F*. La classe Protéo-ND I regroupe notamment des protéines impliquées dans l'assimilation des sucres, avec l'enzyme cytoplasmique E1 du système PTS, le métabolisme des nucléotides PurQ et Gmk ou la nutrition azotée avec le transporteur Opp. Il est retrouvé, dans cette classe, la quinone oxidoréductase, en charge de la production d'énergie par transfert d'électrons, également identifiée comme l'une des protéines clefs de la singularité de *ND5F* par analyse différentielle. La classe Protéo-ND II se distingue par un enrichissement en protéines impliquées notamment dans la traduction (p -value $3,4.10^{-2}$). La comparaison de la distribution moyenne de ces protéines sur les onze classes révèle une surexpression globale chez *ND5F* de ces protéines. Parmi les 14 protéines spécifiques à *ND5F*, à noter la présence des trois protéines différentiellement exprimées LlrE, déhydroquinase synthase et la protéine ribosomale 50S L20. Il est également présent au sein de cette classe, la protéine OsmC inductible par la présence de Na^+ (liaison au régulateur NhaR) et également engagée dans la défense au stress oxydatif causée par l'exposition à des hydroperoxydes organiques [Lesniak et al. 2003]. Deux transporteurs ABC ont également été attribués à cette classe (gènes *ynpE*, *chiO2*). La classe Protéo-ND III, significativement enrichie en protéines impliquées dans la réplication (p -value $3,4.10^{-2}$), comprend 6 protéines dont 4 sont spécifiques à la sélection avec *ND5F* avec des niveaux d'expression intermédiaires pour ces protéines (glycérol-3-phosphate cytidyltransférase, exodésoxyribonucléase VII, l'adénylate kinase et la protéine inconnue L81441). La dernière classe, Protéo-ND-IV, regroupe 11 des 54 protéines, majoritairement des enzymes appartenant au groupe des protéines communes à l'exploration de la diversité protéique des levains (9/11) et deux autres protéines, un transporteur ABC (largement sous-exprimée par *ND5F*) et la GMP réductase.

II.2 Positionnement de la signature de *ND5F*

Pour étudier plus en détail le positionnement de *ND5F* (et notamment des différentes versions des biovariants), une ACP suivie d'une classification ascendante hiérarchique ont été réalisées sur ce jeu de 54 protéines sélectionnées (**Erreur ! Source du renvoi introuvable.**). L'analyse conjointe de l'ACP et de la classification permet d'identifier quatre sous-groupes de souches (groupes A-B-C et D), dont trois sont communs à la précédente analyse de la diversité protéique des levains (**Erreur ! Source du renvoi introuvable.** A, B et C). Le quatrième groupe, le groupe D, est formé exclusivement des répliquats biologiques de la souche *ND5F*, démontrant ainsi la singularité de l'expression protéique de cette souche. De plus, le regroupement des deux biovariants au sein d'un même groupe confirme leur proximité protéique. L'un des répliquats de la souche *S86* est incorrectement classé avec ceux de la souche *UCMA5317*. Cette relative perte de singularité protéique des souches les plus proches, visible également lors de la comparaison des ACPs (Figure III. 4 et Figure III. 8), s'explique probablement par la sélection préférentielle de protéines discriminantes de *ND5F* lors de la classification supervisée par RF.

La localisation de *ND5F* sur le graphique des individus témoigne d'une singularité liée à une position maximale le long de l'axe 2, et positive sur l'axe 1 (Figure III. 8 B). Cette disposition coïncide en particulier aux protéines de la classe Protéo-ND II qui semblent être tout particulièrement discriminantes pour ces deux souches (Figure III. 8 C). Il est possible de rajouter les protéines de la classe Protéo-ND III du fait de leur contribution non négligeable à la seconde dimension, en particulier l'exodésoxyribonucléase VII (gène *xseA*) et l'exonucléase *RexA*. Ces protéines, ainsi que les huit enzymes différenciellement surexprimées, définissent la singularité à l'échelle protéique de la souche. Cependant, le manque d'information fonctionnelle sur ces protéines empêche la projection des conséquences physiologiques de ces modifications d'expression.

Représentation des individus sur les dimensions 1 et 2 (A) avec mention des quatre sous-groupes de souches A-B-C-D. Les ellipses de couleur désignent les intervalles de confiance à 95 % autour des différents réplicats des souches. C) Les 54 protéines sont représentées par leur nom de gènes selon les dimensions 1 et 2, les ellipses en pointillé matérialisent les différentes classes Protéo-ND-I, II-III-IV.

CONCLUSION

Cette étude présente, pour la première fois, une caractérisation comparative de la diversité protéique de levains de *L. lactis* ssp. *lactis* après une croissance en lait et en condition industrielle (matière protéique et matière grasse non nulles). Les 1057 protéines identifiées et quantifiées constituent le jeu le plus complet, à l'heure actuelle, pour examiner l'expression protéique différentielle de ces souches. Ainsi, les protéines cœur de la réponse protéique adaptative en lait ont pu être identifiées lors de la phase stationnaire à pH acide. Parmi les protéines les plus abondantes, sont retrouvées les enzymes impliquées dans les étapes de traduction, qui représentent à elles seules un tiers de ce pool (protéines ribosomales et les ARNt-amino acyl synthases) ou encore les peptidases intracellulaires de la nutrition azotée. Les autres catégories surreprésentées comprennent les enzymes appartenant au métabolisme énergétique, avec l'intégralité des enzymes de la glycolyse, mais également les protéines du métabolisme des nucléotides.

Après la sélection d'un cœur d'une trentaine de protéines (protéines communes à l'analyse sans et avec ND5F, jugées comme robustes pour décrire la diversité d'expression protéique), nous avons pu mettre en évidence une singularité protéique pour chacune des neuf souches étudiées. Cette diversité protéique est en accord avec le haut degré de singularité phénotypique précédemment observé pour chaque levain en lait. Cependant, il est également possible d'identifier une structuration de cette sous-population en trois sous-groupes principaux (groupes A, B et C) en fonction de leur profil d'expression sur les quatre classes de protéines identifiées (Protéo-I à IV). Ces protéines comprennent notamment un enrichissement en enzymes du métabolisme énergétique, impliquées dans la production d'acétolactate, diacétyle/acétoïne et butanediol. Les enzymes de dégradation du citrate sont également présentes (CitC, CitD et CitF), tout comme celles participant aux étapes de nutrition azotée (peptidases et transporteurs d'oligopeptides). Un groupe de protéines interagissant dans la réponse aux différents stress (acide, oxydatif et osmotique) a été également sélectionné. Cependant, la nécessité de procéder à une sélection de variables pour visualiser ces singularités prouve que cette diversité d'expression protéique est limitée à un nombre restreint de marqueurs protéiques. La diversité protéique des souches est donc un phénomène minoritaire au sein de l'expression protéique.

Bien que des recoupements entre marqueurs protéiques et signatures phénotypiques aient pu être réalisés pour les deux souches *DIA-A* et *S87*, l'organisation de la diversité protéique des souches diverge fortement des 4 principaux groupes phénotypiques (I à IV) et de la diversité génotypique. Ce résultat atteste d'un manque de corrélation directe entre la phylogénie, les concentrations protéiques et la régulation de l'activité cellulaire aboutissant au phénotype. Afin de concilier ces trois niveaux de diversité génotype-protéique-phénotype, plusieurs études intégratives selon la problématique biologique (diversité des 9 souches ou typicité aromatique en 4 sous-groupes) ont été menées pour proposer une classification prédictive en adéquation avec la fonctionnalité des levains. En particulier, l'analyse conjointe des 1157 variables de ces trois niveaux de diversité, dite analyse « globale », a permis d'apprécier, pour chaque levain, les potentialités exprimées (à l'échelle protéique ou phénotypique) en accord avec leur singularité génotypique. L'adéquation de cette classification avec celle réalisée à partir des marqueurs sélectionnés lors de l'analyse indépendante de chaque niveau de diversité (intégralité des distances génétiques, 47 protéines et les 20 COVs) a par ailleurs validé la pertinence de ces marqueurs identifiés pour prédire la diversité du génotype au phénotype des 9 souches. Ces deux classifications intégrées démontrent l'importance des variables protéiques comme étant les principaux éléments structurants de la diversité des souches, tandis que l'apport des variables génotypiques et phénotypiques accentue leur singularité.

Dans un but applicatif, il a été également choisi de proposer une classification intégrative en adéquation avec la typicité aromatique des souches et non plus avec la singularité des souches. Pour cela, une sélection des variables génotypiques, protéiques et phénotypiques selon la supervision des 4 groupes phénotypiques précédemment observés a permis d'isoler 81 variables, majoritairement phénotypiques et protéiques, comme prédictives de cette diversité fonctionnelle ciblée. Cette dernière classification intégrée constitue un outil d'aide à la sélection mais également à la compréhension de la typicité des levains.

Le positionnement de *ND5F* au sein de cet échantillon de levains a révélé le haut degré de singularité de cette souche présentant, à minima, près de 175 protéines différenciellement exprimées et formant à elle seule un sous-groupe distinct lors de l'analyse multi-variables. L'analyse protéique a confirmé la proximité des deux biovariants *Danisco* et *Lab* de la souche (1^{er} regroupement en classification ou encore présence d'une seule protéine différenciellement exprimée). Le croisement des méthodes statistiques nous a permis d'identifier un sous-ensemble d'une quinzaine de protéines surexprimées et discriminantes de cette singularité. En particulier, nous citerons des protéines ribosomales dont les conséquences physiologiques restent à élucider, et le régulateur transcriptionnel du système à deux composants LlrE.

CHAPITRE IV :

LEVIER ET VERROUS DE LA PRODUCTION
AROMATIQUE DE *L. LACTIS* SSP. *LACTIS ND5F*

INTRODUCTION

Afin de contrôler la production aromatique de *L. lactis ND5F*, il est nécessaire d'analyser les différents paramètres du procédé (nutritionnels et environnementaux) pouvant influencer cette signature aromatique. Ainsi, ce chapitre vise à mettre en évidence les leviers et verrous de cette production d'arômes d'un point de vue technique, mais également à un niveau métabolique et protéomique.

Cette étude a été scindée en deux volets. Une première partie de l'analyse a été réalisée sur Milieu Chimiquement Défini (MCD) afin de faciliter en particulier l'identification des paramètres impactant la production de pentane-2,3-dione. L'étude de paramètres nutritionnels fins (tels que la concentration d'un composé) est facilitée sur ce milieu de composition connue, et les outils de suivi de biomasse, et d'analyse du métabolisme et de la protéomique sont disponibles.

La seconde partie, plus applicative, s'est articulée sur l'étude globale de l'évolution de la signature de *L. lactis ND5F* en milieu industriel en fonction de quatre paramètres environnementaux sélectionnés (matière grasse, sèche, oxygénation et température), par une approche de plan d'expériences. Elle a abouti à la définition d'un procédé modifié susceptible de renforcer les caractéristiques organoleptiques particulières de ce produit.

I Approche mécanistique du contrôle de la production de pentane-2,3-dione

Bien que la singularité de *L. lactis* ssp. *lactis* ND5F soit liée à une somme complexe de déterminants phénotypiques, l'un des composés clef de la typicité aromatique est la pentane-2,3-dione en raison de l'adéquation des notes aromatiques « beurre » et « crémeuse » de cet arôme avec les produits fermentés. Par conséquent, la production de ce métabolite a fait l'objet d'une étude approfondie afin d'identifier les différents paramètres pouvant l'influencer tant au niveau du procédé que des déterminants mécanistiques. Pour faciliter cette analyse, les essais ont été conduits en Milieu Chimiquement Défini (MCD).

I.1 Paramètres environnementaux sélectionnés

Le choix des différents paramètres nutritionnels et physico-chimiques a été réalisé à partir d'une synthèse bibliographique des voies de production de la pentane-2,3-dione, afin de sélectionner un jeu de conditions susceptible de l'impacter. Ainsi, la pentane-2,3-dione est formée dans le milieu extracellulaire suite à la décarboxylation oxydative non enzymatique et spontanée d' α -acéto-2-hydroxybutanoate [Imhof et al. 1995; Landaud et al. 1998; Ott et al. 1999; Ott et al. 2000]. Cet intermédiaire de la voie de biosynthèse de l'isoleucine est produit par l'acétolactate synthase anabolique nouvellement appelée acétohydroxyacide synthase (abrégée AHAS, *ilvBN*) à partir du pyruvate et de l'oxobutanoate [Godon et al. 1992; Goupil-Feuillerat et al. 1997]. Deux voies ont été distinguées pour la production d'oxobutanoate. La première, identifiée chez *L. lactis*, est la dégradation de la L-thréonine par la thréonine amonia-lyase ou thréonine désaminase (abrégée TD, *ilvA*) [Bell & Turner 1977; Sawers 1998] (Figure IV. 1). La seconde voie, mise en évidence chez *E. coli* W et *E. coli* ATCC 8739, repose sur la présence d'une activité β -méthylaspartate aminotransférase. Cette enzyme serait capable de transférer le groupement amine du 3-méthylaspartate à l'oxoglutarate pour coproduire du L-glutamate et du méthylxaloacétate [Abramsky et al. 1965; Phillips et al. 1972]. Ce dernier peut alors subir une décarboxylation spontanée en oxobutanoate. Le méthylaspartate peut-être synthétisé à partir de L-glutamate (nécessite une glutamate mutase, *mamA*) ou de pyruvate et d'acétate (nécessite une citrate malate synthase, *CimA*). Cette voie de production a été postulée au sein de cultures de *S. thermophilus* et *L. bulgaricus* pour expliquer le marquage en C13 de la pentanedione à partir de pyruvate et d'acétate [Ott et al. 2000]. L'absence d'identification d'une séquence protéique empêche la recherche de cette activité sur le génome de ND5F. De plus, aucune évidence n'existe à notre connaissance quant à la présence de cette activité pour des souches de *L. lactis* sp. *lactis*.

I.1.1 Etude bioinformatique des gènes impliqués dans la production et/ou la régulation de la pentane-2,3-dione

A partir des données de séquençage, il a pu être mis en évidence que tous les gènes impliqués dans la voie de biosynthèse des acides aminés branchés étaient présents sur le chromosome de *ND5F* au sein des opérons *leuABCD* et *ilvDBNCA*, ou isolés, comme le gène *ilvE* codant l'aminotransférase des BCAAs. Cependant, il est connu que la majorité des souches laitières de *L. lactis* ssp. *lactis* possèdent ces gènes mais présentent de nombreuses auxotrophies pour ces acides aminés branchés, à la différence des souches environnementales [Chopin 1993; Godon et al. 1993; Siezen et al. 2008; Siezen et al. 2011]. Cette auxotrophie peut être liée, soit à des mutations de ces gènes, soit à une expression inadéquate. Pour expliquer cette auxotrophie, des comparaisons des séquences nucléotidiques des gènes *leuABCD* et *ilvDBNCA* entre des souches prototrophe (*NCDO 2118*) et auxotrophe (*IL1403*) de *L. lactis* ssp. *lactis* ont été réalisées [Godon et al. 1993]. Les travaux de Godon ont révélé que l'auxotrophie était probablement liée à des mutations ponctuelles qui rendaient les produits de l'expression de ces gènes inactifs. En particulier, deux mutations non-sens au sein des gènes *leuA* et *leuB* ont été identifiées aboutissant à des protéines tronquées de, respectivement, 104 (409 sur 513) et 25 (320 sur 345) acides aminés [Godon et al. 1993].

Dans un premier temps, il a été décidé de déterminer si la souche *ND5F* était prototrophe ou auxotrophe vis-à-vis des BCAAs par une étude de croissance sur milieu chimiquement défini MCD [Cocaign-Bousquet et al. 1995] avec et sans BCAAs. Il s'est avéré qu'après 72h de culture, aucune croissance n'a été observée sur le milieu MCD sans BCAAs, alors que 1 g.L⁻¹ de biomasse a été mesuré sur MCD avec BCAAs, indiquant que *ND5F* est auxotrophe pour les acides aminés branchés. Des essais complémentaires ont été également conduits avec une supplémentation du milieu MCD uniquement en isoleucine, ou encore en valine-leucine, pour évaluer le, ou les, acides aminés responsables de cette auxotrophie. Dans tous les cas, aucune croissance n'a pu être observée, ce qui semblerait indiquer que *ND5F* est auxotrophe pour l'isoleucine et la valine-leucine. Cependant, il ne peut pas être exclu que le déséquilibre des apports en BCAAs dans ces conditions ait inhibé l'expression des gènes relatifs aux acides aminés absents. Pour tenter de préciser quelles sont les étapes limitantes de cette voie de biosynthèse, les séquences nucléotidiques de *ND5F* ont été comparées à celles de la souche *NCDO2118* afin d'évaluer la fonctionnalité des produits des gènes de notre souche d'intérêt, de manière analogue à l'étude de Godon (Figure IV. 1). Les séquences des gènes *ilvA*, *ilvE* et *leuD* sont similaires en tout point à celles de la souche *NCDO 2118*, ce qui laisserait présager de la fonctionnalité de la thréonine désaminase (TD), de l'aminotransférase des BCAAs (BcaaT) et de l'isopropylmalate déshydrogénase (LeuB).

Tableau IV. 1 : Comparaison des séquences protéiques des gènes de biosynthèse des acides aminés branchés des souches *NDCO2118* et *ND5F*.

Nom du gène	Nom de la fonction protéique codée par le gène	Nomenclature EC	Nombre d'acides aminés identiques	Nombre d'acides aminés appartenant à la même famille (Position et transition <i>NCD0 2118</i> → <i>ND5F</i>)	Nombre d'acides aminés manquants	Nombre de mutation non-sens
<i>ilvA</i>	Thréonine désaminase, TD	4.3.1.19	416/416	416/416	0/416	0
<i>ilvB</i>	α -acétolactate synthase large sous-unité	2.2.1.6	569/575	573/575 N°112 : Glutamine → Leucine N°353 : Isoleucine → Thréonine	0/575	0
<i>ilvN</i>	α -acétolactate synthase petite sous-unité	2.2.1.6	155/158	156/158 N°82 : Phénylalanine → Leucine N°126 : Thréonine → Méthionine	0/158	0
<i>ilvC</i>	Acétohydroxyacide isomérase, AHAIR	1.1.1.86	317/320	318/320 N°271 : Glutamate → Alanine N°303 : Isoleucine → Thréonine	0	0
<i>ilvD</i>	Dihydroxyacide déshydratase, DHAD	4.2.1.9	529/571	532/571	21	0
<i>ilvE</i>	Aminotransférase des BCAAs	2.6.1.42	340/340	340/340	0	0
<i>leuA</i>	2-isopropylmalate synthase	2.3.3.13	184/188	184/188 N°23 : Sérine → Cystéine N°38 : Arginine → Glycine N°47 : Sérine → Proline N°126 : Arginine → Lysine	0	0
<i>leuB</i>	3-isopropylmalate déshydrogénase	1.1.1.85	293/320	296/320	24	0
<i>leuC</i>	3-isopropylmalate déshydratase large sous-unité	4.2.1.35	423/426	424/426 N°328 : Aspartate → Alanine N°394 : Asparagine → Tryptophane	0	0
<i>leuD</i>	3-isopropylmalate déshydratase petite sous-unité	4.2.1.35	191/191	191/191	0	0

Aucune mutation non-sens n'a été détectée dans tous les gènes de la voie de biosynthèse des BCAAs, seuls les gènes *ilvD* (codant la dihydroxyacide déshydratase, DHAD) et *leuB* (Isopropylmalate désydrégénase, LeuB) présentent des déplacements du cadre de lecture avec absence de, respectivement, 21 et 23 acides aminés similaires à ceux de la souche auxotrophe *IL1403*. Il est probable que ce décalage puisse altérer la structure 3D des protéines et modifier, voire annuler, l'activité de ces enzymes. Pour examiner ce point, les mutations faux-sens produisant des codons correspondant à des acides aminés d'une autre famille chimique ont été positionnées sur les structures tridimensionnelles identifiées chez *Saccharomyces cerevisiae* et *Escherichia coli* K12. Aucun changement d'acide aminé au sein des séquences de l'AHAS et de l'AHAIR n'est apparu localisé dans le site actif ou dans un site de liaison. Ces observations, ainsi que la présence de pentane-2,3-dione, corroboreraient l'hypothèse que l'AHAS soit au moins partiellement fonctionnelle chez *ND5F*. Sans preuve du contraire, nous considérons également que l'AHAIR serait également probablement partiellement fonctionnelle. Cependant, les régulations allostériques exercées par les BCAAs sur ces enzymes peuvent avoir été modifiées par ces changements de composition. En l'absence de structure 3D des protéines LeuA et LeuC, il est difficile de statuer sur les conséquences fonctionnelles des modifications d'acides aminés observées chez *ND5F* par rapport à *NDCO 2118*. De plus, ces modifications concernent des changements importants en encombrement stérique (arginine en glycine), hydrophobicité (sérine à proline) ou encore la possibilité de former des ponts disulfures (cystéine).

D'après des informations récoltées, il semblerait que les enzymes TD, AHAS, AHAIR, LeuD et BcaaT soient fonctionnelles, à la différence des protéines DHAD et LeuB qui présentent de profondes modifications de séquence et de structure, et qui pourraient ainsi expliquer l'auxotrophie observée pour les trois BCAAs. Aucun élément ne permet de se positionner concernant les protéines LeuA et LeuC. Cependant, puisque les enzymes TD et AHAS semblent fonctionnelles, la production du précurseur de la pentanedione, l'acéto-hydroxybutanoate, doit être possible.

I.1.2 Les paramètres nutritionnels

Les paramètres nutritionnels choisis pour être testés sont en relation avec les enzymes des voies de production des acides aminés branchés (BCAAs) et leur système de régulation, afin de tenter de diriger préférentiellement le flux métabolique vers le 2-acéto-2-hydroxybutanoate, qui est le précurseur par décarboxylation oxydative de la pentane-2,3-dione. A notre connaissance, il n'existe qu'une seule étude dans la littérature rapportant l'influence des BCAAs sur la formation de la pentane-2,3-dione chez des bactéries lactiques. Cependant, celle-ci a été modélisée uniquement dans le cadre de l'association des bactéries modèles du yaourt *Lactobacillus delbrueckii* ssp.

bulgaricus et *Streptococcus thermophilus* [Ott, et al. 2000] sur des souches prototrophes pour les BCAAs.

- Le premier paramètre nutritionnel cible le déplacement d'équilibre de la réaction catalysée par la thréonine déaminase (abrégé TD), qui est l'un des principaux points de contrôle de la synthèse de valine et d'isoleucine [Patek 2007]. En effet, comme présenté, la thréonine est transformée par la TD (gène *ilvA*) en 2-oxo-butanoate. La réaction catalysée étant réversible, un contrôle en faveur de la production de 2-oxo-butanoate peut être envisagé par accumulation du précurseur, la L-thréonine. Ainsi, la démarche expérimentale mise en œuvre repose sur une supplémentation du milieu MCD en thréonine (à 2X) afin de viser une augmentation du pool d'oxo-butanoate et, par extension, du flux catabolique dans les voies de biosynthèse des BCAAs pour augmenter le titre de la pentane-2,3-dione. Cependant, le 2-oxo-butanoate est un intermédiaire réactionnel intervenant également dans le métabolisme de la méthionine et de la cystéine, et du propanoate, ce qui peut suggérer qu'un flux élevé de 2-oxo-butanoate puisse être dirigé vers d'autres voies réactionnelles.

- Le second paramètre nutritionnel vise un ajustement des concentrations en BCAAs afin d'augmenter la concentration de l'acéto-2-hydroxybutanoate, et ainsi favoriser la conversion de ce précurseur en pentanedione. Pour cela, l'activité des enzymes en amont de sa voie de production, TD et AHAS, doit-être favorisée, tandis que l'activité de celles en aval nécessiterait d'être réduite. Deux niveaux de régulation ont été décrits pour ces gènes : transcriptionnel et allostérique (Figure IV. 1). Les gènes de biosynthèse spécifiques à la leucine, codés par l'opéron *leuABCD*, sont sous le contrôle du promoteur P1. Les gènes de l'opéron *ilvDBNCA* impliqués dans la synthèse d'isoleucine, de valine et dans la 1ère réaction de la voie de synthèse de la leucine sont sous le contrôle des promoteurs P1 et P2. Ces deux promoteurs sont réprimés par CodY en présence d'isoleucine. P1 est, de plus, soumis à un deuxième contrôle de l'élongation de la transcription par un mécanisme de terminaison-anti terminaison modulé en fonction de la traduction d'un peptide riche en leucine [Godon et al. 1992; Godon et al. 1993; Goupil et al. 1996]. Ces deux mécanismes ont pour but d'empêcher la production de ces enzymes anaboliques en présence de leucine et isoleucine dans le milieu. De plus, la leucine et l'isoleucine, qui ont chacune une enzyme dédiée pour amorcer leur synthèse, exercent un contrôle simple par retro inhibition respectivement sur LeuA et TD [Changeux 1961; Teng-Leary & Kohlhaw 1973]. La diminution de la concentration exocellulaire en isoleucine et leucine devrait donc permettre de lever la répression transcriptionnelle des gènes *ilvA* et *ilvBN* de l'opéron *ilv*, et diminuer le retro contrôle par l'isoleucine pour favoriser la production des enzymes TD et AHAS. Le promoteur P3 initiant la transcription de l'acétolactate décarboxylase (ALDC, gène *aldB*) est, quant à lui, réprimé, en absence de BCAAs, par la réponse stringente [Goupil-Feuillerat et al. 2000] (pour détails voir Figure IV. 1). En effet, l'ALDC catalyse la conversion de l'acétolactate en acétoïne, or l'acétolactate

est également le précurseur de la synthèse de leucine et valine par l'AHAS. La présence de leucine exerce, au contraire, un contrôle allostérique positif sur l'ALDC et la production d'acétoïne. Il a été démontré que l'AHAS est également inhibée par la valine et l'isoleucine [Umbarger 1978; Ott et al. 2000; Patek 2007].

Dans ce contexte de régulation, l'équilibre des niveaux de leucine, valine et isoleucine, est donc primordial pour contrôler les différentes voies de biosynthèse et l'activité des enzymes, dont dépend le niveau de pentane-2,3-dione. Puisque les trois BCAAs ont été identifiés comme ayant des effets délétères sur les deux enzymes TD et AHAS indispensables à la synthèse de l'acéto-2-hydroxybutanoate, il a été décidé d'étudier l'impact d'une diminution de ces acides aminés sur le titre final en pentane-2,3-dione. Pour aborder de façon dynamique le contrôle des BCAAs sur la synthèse de pentane-2,3-dione chez *L. lactis* ND5F, une limitation en BCAAs a été imposée de façon "naturelle" par consommation des acides aminés au cours de la croissance. Pour cela, la composition du milieu MCD a été modifiée afin de provoquer un épuisement précoce des BCAAs. Cependant, une concentration minimale permettant de satisfaire la croissance cellulaire a été définie (en se basant sur les cinétiques de consommation d'acides aminés en MCD). A partir des analyses préliminaires de dosage des acides aminés dans les surnageants de culture à pH libre en MCD, il a été déterminé que, pour atteindre une DO de 2,80 à pH 4,60, les quantités de BCAAs consommées étaient respectivement pour la Val/Leu/Ile de $600 \pm 50 \mu\text{M}$, $800 \pm 50 \mu\text{M}$ et $500 \pm 50 \mu\text{M}$ (données non présentées). Par conséquent, il a été déduit que les concentrations initiales en Val/Leu/Ile seraient respectivement de 600, 800 et 500 mM, soit une diminution de près de 79, 68 et 78 % par rapport aux concentrations usuelles du MCD (voir Matériels & Méthodes). Le milieu de culture est donc un MCD classique pour lequel la concentration des BCAAs se situe autour de 0,2 X. Ces concentrations sont analogues à celles définies dans le milieu synthétique simplifié MS10 précédemment développé pour les lactocoques [Cocaign-Bousquet et al. 1995].

- Le dernier paramètre évalué au cours de cette étude concerne la disponibilité préférentielle d'acides aminés sous forme d'oligopeptides à la place d'acides aminés libres dans le milieu. L'objectif de ce design expérimental est d'évaluer quel impact une substitution d'acides aminés libres par des peptides peut avoir sur la croissance, le métabolisme et éventuellement les voies de régulation des BCAAs. Ces résultats seront à mettre en perspective avec l'utilisation future de ND5F en co-cultures afin de voir si l'activité protéolytique différentielle d'autres souches influence la production de pentane-2,3-dione. En particulier, il sera analysé s'il existe, pour une même concentration finale en acides aminés, un effet entre l'ajout d'acides aminés libres et l'adjonction d'acides aminés apportés sous forme de peptides. Pour mimer cet effet, des cultures en MCD supplémentées en extrait de levure HypA® (extrait de levure sans acides aminés libres et dont la composition finale des peptides

en acides aminés est connue) ont été réalisées. Le taux de HypA a été ajusté à 0,8 % afin de correspondre à un niveau d'apport d'acides aminés sous cette forme complexe correspondant à approximativement 1 X d'apport sous forme libre par la composition classique du MCD, afin de doubler l'apport azoté global. Ces cultures seront analysées conjointement à des fermentations réalisées sur un nouveau milieu MCD défini pour être équivalent, acide aminé par acide aminé, à la concentration finale du milieu MCD + 0,8 % HypA. De plus, il a été démontré par les travaux de Juillard que les oligopeptides sont la source préférentielle d'acides aminés sur milieu complexe tel que le lait [Juillard et al. 1996; Juillard et al. 1998]. Leur présence pourrait permettre d'augmenter significativement la croissance et fournir assez d'acides aminés pour voir apparaître des phénomènes de catabolisme et de synthèse d'arômes associés [Ayad et al. 1999; Bruinenberg et al. 1992; Novák & Loubiere 2000].

I.1.3 Les paramètres physico-chimiques du procédé

Des cultures sous régulation du niveau d'oxygène ont été programmées afin de favoriser la conversion par décarboxylation oxydative du 2-acéto-2-hydroxybutanoate en pentane-2,3-dione. L'utilisation de ce levier a déjà été longuement décrit dans la littérature pour favoriser de manière similaire la production d'acétoïne et de diacétyle [Cogan et al. 1989; Bassit et al. 1993; Escamilla-Hurtado et al. 1996] par l'utilisation de concentration en oxygène allant de la micro-aérophilie (0-5 %), jusqu'à l'aérobiose à 100 %. Il a ainsi été démontré que la concentration en diacétyle pouvait être multipliée par un facteur 18 sous régulation à 100 % de la concentration en oxygène [Bassit et al. 1993]. Ces effets observés sont liés d'une part au contrôle du métabolisme par la balance énergétique qui contraint à un équilibre entre les réactions productrices et celles consommatrices de NADH, et d'autre part au phénomène de décarboxylation oxydative qui est favorisé. De plus, il a été montré que l'activité de l'acétolactate synthase (ALS) qui convertit le pyruvate en acétolactate était activée par la présence d'oxygène [Condon 1987] et pouvait également participer à l'augmentation observée. Il est possible qu'un phénomène similaire d'activation de l'AHAS existe, bien que non démontré à l'heure actuelle. Une régulation sous 30 % de concentration en oxygène a été choisie ici (100 % étant la concentration saturante en oxygène dans l'eau).

Cependant, il faut noter que les études ont montré que l'aérotolérance de *L. lactis* est souche-dépendante et en relation avec la complexité du milieu de culture [Cocaign-Bousquet et al. 1995]. Sur milieu riche comme le lait, l'impact de l'aération sur la croissance est faible (Chapitre I), à la différence d'un milieu minimum où l'effet est plus marqué [Cocaign-Bousquet et al. 1995].

I.2 Analyse comparative du comportement macro-cinétique

Les différentes cultures ont été réalisées en triplicat biologique en fermenteur avec un suivi en temps réel de la biomasse, du pH, de l'O₂, du REDOX, et des principaux métabolites carbonés et des acides aminés. Les essais ont été conduits à pH libre en environnement micro-aérophile (sans aération mais au contact de l'atmosphère non dégazée en début de fermentation) sauf pour celui à PO₂ régulé. Le milieu MCD a été préparé à 60 g.L⁻¹ de lactose par analogie au lait. La composition détaillée est présentée section Matériels & Méthodes. Les fermentations sont arrêtées lorsque le pH du milieu de culture atteint 4,60 de manière similaire aux expérimentations en lait. Le niveau final de pentane-2,3-dione a également été mesuré à partir des surnageants en fin de culture à 24 h en spectrométrie de masse. Le détail des conditions testées est résumé dans le Tableau IV. 2.

Tableau IV. 2 : Détail des conditions expérimentales (1 à 6) des essais sur MCD de production différentielle de pentane-2,3-dione. (- : Pas de modification)

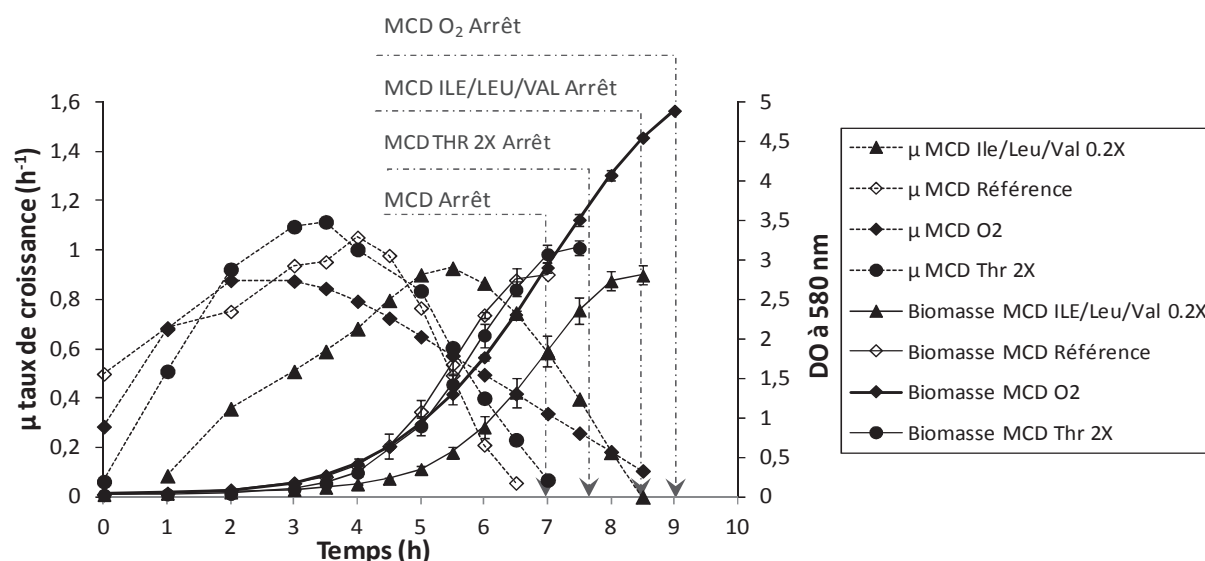
Condition	Référence	Paramètres	
		Nutritionnel	Physico-chimique
Condition 1	La référence : MCD	-	-
Condition 2	MCD-condition1	THR (2X)	-
Condition 3	MCD-condition1	ILE/VAL/LEU (0,2X)	-
Condition 4	MCD-condition1	-	30% PO ₂ (100% saturation)
Condition 5	MCD-condition1	AA (2X)	-
Condition 6	MCD-condition 5	MCD –condition 1+0,8% HypA	-

I.2.1 Cinétiques de croissance et d'acidification

Pour commenter ces résultats, les essais seront analysés face à leurs références respectives afin d'appréhender précisément l'impact du facteur étudié sur la croissance. Les profils de croissance des conditions 1 à 4 sont présentés **Erreur ! Source du renvoi introuvable.** Le critère d'arrêt des cultures étant fonction de la cinétique d'acidification propre à chaque condition, les fermentations ont été stoppées à des temps (entre 7 et 10 h), à des concentrations cellulaires (entre 2,8 et 5 de DO₅₈₀) et à des états physiologiques différents (phase de croissance, de ralentissement ou encore phase stationnaire).

A noter également que l'effet inhibiteur d'un pH faible (en-dessous de 4,8) est vraisemblablement responsable des ralentissements voire des arrêts de croissance constatés. En effet, après vérification, ni le lactose, ni les acides aminés ne sont totalement épuisés à la fin des cultures, excluant un arrêt par carence nutritionnelle. De plus, ce milieu a déjà permis d'atteindre sous régulation de pH à 6,4 une DO de 10 pour ND5F (résultats non présentés), confirmant que les autres composés non dosés tels que les vitamines ou cofacteurs sont en quantité suffisante pour supporter une telle biomasse.

Ainsi, au vu de son importance, le comportement d'acidification sera donc discuté parallèlement afin



d'explicitier au mieux les profils de biomasse obtenus (Figure IV. 3).

Figure IV. 2 : Evolution de la biomasse et du taux de croissance de *L. lactis ssp. lactis ND5F* au cours de la culture discontinue à pH libre dans différents MCD. Les données courbes sont obtenues par lissage. Les écart-types sont issus de trois réplicats biologiques.

Pour l'intégralité des fermentations, la croissance démarre dès l'inoculation, sans phase de latence. La fermentation de *L. lactis ND5F* sur MCD à pH libre permet d'atteindre à pH 4,60 une DO de 2,8 en 7 h de culture, le taux de croissance maximal en phase exponentielle étant de $1,05 \text{ h}^{-1}$ à 4 h. Ce taux maximal de croissance sur milieu synthétique est similaire à celui obtenu pour des souches laitières telles que *L. lactis IL1403* avec $1,00 \text{ h}^{-1}$ en présence d'un environnement micro-aérophile [Cocaign-Bousquet et al. 1995].

L'apport de thréonine permet d'augmenter légèrement la biomasse formée par rapport au MCD classique de 12 %, ainsi que le taux de croissance maximal du même facteur (**Erreur ! Source du renvoi introuvable.**). Le temps de fermentation est également supérieur de 30 minutes. Au vu de l'effet constaté sur la croissance, l'apport supplémentaire de thréonine permet vraisemblablement de limiter l'utilisation d'aspartate et de méthionine pour fournir la cellule en thréonine. Par conséquent, la diminution des besoins de synthèse en thréonine doit permettre de limiter l'utilisation de l'oxaloacétate vers ces voies de biosynthèse et de redistribuer le flux vers d'autres nœuds métaboliques nécessaires à la croissance.

La diminution des BCAAs à 0,2 X a provoqué une diminution du taux de croissance maximal de 10 % par rapport à la fermentation type en MCD, bien que le niveau final de biomasse à l'arrêt de culture, une demi-heure plus tard, soit identique (**Erreur ! Source du renvoi introuvable.**). Il est observé

également un décalage temporel dans la cinétique d'acidification lié très probablement à une production d'acide lactique décalée dans le temps (Figure IV. 3).

Ce niveau final de biomasse confirme donc que les concentrations choisies pour ce milieu de culture étaient adéquates pour satisfaire les besoins nutritionnels. La diminution du taux maximal de croissance révèle cependant une baisse de l'efficacité énergétique due à cette diminution.

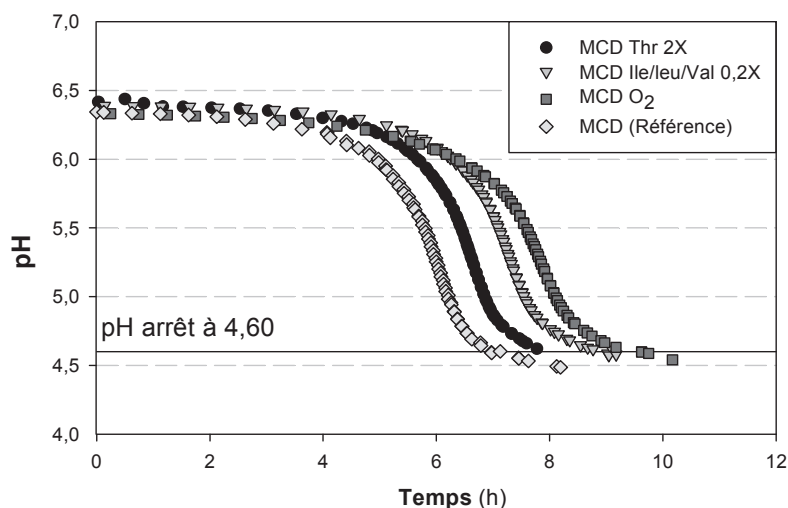
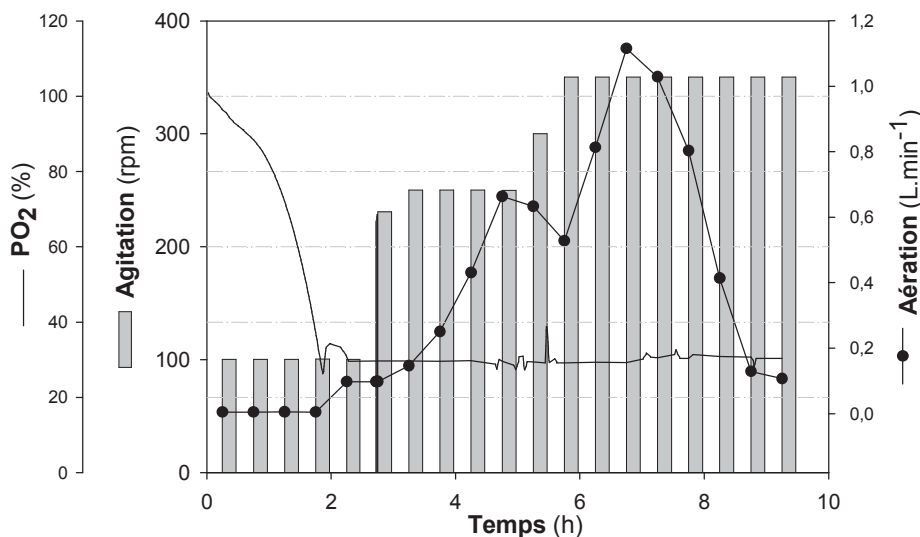


Figure IV. 3 : Evolution du pH au cours de la culture discontinue de la souche *L. lactis ssp. lactis ND5F* à pH libre dans différents MCD.

La mise en place d'une régulation de la concentration en oxygène à 30 % pendant la phase de croissance a été réalisée en contrôlant simultanément le couple aération-agitation pour satisfaire la consigne de la pression partielle en oxygène. Il a été fixé comme limites maximales 1 L.min⁻¹ pour l'aération et 500 rpm pour l'agitation. Comme présenté Figure IV. 4, la mise en place de la régulation de la PO₂ a débuté lorsque la PO₂ initialement à 100 % a diminué à 30 %, à 2,25 h. L'analyse de la croissance sous condition aérobie révèle qu'une DO maximale de 5,9 est atteinte après 9 h de culture, soit une augmentation d'un facteur 2 par rapport à la référence (Figure IV. 3).



[THESE CONFIDENTIELLE]

Figure IV. 4 : Descriptif du suivi des consignes des valeurs du couple agitation-aération lors de la régulation de la PO₂ à 30 % en MCD.

Cependant, le taux maximal de croissance est plus faible de 15 % à 0,85 h⁻¹, par rapport à la référence MCD en environnement micro-aérophile. Ces résultats sont en accord avec les taux de croissance mesurés sur la souche *L. lactis* ssp *cremoris* MG 1363 sous environnements micro-aérophile et aérobic sur milieu complexe, où une décroissance de 12 % était également observée [Nordkvist et al. 2003]. De plus, à la différence du profil en MCD en micro-aérophilie, en aérobiose, la phase de ralentissement de la croissance se poursuit sur une période plus importante de 5 h. Ce phénomène est à l'origine de la biomasse finale plus importante. Cette poursuite de croissance sur une période plus longue est une conséquence indirecte de l'oxygénation du milieu. En effet, le retard dans l'acidification qu'elle provoque (7 h pour perdre 0,5 unité de pH au lieu de 5 h sur MCD classique) permet une croissance plus longue et une biomasse plus élevée.

Les autres conditions nutritionnelles analysées concernent la modification globale de l'apport en source azotée. Les résultats présentés Figure IV. 5 révèlent que toute augmentation de la quantité d'acides aminés (libres ou sous forme complexe) se traduit par une biomasse finale supérieure de respectivement 18 et 100 % par rapport à la composition MCD classique. Cependant, lors de la multiplication par deux des quantités d'acides aminés libres du MCD, le taux maximal de croissance n'est que de 0,74 h⁻¹. Ce résultat indique que la concentration (doublée par rapport à la référence) de l'un au moins des acides aminés a atteint une valeur légèrement inhibitrice pour la croissance.

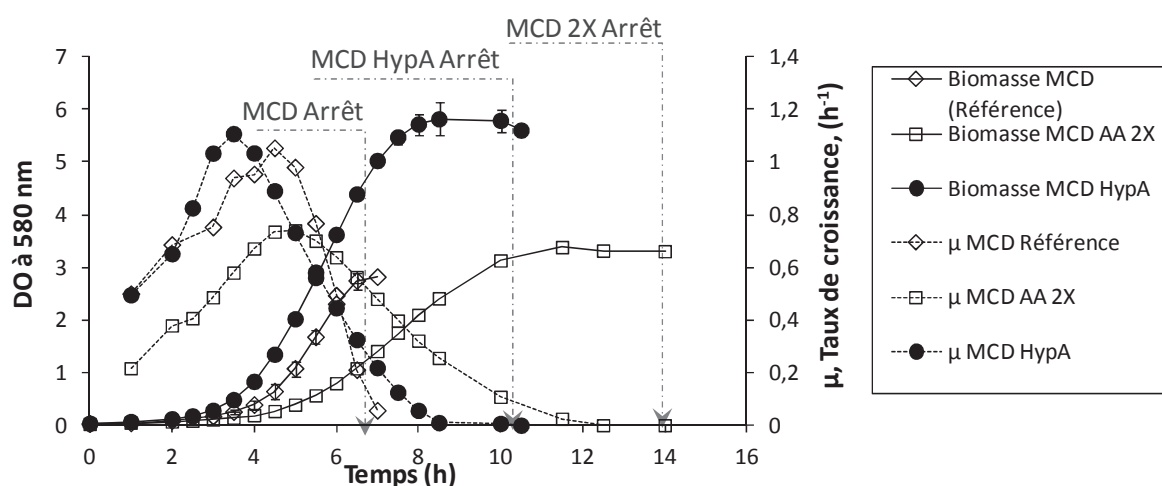


Figure IV. 5 : Evolution de la biomasse et du taux de croissance de *L. lactis* ssp. *lactis* ND5F au cours de la culture discontinue à pH libre dans différents MCD. Les données courbes sont obtenues par lissage. Les écarts-types sont issus de trois répliquats biologiques, sauf pour MCD AA 2X en duplicat.

A noter également qu'il a été impossible d'atteindre un pH de 4,60 après 14 h de culture en présence d'une concentration doublée de chaque acide aminé (Figure IV. 6).

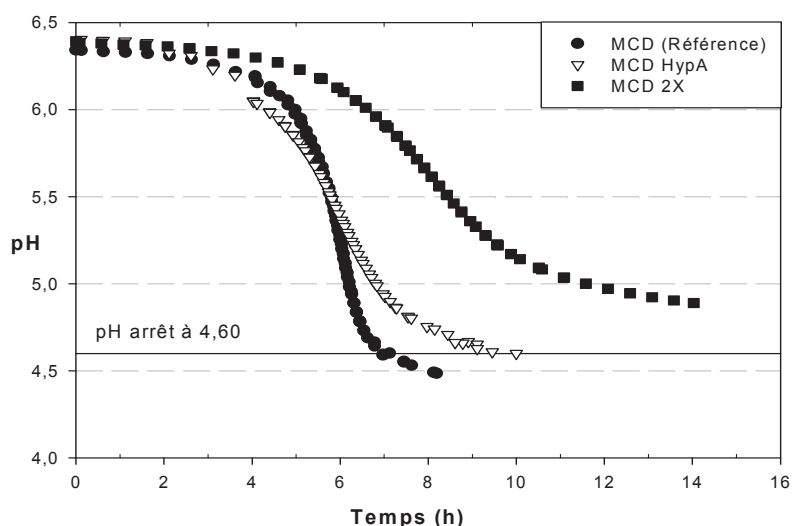


Figure IV. 6 : Evolution du pH au cours de la culture discontinue à pH libre de la souche *L. lactis* ssp. *lactis* ND5F dans différents MCD.

Trois hypothèses sont possibles pour expliquer ce phénomène, i) un pouvoir tampon augmenté par le doublement des acides aminés, ii) une moindre production d'acide lactique (hypothèse réfutée sur la base des dosages), et iii) une production de molécules basiques, du type NH_3 , par désamination de différents acides aminés qui contrecarre la baisse du pH. Il a donc été décidé de stopper la fermentation à environ pH 4,8 pour éviter des phénomènes de lyse qui auraient pu affecter les résultats de l'analyse protéomique sur ces surnageants de fin de culture.

L'apport des acides aminés sous forme peptidique permet, quant à lui, d'augmenter légèrement le taux de croissance maximal (de 7 %), mais surtout de soutenir la croissance plus longtemps. Comme sur lait, ces résultats confirment que chez *L. lactis*, l'apport de source azotée sous forme peptidique est préférentiel pour augmenter la biomasse (meilleur rendement énergétique). Toutefois, il est possible que se combine également un effet tampon par les peptides, qui ralentirait la décroissance du pH pour une même concentration en acides produits, et atténuerait les effets inhibiteurs du pH (résultats à confirmer avec l'étude métabolique par comparaison des pHs lus à une même concentration de lactate). Ces résultats sont cohérents avec l'apparente diminution de la vitesse maximale d'acidification d'un facteur deux. Autre remarque, cet extrait de levure doit vraisemblablement contenir des vitamines et autres facteurs de croissance (pas d'indication sur la fiche fournisseur). Or il est connu que ces composés peuvent être à l'origine de cette prolongation de la croissance. Le travail avec ce substrat complexe ne nous permet pas d'exclure cette hypothèse.

I.2.2 Profils de consommation d'oxygène et REDOX

Les profils REDOX sont similaires pour toutes les fermentations, quelles que soient les conditions nutritionnelles. D'un potentiel Eh_7 positif à l'inoculation (15 ± 20 mV), le milieu de culture passe à potentiel Eh_7 négatif (-320 ± 20 mV), preuve de l'activité réductrice de *L. lactis ND5F*. Les profils de consommation d'oxygène sont similaires et révèlent une disparition de l'oxygène entre 2 et 2,5 h de fermentation, tout comme observé sur lait. Logiquement, les fermentations sous régulation de la PO_2 présentent un Eh_7 positif tout le long de la culture et l'effet oxydant de l'apport constant d'oxygène masque toute réaction de réduction pouvant avoir lieu.

I.2.3 Métabolismes carboné et azoté

Dans un premier temps, les profils de consommation et de production de métabolites sont exposés, et les rendements globaux de production calculés pour comparer le comportement de la souche en fonction du milieu. Le cas particulier de la production aromatique de la pentane-2,3-dione sera ensuite traité indépendamment. Les profils métaboliques des différentes conditions MCD (Référence), MCD Thr 2X, MCD Ile/Leu/Val 0,2X et MCD O_2 sont présentés Figure IV. 7. A l'arrêt de culture, il reste encore plus de 150 mM de lactose dans le milieu sur les 180 mM présents initialement pour toutes les conditions. Pour les fermentations sur MCD, MCD Thr 2X et MCD Ile/Leu/Val 0,2X, les profils de consommation de lactose et de production de produits fermentaires sont similaires. Le lactate est le produit majoritaire (65 ± 7 mM inter conditions) et les autres produits, formiate, acétate et éthanol, sont présents dans les proportions 2 :1 :1 respectivement (en concentration inférieure à 1,5 mM). En présence d'oxygène (Figure IV. 7 D), comme attendu, il est observé une diminution importante de la production de lactate (42 ± 6 mM) et une production massive d'acétoïne qui compte pour moitié du lactate produit.

Les cinétiques de consommation du lactose et de formation des produits des essais MCD-HypA et MCD AA 2X sont similaires à celle sur MCD, et caractéristiques d'un métabolisme homo fermentaire (données non représentées). Afin de mieux déterminer l'effet de ces différentes conditions sur le métabolisme de *L. lactis ND5F*, les bilans stœchiométriques ainsi que le bilan carboné vont être calculés et comparés. Les rendements en biomasse seront également présentés. Cette analyse fine du métabolisme permettra d'explicitier au mieux les différents niveaux de biomasse obtenus, de vérifier en fonction des bilans si certains composés ont été omis, et de faire le lien entre métabolisme azoté et carboné.

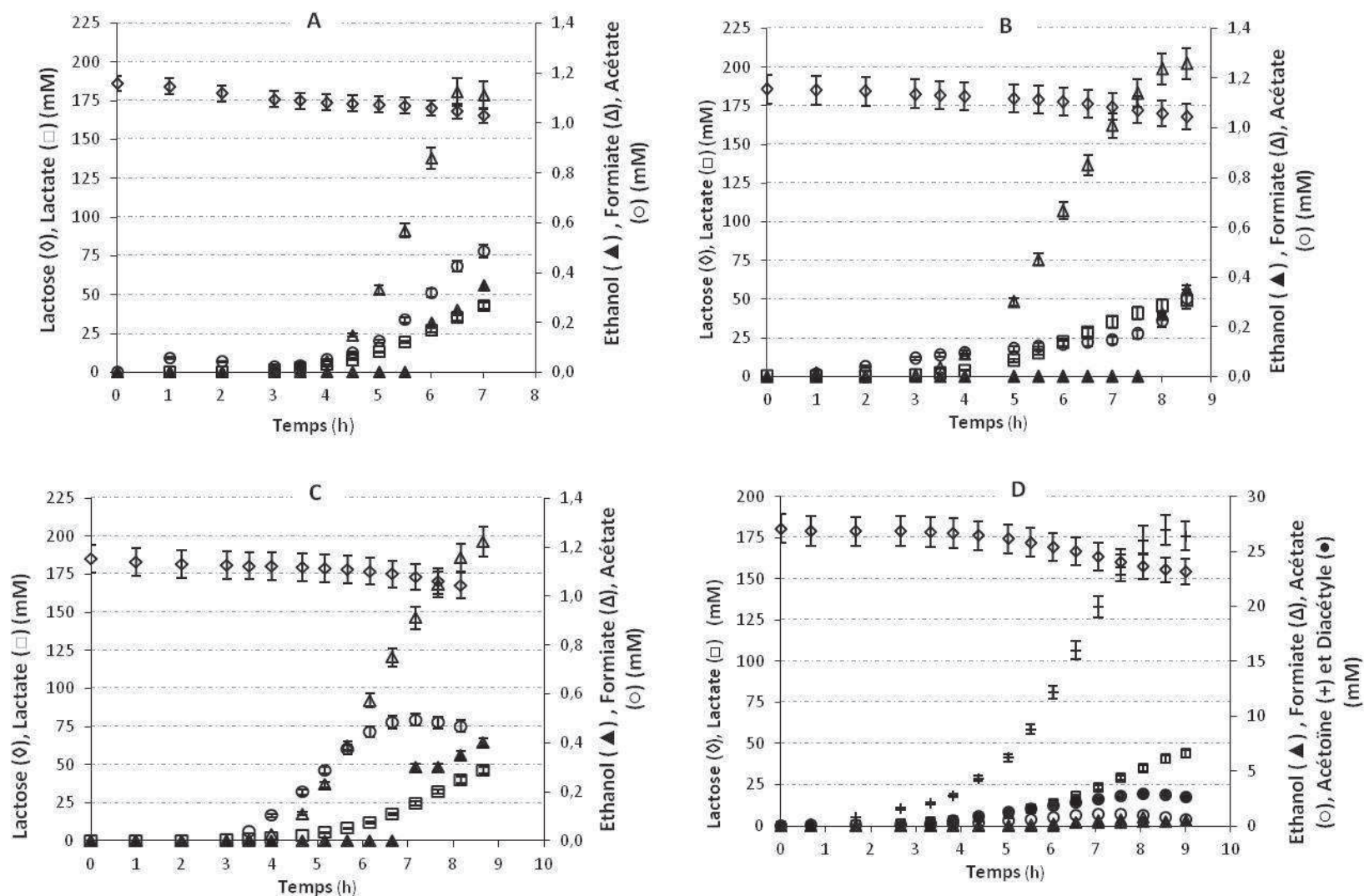


Figure IV. 7 : Cinétique de consommation du substrat (lactose à 60 g.L⁻¹) et de formation des produits (lactate, formiate, acétate, éthanol). Cultures en batch de *L. lactis* ssp. *lactis* ND5F à pH libre sur MCD, la référence (A), sur MCD THR 2X(B), sur MCD ILE/LEU/VAL 0,2X (C) et MCD O₂ (D). Résultats moyens de trois réplicats. Les productions d'acétoïne et diacétyle ne sont mentionnées que si leur niveau est supérieur à 2 mM.

[THESE CONFIDENTIELLE]

La quantité maximale de biomasse par rapport au lactose consommé est atteinte sur le milieu MCD supplémenté en Thr-2X, mais avec une valeur non significativement différente de celles correspondant aux milieux MCD (avec et sans aération) et MCD avec HypA, plus faibles (Tableau IV. 3). Ce résultat met l'accent sur le rôle de la thréonine dans l'efficacité biosynthétique, lié vraisemblablement à sa position d'intermédiaire réactionnel primaire dans la voie de biosynthèse des BCAAs. En effet, le rendement obtenu en biomasse dans le cas du milieu avec une concentration en BCAAs diminuée à 0,2X est faible, révélant l'importance de ces acides aminés pour lesquels la souche est auxotrophe. Logiquement, la présence de peptides a renforcé l'apport azoté et favorisé le rendement en biomasse produite.

Tableau IV. 3 : Bilans stœchiométriques carbonés et azotés obtenus dans les différentes conditions de culture en MCD de *L. lactis* ND5F. Les valeurs exprimées en millimoles sont ramenées à une concentration de 50 mmol de lactose (=100 mmol de glucose). ND : Non déterminé et - : Absence. Valeurs calculées à partir ^a des profils de consommation des acides aminés, ^b de la biomasse formée, du rapport 1 uDO = 0,33 g.L⁻¹, ^c à partir des produits formés et ^[1] de la masse molaire de la biomasse à 129 g.mol⁻¹ [Novak et al. 1997].

	MCD	MCD- THR 2X	MCD Ile/Leu/Val 0,2X	MCD O ₂	MCD AA 2X	MCD HypA	MCD avec <i>L.lactis</i> <i>NDCO 2118</i> ^[1]
Substrat							
Lactose (glucose)	50	50	50	50	50	50	100
Acides Aminés ^a	16,6± 0,54	18,6± 0,82	16,6± 0,36	16,6± 0,33	16,4± 0,44	ND	22,6± 0,91
Produits							
Biomasse ^b	20,0± 0,9	23,5± 0,7	16,4± 0,6	23,2± 0,7	17,3± 0,5	22,3± 0,7	25,6
Lactate	119,2± 4,5	139,8± 5,0	111,7± 4,1	81,1± 6,5	111,7± 4,2	123,0± 8,1	170,0
Formiate	3,1± 0,2	3,6± 0,3	2,94± 0,5	1,4± 0,4	2,9± 0,4	6,0± 0,3	12,0
Acétate	1,4± 0,3	0,9± 0,2	1,1± 0,3	1,1± 0,2	1,1± 0,4	3,2± 0,5	5,0
Ethanol	2,4± 0,5	1,0± 0,2	1,0± 0,2	0,8± 0,2	0,9± 0,2	3,1± 0,3	4,0
Acétoïne	4,4± 0,4	3,0± 0,3	2,0± 0,4	48,8± 4,5	5,1± 0,5	3,2± 0,4	-
Diacétyl	2,1± 0,4	2,1± 0,4	1,1± 0,3	4,9± 0,7	1,1± 0,6	0,7± 0,3	-
CO ₂ ^c	15,8	11,1	9,06± 4,5	108,4	14,5	14,2	-

Comme déjà évoqué lors de la présentation des profils de biomasse, la faible valeur de biomasse obtenue sur MCD AA 2X résulte d'une inhibition partielle du taux de croissance du fait du doublement de la concentration de tous les acides aminés. Ces résultats sont cohérents avec les données bibliographiques obtenues sur glucose pour la souche *NDCO 2118*.

L'analyse de la consommation en acides aminés n'a pas été réalisée sur le milieu complexe MCD-HypA du fait des autres sources d'azote (peptides) non quantifiables. La consommation globale en

acides aminés est équivalente sur les milieux MCD (avec ou sans oxygénation), et MCD Ile/leu/val 0,2X ; elle est légèrement supérieure sur MCD Thr 2X (Tableau IV. 3). L'observation du profil de consommation de chaque acide aminé révèle que la quantité de thréonine consommée est plus importante dans l'essai où elle est supplémentée, mais également dans le cas de l'essai avec BCAAs plus faible, illustrant le besoin de compenser ces acides aminés (données non présentées). Cette consommation est inférieure à celle mise en évidence sur le même milieu pour *NDCO 2118* bien que cette souche d'origine végétale soit considérée comme moins exigeante en acide aminé par rapport à une souche d'origine laitière. Cette étude confirme donc que les exigences nutritionnelles en acides aminés sont souches dépendantes et demandent une étude approfondie sur chaque souche utilisée pour comprendre au mieux ses besoins.

Indépendamment du milieu MCD utilisé, le produit majoritaire du métabolisme de *L. lactis ND5F* reste l'acide lactique. Seule la présence d'oxygène permet de passer d'un flux quasi unidirectionnel du pyruvate vers le lactate vers la production intense de diacétyle et surtout d'acétoïne. En effet, en aérobiose, l'activité de la NADH oxydase rentre en compétition avec la LDH comme voie alternative de régénération de NADH en utilisant l'oxygène comme accepteur d'hydrogène. A noter que le rapport diacétyle sur acétoïne est de 1 : 10, révélant d'une part le rôle prépondérant de l'environnement oxydant sur la décarboxylation oxydative chimique de l'acétolactate, et d'autre part l'activité de l'acétolactate décarboxylase (ALDC). Il est également observé une réduction de la production de ces deux molécules aromatiques lors de la diminution des BCAAs du milieu MCD. Ce résultat confirme la régulation négative de l'ALDC chez *L. lactis ND5F* en absence de leucine, afin de favoriser l'utilisation « anabolique » de l'acétolactate.

Pour préciser en détail les états métaboliques (homolactique, mixte), et vérifier si des composés produits n'ont pas été omis lors de l'analyse, les bilans élémentaires en carbone ont été calculés (Tableau IV. 4). Le bilan carboné global a été établi en tenant compte de la consommation de la source carbonée et des acides aminés, ainsi que de la formation des produits et de la biomasse. Les valeurs diffèrent en fonction des conditions (de 99,5 % en présence d'oxygène sur MCD à seulement 66,1 % sur MCD avec BCAAs à 0,2X). Ainsi, à l'exception de la condition sous régulation d'oxygène, les autres bilans révèlent un déficit carboné. Ce déficit du bilan carbone est très probablement lié à la faible consommation en lactose mesurée à pH libre (entre 12 et 15 % seulement du lactose initial, valeur comparable au milieu lait), et à l'imprécision sur cette mesure qui a pour conséquence de surévaluer la quantité de lactose consommée lorsque celle-ci est ramenée pour comparaison à un taux de 50 mM. Afin d'exclure un effet significatif de la production éventuelle d'exopolysaccharides,

les surnageants de fin de fermentation ont été traités au TCA, mais aucune production significative n'a été mesurée.

A la différence de *NDCO 2118* (certes en anaérobiose stricte et à pH régulé), le rendement de lactate sur la source de carbone est apparemment égal ou inférieur à 70 %, ce qui pourrait indiquer que le métabolisme de *L. lactis ND5F* n'est pas homolactique (au-dessus de 80 %) sur ces substrats. Or la présence d'autres produits issus du métabolisme glycolytique, tels que le formiate, l'acétate et l'éthanol, reste mineure (entre 1 et 3 % en rendement global sur le lactose), démontrant que le métabolisme est bien de type homolactique. Ces observations corroborent les imprécisions de mesure du lactose. En aérobiose, le rendement $R_{\text{lactate/lactose}}$ est logiquement de 40 % pour un bilan catabolique au-dessus de 95 %, illustrant le poids de la production de diacétyle et acétoïne dans cette condition.

Tableau IV. 4: Bilans carbone (global et catabolique) et rendements de production de lactate obtenus durant la culture de *L. lactis ND5F* dans les différentes conditions.^[1] Données de [Novak et al. 1997]. ND : Non déterminé. ^a : Bilan carbone global calculé, ^b : Bilan carbone catabolique : quotient des produits formés sur la source de C consommée.

$R_{\text{lactate/source C}}$ Rendement lactate sur	MCD	MCD THR, 2X	MCD Ile/Leu/Val	MCD O ₂	MCD AA, 2X	MCD HypA	MCD avec <i>L.lactis NDCO</i> 2018 ^[1]
lactose/glucose (C _{mol} /C _{mol})	74,2	83,0	66,1	99,5	77,3	ND	87,7
Bilan C catabolique (%) ^b	68,0	76,4	60,9	95,3	73	69,52	87,0
$R_{\text{lactate/source C}}$ Rendement lactate sur lactose/glucose (C _{mol} /C _{mol})	0,60	0,70	0,56	0,41	0,65	0,61	0,85

I.3 Impact sur la production de la pentane-2,3-dione

Cette étude ayant pour objectif la détermination des facteurs influençant la production de pentane-2,3-dione, seule cette dernière a été dosée parmi les composés volatils. Chaque surnageant de fin de fermentation a été immédiatement congelé (-20°C) puis quantifié en triplicat par la méthode SPME-GC/MS-MS mise au point. Les résultats des différents réplicats techniques ont été moyennés pour une même culture. Ne sont présentés

Tableau IV. 5 que les concentrations et écarts-type des triplicats biologiques pour les différentes conditions. L'analyse statistique démontre la présence de différences significatives entre les concentrations mesurées en pentane-2,3-dione au sein des différentes conditions (Analyse ANOVA avec $p < 0,001$). Cependant, seul l'essai en présence de thréonine permet d'obtenir une

	MCD	MCD- THR 2X	MCD Ile/Leu/Val 0,2X	MCD O ₂	MCD AA 2X	MCD HypA
Concentration moyenne ($\mu\text{g.L}^{-1}$)	32,0	133,9	17,1	40,9	47,2	31,3
Ecart-type	9,2	18,1	2,3	9,8	2,0	13,8

concentration significativement plus élevée d'un ordre 4 par rapport à la composition en MCD classique.

Tableau IV. 5 : Concentration en pentane-2,3-dione (en $\mu\text{g.L}^{-1}$) au sein des surnageants en fin de culture de *L. lactis* ND5F en fonction des conditions testées.

Cette observation confirme que le doublement de la concentration de thréonine a donc probablement apporté un supplément de 2-oxobutanoate et, grâce à la présence de pyruvate, participé à la synthèse du précurseur de la pentane-2,3-dione, le 2 acéto-hydroxybutanoate. Ce résultat pourrait permettre d'envisager une stratégie de sélection de souches pour une utilisation en co-culture avec ND5F basée sur un criblage de souches caractérisées par une libération plus importante de thréonine libre dans le milieu. Ces résultats sont confirmés par la tendance observée sur l'essai MCD-AA 2X qui comprend, en plus des autres acides aminés, un taux de thréonine multiplié par deux. La production mesurée de pentane-2,3-dione est supérieure au milieu MCD de référence, mais l'impact est fortement diminué par rapport au milieu MCD -Thr 2X.

La présence d'un environnement oxydant n'a pas permis d'augmenter significativement la production de pentane-2,3-dione, bien que la valeur moyenne de pentane-2,3-dione détectée soit supérieure de 28 % par rapport à la référence. Il est possible que la faible quantité de 2 acéto hydroxybutanoate disponible (à la différence de l'acétolactate) masque l'effet positif de la présence d'oxygène sur la décarboxylation oxydative, ou que l'aspect évolution temporelle soit nécessaire pour augmenter le rendement comme observé dans l'étude d'Ott [Ott et al. 2000]. Il serait donc intéressant de combiner les deux approches (ajout de thréonine et régulation oxygène) pour vérifier si l'effet nutritionnel est ainsi décuplé.

L'utilisation de HypA n'a pas d'effet significatif sur la concentration en pentane-2,3-dione. Ce résultat pourrait s'expliquer par l'augmentation du taux de leucine et d'isoleucine dans le milieu qui aurait

réprimé la transcription de l'opéron *ilvDBNCA*. Par ailleurs, il est envisageable que sur d'autres COVs issus du catabolisme de acides aminés, des variations auraient pu être observées en raison de la disponibilité accrue de ces sources d'azote.

L'une des observations les plus surprenantes est que la diminution des concentrations en BCAAs paraît avoir un effet négatif (diminution significative d'un facteur 2) sur la production de pentane-2,3-dione. Plusieurs hypothèses peuvent être avancées pour expliquer ce résultat non attendu. Premièrement, il est possible que la carence imposée ait réduit l'opportunité de décarboxylation oxydative du 2 acéto-hydroxybutanoate, ce dernier ayant été dirigé vers les enzymes terminales de la voie de biosynthèse pour satisfaire les besoins en isoleucine. De plus, la diminution de la concentration en BCAAs initialement ciblée pour augmenter l'activité des enzymes TD et AHAS a pu également être contrebalancée par l'augmentation de l'activité de l'enzyme AHAIIR en aval de l'AHAS, qui suivent les mêmes principes de régulation. Cette dernière est théoriquement active d'après notre étude de fonctionnalité du gène *ilvC* par bioinformatique.

Les mutations observées dans les gènes correspondant ont pu également modifier les régulations allostériques de ces enzymes de telle sorte que la diminution en BCAAs ait augmenté l'activité de l'AHAIIR au détriment de l'AHAS, ce qui limiterait l'effet levier voulu par la diminution des concentrations. Il est également probable qu'une limitation de la disponibilité en pyruvate, réparti entre son utilisation dans le flux catabolique de la lactate déshydrogénase pour régénérer le NADH (métabolisme homolactique dans cette condition), mais également comme substrat pour l'ALS, explique une activité plus faible de l'AHAS. De plus, suite à la diminution en BCAAs, une compétition a pu avoir lieu au niveau de l'AHAS entre la voie de biosynthèse de la valine et de la leucine, et celle de l'isoleucine que nous souhaitons favoriser. Cette dernière réclame, en plus du pyruvate, une disponibilité en oxo-butanoate qui pourrait potentiellement, là encore, être limitant dans ces conditions. Il est à noter cependant que l'AHAS présente une plus grande affinité pour le pyruvate que l'ALS [Benson et al. 1996].

Toutefois, aucune information n'est disponible quant au rapport de concentration de ces deux protéines, bien que les enzymes cataboliques chez *L. lactis* soient vraisemblablement largement exprimées par rapport à celles anaboliques en MCD comme observés sur lait. Dans ce contexte, l'analyse protéomique est donc tout à fait adéquate pour visualiser si une expression des protéines de la voie de biosynthèse de l'isoleucine a été caractérisée, et comment cette dernière a évolué en fonction des conditions.

I.4 Réponses protéomique comparatives

L'analyse des protéines a été conduite sur les culots cellulaires à partir des six conditions précédemment étudiées. Les prélèvements ont été effectués à pH 4,60.

I.4.1 Présentation des protéines cibles

Une liste de 19 protéines a été définie en se rapportant aux principales enzymes du métabolisme carboné en aval du nœud du pyruvate (LDH, PFL, PDH), ainsi qu'aux voies de biosynthèse des acides aminés branchés et du catabolisme de la thréonine dont est issue la pentane-2,3-dione (Tableau IV. 6). La présence des gènes de ces enzymes a été vérifiée au préalable sur le génome de *ND5F*. Il est à noter que quatre protéines incluses dans cette analyse (AHAS, LeuB, LeuC et l'histidinol phosphate transaminase) n'ont pas été dosées lors de l'analyse protéomique en lait. Cependant, la présence ou l'absence des enzymes lors de l'analyse protéique en lait ne conditionne en rien l'expression de ces enzymes dans ces nouvelles conditions en MCD. Il a été choisi d'étudier également l'expression de quatre aminotransférases du fait de leur implication dans la réorientation des flux métaboliques des acides aminés. En effet, elles participent à la fois à l'étape terminale de la biosynthèse des acides aminés et à l'étape initiale de leur catabolisme en de nombreux COVs. L'analyse des composés volatils sur lait avait également souligné le rôle de l'étape de lipolyse dans la production aromatique des souches laitières, et notamment sa très faible activité, caractéristique de la signature de *ND5F*. Pour étudier plus en détail ces enzymes, deux protéines intervenant dans la production de corps cétoniques ont donc été sélectionnées, l'hydroxyméthylglutaryl-CoA synthase et l'acétyl Coenzyme A acétyltransférase.

Dans le cas présent, la problématique de l'étude se focalise sur une analyse comparative de la quantité d'un nombre défini de protéines, ce qui justifie le haut niveau d'identification protéique avec une FDR de 1%. L'analyse des résultats d'identification de cette liste de 19 protéines sélectionnées révèle que 8 d'entre elles n'ont jamais été identifiées, quelle que soit la condition de culture :

- les deux aminotransférases impliquées dans le métabolisme de l'histidine et des acides aminés aromatiques ;
- trois des quatre enzymes spécifiques de la voie de biosynthèse de la leucine codées au sein de l'opéron *leu*, qui permettent de la synthétiser à partir de céto-isovalérate: LeuC, LeuD, LeuB ;
- trois des cinq protéines codées par l'opéron *ilv*, impliquées dans la biosynthèse des BCAAs : TD (*ilvA*), AHAS (*ilvBN*) et DHAD (*ilvD*) ;

- l'acetyl coenzyme A acétyltransférase (*fadA*), enzyme au cœur du système de synthèse/dégradation des acides gras.

Tableau IV. 6 : Liste des dix-neuf protéines sélectionnées pour l'analyse protéomique comparative avec leur positionnement métabolique et leur gène respectif.

Voie métabolique	Nom de la protéine	Nomenclature EC	Abréviation usitée de la protéine	Non du gène
Directement en aval du pyruvate	Lactate déshydrogénase	1.1.1.27	LDH	<i>ldh</i>
	Pyruvate formiate lyase	2.3.1.54	PFL	<i>pfl</i>
	Pyruvate déshydrogénase E1 component	1.2.4.1	PDH	<i>pdhA</i>
En aval du pyruvate, et précurseur de acétolactate et/ou de 2-acéto-2-hydroxybutanoate	Acétolactate synthase	2.2.1.6	ALS	<i>als</i>
	Acétohydroxyacide synthase	2.2.1.6	AHAS	<i>ilvBN</i>
Voie de biosynthèse de la leucine à partir de céto-isovalérate	Isopropylmalate déshydratase grande sous-unité	4.2.1.35	LeuC	<i>leuC</i>
	Isopropylmalate déshydratase petite sous-unité	4.2.1.35	LeuD	<i>leuD</i>
	Isopropylmalate déshydrogénase	1.1.1.85	LeuB	<i>leuB</i>
En aval de thréonine, et précurseur de 2-oxobutanoate	L-thréonine synthase	4.2.3.1	*	<i>thrC</i>
	L-thréonine déaminase	4.3.1.19	TD	<i>ilvA</i>
Voie de synthèse de l'isoleucine à partir de 2-acéto-2-hydroxybutanoate, ou de valine à partir d'acétolactate	Cétoacide réducto isomérase	1.1.1.86	AHAIR	<i>ilvC</i>
	Dihydroxyacide déshydratase	4.2.1.9	DHAD	<i>ilvD</i>
	Aminotransférase des acides aminés branchés	2.6.1.42	BcaaT	<i>ilvE</i>
Autres aminotransférases	Aspartate aminotransférase	2.6.1.1	TA	<i>aspB</i>
	Glutamate synthase	1.4.1.13	GS	<i>gltB</i>
	Histidinol phosphate transaminase	2.6.1.9	*	<i>hisC</i>
	Aminotransférase des acides aminés aromatiques	2.6.1-	AraT	<i>araT</i>
Synthèse et dégradation des corps cétoniques (acides gras) et dégradation de la valine/leucine/isoleucine	Hydroxymethylglutaryl-CoA synthase	2.3.3.10	*	<i>hmcM</i>
	Acétyl coenzyme A acétyltransférase	2.3.1.9	*	<i>fadA</i>

La détection de l'enzyme AHAI (ilvC) confirme que les gènes de l'opéron *ilvDBNCA* ont bien été transcrits, et qu'à minima leur ARNm devrait être détectable.

Afin de mieux appréhender l'absence de ces protéines, quelques éléments de réflexion sont proposés ci-après. Les prélèvements pour les protéomes ont été réalisés sur des cellules en phase de ralentissement de croissance, voire en phase stationnaire, donc faiblement actives au niveau transcriptionnel et traductionnel. Si les protéines ont été produites pendant les phases exponentielles de croissance, la diminution de leur expression en phase stationnaire peut expliquer que leur présence soit en dessous du niveau d'identification très élevé (FDR 1 %) que nous avons choisi. Dans l'analyse de Dressaire en 2011, le taux de FDR choisi était de 20 % dans le cadre d'une analyse globale [Dressaire et al. 2011]. Le taux sélectif choisi dans notre étude se justifie du fait du besoin de précision dû à la spécificité de notre étude sur des protéines cibles. De plus, les cultures réalisées ont été conduites à pH libre et il est donc possible que l'adaptation au stress acide masque, par son ampleur dans la dynamique de réponse enregistrée, des protéines cibles moins exprimées. Toutefois, des analyses de protéomique comparative ont déjà été menées par notre équipe en MCD sur des prélèvements en phase stationnaire à pH 4,2 [Mazzoli et al. 2010] et la majorité de ces protéines avait été précédemment détectée sur lait à pH 4,60, ce qui valide la faisabilité d'un tel choix. Il est également possible que les modifications nutritionnelles et/ou physico-chimiques appliquées n'aient pas conduit à des différences significatives de l'expression des protéines et que le niveau basal faible de ces protéines anaboliques explique ce résultat. La complémentarité des approches transcriptomique et protéomique prend, une fois de plus dans cet exemple, tout son sens pour voir si les cibles de régulation visées ont, ou n'ont pas, été affectées. Cependant, seule la mesure *in vitro* de l'activité de ces enzymes en phase exponentielle permettrait *in fine* de vérifier la présence de ces enzymes et leur état actif. Une autre remarque concerne l'activité d'acétohydroxyacide synthase que partagent les protéines ALS et AHAS. Il est possible que, même si ND5F ne produit que peu ou pas d'AHAS fonctionnelle, l'ALS puisse participer avec une moindre efficacité à la synthèse d'acéto-hydroxybutanoate.

I.4.2 Dosage des niveaux de protéines

Les résultats de l'analyse protéomique sont présentés ci-dessous, même si l'absence des protéines d'intérêt, AHAS, TD et autres enzymes impliquées dans la biosynthèse des BCAAs, diminue la portée des conclusions (Tableau IV. 7). Une fois les résultats normalisés, un test de Student a été effectué pour chacune des protéines afin de mettre en évidence les différences significatives entre les niveaux de protéines de la condition MCD (référence) et ceux des autres conditions testées. Un seuil de significativité correspondant à une *p*-value inférieure à 5 % a été retenu. Aucun filtre supplémentaire

n'a été rajouté sur la sélection des *p*-values significatives. Cependant, pour aider à l'analyse, les protéines dont le rapport d'expression entre la condition testée et la référence est supérieur à 1,5 ou inférieur à 1/1,5 (soit 0,67) seront soulignées pour accentuer la significativité de ces différences. Ainsi, si le ratio est inférieur à une valeur de 1/1,5, les protéines seront considérées comme sous exprimées dans la condition testée, et si le ratio est supérieur à une valeur de 1,5, les protéines seront surexprimées.

Il est observé des *p*-values significatives entre la référence et le milieu MCD-THR2X pour la LDH, l'AHAIR et l'aminotransférase des acides aminés aromatiques. D'après l'analyse du ratio d'expression, il y aurait une tendance à la sous-expression de ces trois enzymes dans les conditions testées par rapport à la référence. La diminution de la quantité d'AHAIR, enzyme consommatrice du précurseur de la pentane-2,3-dione, pourrait également expliquer l'augmentation du produit d'intérêt observée. L'absence de quantification de la TD ne permet pas d'envisager de changements significatifs de son expression du fait de la supplémentation en thréonine.

Dans le milieu MCD-Ile/Leu/Val à 0,2 X, il semblerait que seule l'aminotransférase des acides aminés branchés (BcaaT) soit significativement surexprimée comparativement au milieu MCD de référence pendant la phase stationnaire. Cette réponse était attendue et s'inscrit dans la logique d'une activation des voies de biosynthèse des BCAAs pour répondre aux besoins anaboliques suite à leur diminution dans le milieu. Cependant, cette activation des voies de biosynthèse ne peut être généralisée du fait de l'absence de confirmation de la surexpression de l'AHAIR. Bien que le ratio d'expression ne soit pas en-dessus de 0,67 pour la PFL (0,80), la tendance à la sous-expression s'inscrit dans une logique probable de redirection du flux de pyruvate vers l'AHAS pour répondre aux besoins cellulaires.

La majorité des variations d'expression des protéines d'intérêt a été observée lors de la culture MCD sous régulation d'oxygène (6 variations de ratio d'expression significatives sur les 11 protéines identifiées de la liste). En aval du pyruvate, il est logiquement retrouvé une sous-expression d'un facteur trois de la PFL (sensible à l'oxygène) et d'un facteur 1,6 de la LDH conformément aux données bibliographiques connues. De plus, l'activité de la pyruvate déshydrogénase (PDH) est activée en présence d'oxygène d'un facteur 2,38 en accord avec la bibliographie [Condon 1987]. Toujours en aval du pyruvate, et bien que non significative au niveau du filtre du ratio d'expression, il est possible de signaler aussi l'augmentation de l'expression de l'ALS d'un facteur 1,4 également démontrée par des études antérieures. Ces résultats démontrent la cohérence des informations issues de cette analyse.

Tableau IV. 7: Réponses protéomiques comparatives de *L. lactis* ND5F dans différentes conditions de culture face à la référence MCD. Les *p*-values du test de Student ne sont présentées que si *p*-value<0,05. Pour chaque *p*-value significative, le ratio d'expression entre la condition testée et la condition MCD de référence est présenté. Lorsque celui-ci est supérieur à 1,5 ou inférieur à 0,67, une flèche indique s'il y a sur ou sous expression. Abbréviation des protéines, voir Tableau IV. 6.

Catégorie fonctionnelle	Protéines	MCD-THR 2X		MCD-Ile/leu/val 0,2X		MCD-O ₂	
		p-value	Ratio d'expression	p-value	Ratio d'expression	p-value	Ratio d'expression
Directement en aval du pyruvate	LDH	0,002	0,79			4.10 ⁻⁵	0,62 ↘
	PFL			0,006	0,80	6.10 ⁻³	0,33 ↘
	PDH					1,8.10 ⁻⁹	2,38 ↗
Voie de synthèse de l'isoleucine ou de la valine/leucine.	ALS					1,2.10 ⁻³	1,39
	AHAIR	0,001	0,71			3,6.10 ⁻⁶	0,48 ↘
	BcaaT			0,000	1,64 ↗		
Métabolisme de la thréonine	L-Thréonine synthase					4.10 ⁻³	0,76
Autres aminotransférases	AraT	0,041	0,79			4,6.10 ⁻⁴	0,63 ↘
	GS					2,8.10 ⁻⁵	0,54 ↘
	Transaminase A					0,04.10 ⁻²	0,80
Métabolisme des acides gras	Hydroxyméthyl glutaryl-CoA synthase						

L'analyse des autres protéines montre que la présence d'oxygène a provoqué une diminution d'un facteur 2 de l'expression de l'AHAIR de la voie de biosynthèse des BCAAs (ratio d'expression de 0,48). Ce résultat, non rapporté précédemment dans la littérature, demande une confirmation expérimentale avant de pouvoir aller plus avant dans l'interprétation. Une hypothèse serait qu'en contrepartie de l'augmentation de l'ALS, et donc du flux de production d'α-acétolactate, une diminution de la quantité d'AHAIR éviterait un flux trop important vers les voies de biosynthèse des BCAAs. Cependant, ce résultat n'est pas confirmé par une modification de l'expression de la BcaaT dans ces mêmes conditions, ce qui implique une prudence dans l'interprétation.

La présence d'oxygène a également pour conséquence de diminuer respectivement d'un facteur 1,6 et 1,8 les quantités d'aminotransférase des acides aminés aromatiques (AraT) et de glutamate synthase (GS). La glutamate synthase catalyse la production de glutamine et d'α-cétoglutarate (KG) à partir de deux molécules de glutamate, et inversement. Il n'y a pas d'apport de glutamate dans le MCD, or comme toutes les bactéries lactiques, ND5F est auxotrophe pour le glutamate (résultat non présenté). Sa production est donc assurée par les glutaminases (enzymes ubiquitaires) qui garantissent un flux non limitant de glutamate à partir de la glutamine. L'absence de KG apporté par

le milieu et la non-fonctionnalité des enzymes du cycle des acides tricarboxyliques [Lapujade et al. 1998] provoque une extrême limitation en KG chez les bactéries lactiques, et exclu toute possibilité de synthèse de glutamate à partir de l'activité de la GS. Les hypothèses suivantes peuvent être avancées pour expliquer la diminution de la GS en présence d'O₂: i) l'augmentation de la croissance d'un facteur 2 par rapport à la condition de référence nécessite de consommer notamment plus de glutamate et de glutamine, ce qui peut conduire à une valeur limitante de ces deux éléments, qui pourrait induire une baisse de GS, ii) l'oxygénation induit la NADH oxydase qui contrôle l'équilibre redox via l'oxydation du NADH en NAD⁺. Le couple NADH/NAD étant impliqué dans la réaction enzymatique catalysée par la GS, on pourrait postuler qu'il exerce un contrôle sur la synthèse de GS. L'aminotransférase des acides aminés aromatiques (AraT) catalyse la transamination des acides aminés aromatiques en cétoacides par couplage avec la consommation de KG pour produire du glutamate. La limitation de la présence de KG est supposée être exacerbée par la consommation plus importante de Glu/Gln sous oxygénation (augmentation de la croissance sous O₂), ce qui pourrait expliquer la baisse observée dans la quantité de l'AraT.

Dans le milieu MCD-hypA comparativement au milieu MCD-AA 2X, seule la teneur de la BcaaT est significativement inférieure d'un facteur 2. La présence de peptides riches en BCAAs peut expliquer une réduction de l'activité de l'étape terminale de la biosynthèse des BCAAs.

II Modélisation de la singularité de la signature de *L. lactis* ssp. *Lactis* ND5F en matrice lait

La croissance des bactéries en général, et de *L. lactis* en particulier, dépend des conditions nutritionnelles (source de carbone, d'azote, de vitamines..) et de conditions physico-chimiques (pH, température, osmolarité). A la différence des cultures en conditions de laboratoires, la mise en œuvre des levains de *L. lactis* dans des processus industriels est souvent sous-optimale. Ainsi, les bactéries doivent faire face aux modifications de leur environnement nutritionnel ou physico-chimique (thermique, oxydatif, acide, osmotique) [Rallu et al. 1996; Bunthof et al. 1999; Sanders et al. 1999]. Au regard du domaine applicatif de *Lactococcus lactis* ssp. *lactis* ND5F restreint à la production de lait fermenté, les principaux facteurs de stress rencontrés par la souche sont l'acidité (production de lactate par la souche elle-même), l'oxydation (en début de croissance et conditionnement dans un environnement gazeux atmosphérique), la carence nutritionnelle, et l'hypothermie (stockage à 4°C). L'objectif de ces travaux est de modéliser le comportement de la souche *L. lactis* ssp. *lactis* ND5F en matrice lait face aux modifications simultanées de différents paramètres environnementaux afin de comprendre, prédire et ainsi contrôler, la typicité « crémeuse » de ND5F au cours du procédé.

II.1 Présentation de la démarche expérimentale

En particulier, il a été souhaité évaluer l'influence de paramètres technologiques contrôlables afin de proposer une optimisation future du procédé. Pour répondre à cette problématique applicative, la mise en équation des déterminants phénotypiques de ND5F et l'analyse des surfaces de réponses par la méthodologie des plans d'expériences a été appliquée pour planifier et étudier les différents essais de fermentations en environnement contrôlé.

II.1.1 Variables réponses pour décrire la signature en lait

La typicité de la souche ayant été définie à l'échelle phénotypique comme un assemblage de marqueurs métaboliques et physiologiques, il va être analysé individuellement puis conjointement plusieurs réponses pour obtenir une modélisation représentative de la signature phénotypique de la souche. Par la suite, cette réponse globale sera soumise à une optimisation. Il a été choisi en particulier d'examiner 20 réponses pour modéliser le comportement de la souche *L. lactis* ssp. *lactis* ND5F en matrice lait. Le choix de ces variables s'est fait à partir des connaissances acquises lors de l'analyse de la diversité des signatures phénotypiques des levains laitiers (voir chapitre II). Les

réponses ont ainsi été sélectionnées au sein des 14 marqueurs qui discriminent le plus la typicité de ND5F, auxquels ont été rajoutées 6 réponses pour élargir le spectre des informations collectées d'un point de vue physiologique (vitesse de réduction, potentiel REDOX) et métabolomique pour couvrir les grandes voies métabolique de production d'arômes : lipolyse, protéolyse et glycolyse. Le détail des réponses collectées est présenté Tableau IV. 8, ainsi que leur relation par rapport à la signature de ND5F. Certains marqueurs, bien que participant à la typicité de ND5F, n'ont été pas retenus dans cette liste du fait de l'absence de variation du composé (citrate) ou de forte corrélation avec un autre facteur (comme viabilité et biomasse à 14 J).

Tableau IV. 8 : Caractéristiques des vingt réponses sélectionnées pour construire le profil organoleptique de *L. lactis* ND5F. * Biomasse métaboliquement active. Pour précision, les réponses COVs sont analysées en

Variables	Famille de	Moment	Relation par rapport à la signature de ND5F	Variable

quantification relative.

			Présence	Faible présence ou Absence
Lactose (mM)	Sucre	J14	OUI	OUI
Lactate (mM)	Acide organique	J14		OUI
Acétate (Intensité relative)		J14		OUI
Vitesse maximale d'acidification , $V_{\max}^a$ (10^3 upH.min ⁻¹)		Phase de croissance	OUI	OUI
Temps où $V_{\max}^a$, $T_{\max}^a$ (h)		Phase de croissance	OUI	OUI
Durée de fermentation, D_{Tf} (h)	Indicateur technologique	Phase de croissance	OUI	OUI
Vitesse maximale de réduction, $V_{\max}^{\text{red}}$ (10^3 mV.min ⁻¹)		Phase de croissance	OUI	-
Eh_{7Tf} , Potentiel redox (mV)		Phase de croissance	OUI	-
Biomasse, B_{T14} (UFC, mL ⁻¹)*	Indicateur	J14		OUI
Glutamate (μM)	Acide aminé	J14		OUI
Glycine (μM)		J14		OUI
Pentane-2,3-dione		J14	OUI	OUI
Méthanthiol		J14	OUI	OUI
Diméthyldisulfide, DMDS		J14		OUI
Furanméthanol		J14		OUI
Acétate d'éthyle	COV	J14	OUI	-
Méthylbutanal		J14		OUI
Diacétyl		J14		OUI
Acétaldéhyde		J14		OUI
Acide hexanoïque		J14		OUI

II.1.2 Sélection des facteurs et du domaine expérimental

Pour comprendre l'évolution de la signature métabolique, physiologique et aromatique, de *L. lactis* ssp. *lactis* ND5F, quatre facteurs de variabilité de l'environnement fermentaire ont été sélectionnés et étudiés simultanément :

- Le premier paramètre concerne la température de fermentation qui affecte plusieurs réponses telles que la croissance bactérienne en lait fermenté [Radke-Mitchell & Sandine 1986], la rhéologie du lait fermenté [Béal & Corrieu 1994] et la production aromatique (acétaldéhyde, acétate, 2-butanone, acétone, et diacétyl) [Gardini et al. 1999]. Ce paramètre est d'autant plus important que l'utilisation en culture mixte avec des ferments thermophiles impose des températures de production au-delà de 35°C et éloignées de l'optimum mésophile de croissance de ND5F.

- L'apport en matière protéique, et de manière générale, le taux en matières sèches du lait non lipidique, est ajusté par le pourcentage en poudre de lait écrémé solide. Ce taux intervient sur des paramètres rhéologiques (viscosité, fermeté, synérèse) [Becker & Puhán 1989; Biliaderis et al. 1992; Gastaldi et al. 1997] ainsi que sur la croissance bactérienne [Tamine & Deeth 1999] et les propriétés sensorielles [Biliaderis et al. 1992; Gardini et al. 1999].

- Le pourcentage en matière grasse du lait interagit sur l'activité acidifiante, la rhéologie finale du produit [Gardini et al. 1999; Torriani et al. 1996] et l'activité lipolytique des bactéries [Urbach 1995].

- Enfin, l'environnement gazeux (oxydant-ou réducteur), et notamment la concentration en oxygène dissout, influence exponentiellement le potentiel d'oxydoréduction (Eh) [Tengerdy 1961], la cinétique d'acidification et la croissance [Jeanson et al. 2009], et modifie le métabolisme, comme vu précédemment. De plus, il a été démontré que différents Eh₇ appliqués à des cellules de *Lactococcus lactis* en phase stationnaire provoquaient des modifications dans les voies de catabolisme des acides aminés, et donc des signatures aromatiques différentes [Kieronczyk et al. 2006]. Ces facteurs ont déjà été testés sur les bactéries lactiques suivantes, *Streptococcus thermophilus*, *Lactobacillus acidophilus* ou *Lactobacillus delbrueckii* ssp. *bulgaricus* [Torriani et al. 1996; Gardini et al. 1999; Kristo et al. 2003], mais peu investigués, et de plus, indépendamment les uns des autres.

Après concertation avec la société Yoplait, les plages d'étude de ces différents facteurs ont été définies afin de correspondre aux réalités industrielles des domaines d'utilisations de la souche *L.*

lactis ND5F, vis-à-vis du type de substrat laitier utilisé ou de la température de fermentation (Tableau IV.9).

Tableau IV. 9 : Plage d'étude des quatre facteurs étudiés dans le plan d'expériences de surface de réponses.

Facteurs	Limite inférieure (-1)	Point au centre(0)	Limite supérieure (+1)
MS : Matière sèche (g.Kg ⁻¹)	33	43	53
MG : Matière grasse (g.Kg ⁻¹)	0	20	40
TP : Température (°C)	28	34	40
O ₂ : Oxygénation (PO ₂ du lait avant inoculation)	0	50	100

II.1.3 Définition du modèle empirique et construction de la matrice expérimentale

Les modèles recherchés pour l'étude de surface de réponses correspondent à des modèles polynomiaux du second degré. Cette hypothèse sera vérifiée à posteriori grâce à une analyse statistique du modèle obtenu, afin de confirmer que la surface de régression donne une interprétation utilisable du phénomène réel. Pour chaque réponse, les quatre facteurs étudiés, i) matière sèche (MS), ii) matière grasse (MG), iii) PO₂ du lait avant inoculation (O₂) et iv) température (TP) sont présumés avoir une influence statistiquement significative sur les variations observées. En l'absence de toute information sur la fonction qui lie la réponse aux facteurs, il est admis, à priori, une loi d'évolution dont la formulation la plus générale est la suivante :

$$Y = f(X_{MS}, X_{MG}, X_{O_2}, X_{TP})$$

Equation IV. 1

Dans ce contexte, pour chaque réponse étudiée (Y), le modèle postulé comprendra p coefficients (β) à identifier qui sont présentés par l'équation suivante :

$$MS \times O_2$$

Equation IV. 2

Le plan d'expériences devra donc être constitué d'au minimum 15 expériences pour identifier les 15 coefficients. Il s'agit alors d'identifier le nombre d'expériences permettant d'obtenir la meilleure matrice possible, puis d'augmenter sa taille afin d'observer si la qualité de la matrice ainsi obtenue justifie le surcoût engendré par l'ajout d'un traitement supplémentaire. Ce travail constitue l'étape de construction du plan d'expériences. Dans le cas présent, il a été choisi de réaliser un plan de surface de réponses en mesurant les réponses pour les trois niveaux intermédiaires définis (-1, 0, +1) sur chacun des facteurs. Après analyse des différentes alternatives à notre disposition, il a été décidé d'effectuer un plan dit Plan de Box-Behnken (qui comprend, de base, 27 expériences). Ce plan a été adapté pour tenir compte des impératifs expérimentaux, dont notamment une limite de production

de quatre types de laits différents au maximum par semaine. Afin de prendre en compte également les variations liées aux différences de batch de productions de laits, et d'estimer au mieux la variabilité biologique et technique tout au long de cette période expérimentale, chaque journée d'expérimentation a compris la répétition d'une même expérience qui va permettre lors de l'analyse d'estimer l'erreur expérimentale. Cette expérience correspond au point au centre ; fermentation pour laquelle les 4 facteurs sont affectés de la valeur au centre de leur domaine d'étude (0, 0, 0, 0). Ainsi, cette expérience a été répétée huit fois pour les journées de fermentations définies. Pour finir, l'ordre des expériences (32 au total) a été 'randomisé' par l'intermédiaire d'un tirage au sort afin de rajouter une part d'aléatoire au sein du planning contraint des productions de laits. Au total, ce sont donc 32 fermentations à pH libre qui ont été exécutées sur différents laits et différentes conditions (température, oxygène). La matrice est présentée dans la section Matériels & Méthodes. Les cultures ont été stoppées lorsque le pH 4,60 a été atteint, pour mimer le procédé industriel, et conservées 14 J à 4°C pour mesures et dégustation.

II.2 Analyse des réponses individuelles et modélisation des surfaces de réponses

II.2.1 Modélisations des réponses individuelles et évaluation de la qualité des modèles

L'analyse de chaque réponse individuelle a été découpée en deux parties. La première a consisté en une résolution mathématique afin d'identifier les valeurs des coefficients β pour chaque modèle à partir des résultats des 32 expériences, par application de la méthode des moindres carrés. Elle a été ensuite suivie d'une étude statistique afin d'analyser, pour chaque modèle établi, la significativité des coefficients du modèle, le pourcentage de variance expliquée et la qualité prédictive. Enfin, il a été vérifié les hypothèses d'analyses : normalité de la distribution des résidus, homoscedasticité de la variance, et présence de valeurs aberrantes pour les modèles retenus.

Au vu du nombre important de réponses étudiées, les différentes étapes de l'analyse des résultats détaillées dans la section Matériel & Méthodes ne seront pas reprises dans cette section. Les équations simplifiées³ pour chacune des 20 réponses sont présentées par type d'indicateur. Pour chaque réponse, l'équation du modèle, la qualité descriptive du modèle en calculant les coefficients de corrélation R , R^2 et $R^2_{\text{ajusté}}$ qui tient compte d'une pénalisation en fonction du nombre de coefficients, sont déterminés. La qualité prédictive de chaque modèle est ensuite estimée à partir du

³ Il a été choisi délibérément de présenter des modèles simplifiés pour chaque réponse en ne conservant que les coefficients aux termes significatifs dans le modèle, même si d'autres coefficients du marqueur

coefficient de corrélation prédictif Q^2 . L'analyse des modèles a été conduite avec une mise en perspective biologique afin de caractériser l'impact de chaque facteur.

a. Sucre et acides organiques

La concentration résiduelle en lactose à 14 J n'est significativement dépendante que de la quantité de matière sèche du lait, c'est-à-dire du pourcentage de poudre de lait écrémé ajouté dans la préparation du lait (Tableau IV. 10). En effet, la poudre de lait écrémé, en plus de contenir une source protéique, comprend du lactose dont l'influence sur la préparation du lait est souvent sous-estimée. Le modèle simplifié présenté permet de prédire à près de 70 % cette concentration finale dans les produits fermenté par ND5F. En toute logique, les concentrations de lactate et acétate, produits du catabolisme du lactose, dépendent principalement du pourcentage en matière sèche et donc du pourcentage en lactose du lait. Le facteur température contribue également à déterminer le titre final en lactate, avec un effet linéaire et quadratique négatif lors de l'augmentation des valeurs vers des températures plus thermophiles. Cette diminution de la concentration en lactate en fonction de la température peut éventuellement traduire une diminution de l'activité optimale de croissance et d'acidification, à confirmer sur les autres réponses. Une interaction entre la concentration initiale en oxygène et la température est identifiée, bien que cette participation ne contribue qu'à expliquer 8 % du phénomène. Cependant, il est connu et il a été prouvé que la présence d'oxygène concourt à un shift du métabolisme homolactique vers un métabolisme mixte et à une réduction du flux de pyruvate pris en charge par la LDH. Le pourcentage en matière grasse du lait intervient sur la concentration finale en acétate (effet linéaire et interaction avec l'oxygène), mettant ainsi en lumière la participation des voies de lipolyse à cette production. Là encore, l'oxygène apparaît comme un paramètre modulateur de cette concentration. Les deux équations relatives au lactate et à l'acétate présentent des scores de qualité de détermination basés sur le R^2_{adj} de respectivement 73 et 50 % (Tableau IV. 10), indiquant dans le cas de l'acétate qu'un ou plusieurs autres facteurs non pris en compte dans cette étude doit(vent) interagir avec les voies de production d'acétate.

b. Indicateurs technologiques et physiologiques

A l'exception du potentiel d'oxydoréduction à l'arrêt de culture, les cinq autres modèles établis qui décrivent le comportement d'acidification (V^a_{max} , T^a_{max}), la vitesse maximale de réduction du lait (V^{red}_{max}), la durée de fermentation (D_{Tr}) et la biomasse finale à 14 J, présentent des R^2_{adj} supérieurs ou égaux à 0,67 (Tableau IV. 11). Ce pourcentage atteste de la qualité descriptive des modèles simplifiés proposés et de la pertinence, pour ces réponses étudiées, des quatre facteurs sélectionnés.

Tableau IV. 10 : Caractérisation des modèles déterminés pour chaque réponse individuelle de *L. lactis* ssp. *lactis* ND5F à partir des quatre facteurs étudiés. Cas des réponses sucre et acides organiques. R : coefficient de corrélation, R^2 coefficient de détermination, R_{adj}^2 coefficient R^2 ajusté en fonction du nombre de coefficients pris en compte et Q^2 , coefficient prédictif. ε : erreur du modèle.

Réponse	Equation	Pourcentage de chaque effet			R	R^2	R_{adj}^2	Q^2
		Effet I	Effet II	Interaction				
Lactose (mM)	$[Lactose_{T14}] = 164,57 + 34,30 \times X_{MS} + \varepsilon$	100	-	-	0,86	0,73	0,72	0,68
Lactate (mM)	$[Lactate_{T14}] = 83,18 + 21,24 \times X_{MS} - 9,21 \times X_{TP} - 13,02 \times X_{TP}^2 + 13,14 \times X_{TP} \times X_{O_2} + \varepsilon$	76,63	15,14	8,23	0,87	0,76	0,73	0,57
Acétate (quantification relative en GS-MS)	$[Acétate_{T14}] = 0,431 + 0,078 \times X_{MS} + 0,123 \times X_{MG} - 0,080 \times X_{O_2}^2 + 0,128 \times X_{MG} \times X_{O_2} + \varepsilon$	69,40	12,70	17,90	0,77	0,60	0,50	0,27
	Equation	Pourcentage de chaque effet						
		Effet I	Effet II	Interaction				

$[V_{\max}^a] = -7,92 - 1,18 \times X_{O_2} + 1,02 \times X_{TP} + 1,04 \times X_{O_2}^2 + 1,56 \times X_{TP}^2 - 1,08 \times X_{MS} \times X_{TP} + 1,23 \times X_{MG} \times X_{TP} + \varepsilon$	49,82	42,28	7,90
$[T_{\max}^a] = 4,35 - 0,56 \times X_{O_2} + 1,03 \times X_{O_2}^2 + 2,23 \times X_{TP}^2 + \varepsilon$	8,10	91,90	-
$[V_{\max}^{red}] = -13,87 + 2,72 \times X_{TP} + 5,24 \times X_{O_2}^2 + 3,37 \times X_{TP}^2 - 5,85 \times X_{MS} \times X_{O_2} + \varepsilon$	17,85	54,75	27,40
$EH_{TF}^7 = 164,47 + 63,14 \times X_{TP} + \varepsilon$	100	-	-
$D_{TF} = 9,70 + 4,05 \times X_{TP} + 4,09 \times X_{TP}^2 - 2,80 \times X_{MG} \times X_{TP} + \varepsilon$	48,18	44,14	7,68
$0,57 \times X_{MP} + 0,61 \times X_{MG} - 0,42 \times X_{O_2} - 0,64 \times X_{TP} + 0,81 \times X_{MS}^2 - 0,49 \times X_{MS} \times X_{O_2} + 0,61 \times X_{MG} \times X_{TP} + 1,55 \times X_{TP} \times X_{O_2} + 0,73 \times X_{MS} \times X_{TP} + \varepsilon$	44,55	14,49	20,96

Tableau IV. 11 : Caractérisation des modèles déterminés pour chaque réponse individuelle de *L. lactis ssp. lactis ND5F* à partir des quatre facteurs étudiés. Cas des réponses liées à des indicateurs physiologiques et technologiques. R : coefficient de corrélation, R² coefficient de détermination, R²_{adj} coefficient R² ajusté en fonction du nombre de coefficients pris en compte et Q², coefficient prédictif. En gris, un modèle jugé peu descriptif de la réponse étudié. ε: erreur du modèle.

En particulier la vitesse maximale d'acidification ($V_{\max}^a$) a été modélisée par une équation faisant intervenir les 4 facteurs étudiés. Ce modèle reflète ainsi la complexité biologique de cette activité, dépendante à la fois des capacités acidifiantes de la souche en fonction de son environnement, mais également de l'effet tampon du milieu. Il est à noter que les valeurs de la réponse $V_{\max}^a$ sont négatives pour décrire l'acidification, et que par conséquent, le modèle a été construit pour décrire cette évolution. Ainsi tout coefficient négatif est à analyser comme augmentant cette vitesse. Les termes au carré et d'interactions sont analysés en premier car ils conditionnent l'interprétation des effets linéaires. Ainsi, l'oxygénation initiale du lait et la température sont les deux seuls termes au carré de ce modèle, et ils sont responsables de l'explication des variations courbes de la surface de réponse. Ces deux effets ont une action négative sur l'augmentation de la $V_{\max}^a$ (coefficients positifs). Ces résultats sont en adéquation avec la littérature ; il a ainsi été démontré qu'une diminution initiale de l'oxygène dissout permet d'augmenter l'activité acidifiante de *Lactococcus lactis* en lait [Jeanson et al. 2009]. Cet effet est également retrouvé sur le temps maximal avant d'atteindre la $V_{\max}^a$ qui est augmenté par la présence d'oxygène (coefficient linéaire négatif). Le modèle de la $V_{\max}^a$ contient également deux effets d'interaction à deux facteurs impliquant la matière sèche et la matière grasse liés à la température de fermentation. Ainsi, l'effet de la température dépend du pourcentage de matière sèche et de matière grasse. Ces résultats sont en partie similaires à ceux obtenus par Kristo sur l'interaction entre le pourcentage de matière sèche et la température vis-à-vis de l'activité acidifiante [Kristo et al. 2003]. Cependant, ces paramètres nutritionnels ont des effets antagonistes sur la réponse du fait du signe des coefficients (augmentation pour MS et diminution dans le cas de MG). L'interprétation des deux effets linéaires des facteurs oxygénation initiale et température demande une certaine prudence au vu de leur participation à d'autres effets. Une hiérarchisation des effets en fonction de leur contribution indique que les interactions représentent des effets mineurs avec moins de 8 %. L'analyse des surfaces de réponses par facteurs deux à deux, présentée ci-dessous, participe à clarifier les conséquences de chacun (Figure IV. 8).

La surface de réponse Figure IV. 8 A, confirme la présence d'un minimum à partir des facteurs oxygène et température sur ce domaine d'étude (soit un point où $V_{\max}^a$ serait la plus importante). Alors que la matière sèche n'a que peu d'incidence sur la $V_{\max}^a$ lors de l'interaction avec l'oxygène (Figure IV. 8 B), elle interagit plus fortement avec la température pour en moduler l'impact (Figure IV. 8.C).

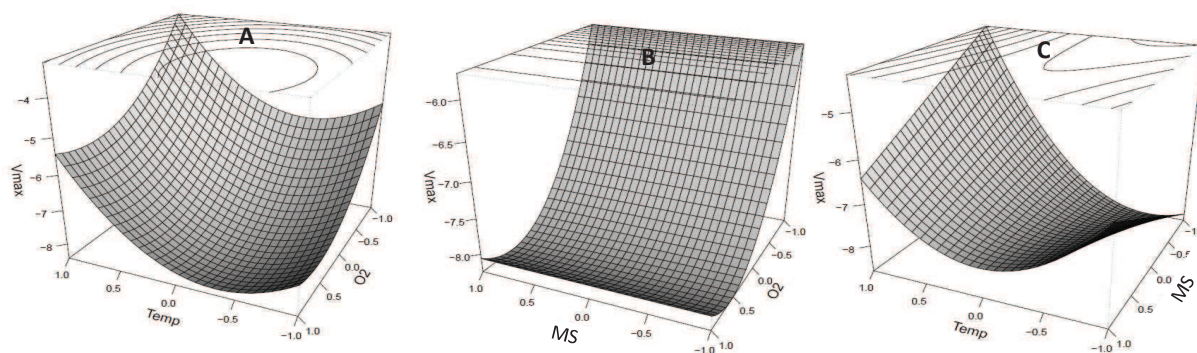


Figure IV. 8 : Surfaces de réponses (perspectives et graphique de contours) de la vitesse maximale d'acidification, $V_{\max}^a$ de *L. lactis* ssp. *lactis* ND5F en lait en fonction des trois facteurs sélectionnés : température (Temp), matière sèche (MS) et PO_2 initiale (O_2). A) $V_{\max}^a$ en fonction de la température et de la PO_2 initiale, B) $V_{\max}^a$ en fonction de la matière sèche et de la PO_2 initiale et C) $V_{\max}^a$ en fonction de la température et de la matière sèche. Le troisième facteur est assigné par défaut à la valeur au centre de son domaine pour chaque surface de réponse.

Concernant la vitesse maximale de réduction ($V_{\max}^{\text{red}}$), elle aussi négative, la modélisation obtenue peut être considérée comme fiable, avec un R^2_{adj} de 0,81. Le facteur température intervient au niveau des coefficients linéaires et quadratiques pour traduire une diminution de $V_{\max}^{\text{red}}$ (coefficients positifs). Cette observation est vraisemblablement liée à deux phénomènes concomitants : i) le calcul mathématique du Eh_7 qui intègre une correction par une diminution du potentiel de référence de E_r en fonction de la température et ii) l'effet négatif sur l'activité cellulaire de la température qui diminue la réduction du milieu. L'oxygène, en tant que principal oxydant du milieu, a logiquement un effet négatif (terme au carré positif) sur cette réponse. Cependant, son impact sur la $V_{\max}^{\text{red}}$ s'effectue également en interaction avec la composition en matière sèche du lait, et donc il semblerait que l'apport protéique contribuerait à renforcer, par l'effet sur la croissance, l'activité réductrice de la bactérie.

La durée de fermentation, où temps nécessaire pour atteindre le pH de 4,60, est un paramètre technologique important dans l'optimisation des coûts de production. D'après le modèle (R^2_{adj} de 0,67), toute augmentation de température prolonge la durée de fermentation et donc l'immobilisation des cuves (coefficients positifs pour les termes linéaires et quadratiques) en réduisant l'activité acidifiante de la souche (augmentation de $T_{\max}^a$ et diminution de $V_{\max}^a$). Une interaction, de contribution faible à la variance expliquée des données (8 %), a été identifiée entre la matière grasse et la température. Le coefficient négatif de cette interaction pourrait éventuellement être interprété comme un marqueur de l'effet protecteur de la matière grasse face à l'augmentation de la température.

La réponse relative à la quantité de biomasse métaboliquement active à 14 J, ou B_{T14} , est le modèle établi le plus complexe parmi les vingt étudiés. Il est composé de neuf coefficients significatifs. Le R^2_{adj} révèle que l'équation proposée est capable de résumer près de 83 % de la variabilité des observations, ce qui est un score tout à fait satisfaisant pour une variable aussi complexe. En particulier, l'ensemble des quatre facteurs sélectionnés contribue en tant que coefficients linéaires, quadratique ou termes d'interactions, ce qui reflète l'interdépendance de ces différents facteurs et la complexité du phénomène biologique. De plus, l'évaluation de la qualité prédictive donnée par le Q^2 s'élève à 63 %. Ce paramètre physiologique résume à lui seul les principales conclusions déduites des autres réponses métaboliques et technologiques. L'augmentation du pourcentage en matière grasse et en matière sèche démontre un effet bénéfique sur la croissance cellulaire et la survie à 14 J à 4°C dans le lait fermenté pour ND5F. La présence initiale d'oxygène et l'augmentation de la température diminuent, elles, la biomasse. Seule l'analyse des surfaces de réponses des facteurs deux à deux permettrait de prendre en compte la modulation de ces effets par les coefficients d'interactions identifiés et de conclure quant à un optimum de fonctionnement possible pour maximiser la biomasse.

c. Acides aminés et composés organiques volatils

A la différence des précédentes réponses étudiées, la qualité descriptive des modèles construits pour ce groupe d'indicateurs de la signature organoleptique de ND5F est faible, avec des R^2_{adj} compris entre 0,24 et 0,57 qui ne permettent pas de disposer d'une grande fiabilité dans chacune de ces équations (Tableau IV. 12). Ces taux illustrent la difficulté à modéliser ces réponses à partir des facteurs étudiés. Ainsi, il semblerait que la variabilité de ces réponses repose sur d'autres facteurs plus adéquats pour décrire ce phénomène. Il peut être cité le potentiel d'oxydoréduction connu pour affecter la production aromatique des souches de lactocoques, sur lequel la PO_2 initiale n'interagit que transitoirement lors de la maturation, ou encore la composition en peptides résiduels dans le lait (source d'azote et précurseurs du catabolisme des acides aminés). Une option serait d'intégrer dans la modélisation des paramètres propres à l'activité cellulaire comme l'acidification et le dosage d'activité enzymatique, plus proches peut-être de la réponse. Cependant, ces paramètres sont, pour la plupart, des facteurs non contrôlables et apparaissent donc comme difficilement intégrables dans une surveillance, voire une optimisation, du procédé industriel. Néanmoins, d'un point de vue de la compréhension métabolique, ces facteurs pourraient aider à mieux appréhender les paramètres contrôlant ces réponses.

Lorsqu'il est comparé la variabilité du point au centre (expérience répliquée 8 fois tout au long de la période expérimentale) à celle des expériences du plan, il est observé qu'elle représente entre 30 et 70 % de la variabilité des réponses COVs. Ce pourcentage très haut révèle que d'autres sources de variabilité non prises en compte, et donc non contrôlées dans cette expérience, ont participé à hauteur équivalente avec les 4 facteurs, aux réponses observées. Ce résultat peut être interprété comme l'expression de la variabilité biologique de la souche *L. lactis ssp. lactis ND5F* à son environnement. En effet, cette vision est également confortée par une variabilité des réponses observées, plus faibles que pour les autres réponses (distribution comprise en moyenne entre ± 30 % autour de la médiane pour ces COVs, contre des distributions pouvant aller jusqu'à plus de 100 % pour les autres réponses), ce qui témoigne d'un impact réduit de ces facteurs sur le domaine d'étude. Cependant, bien qu'expliquant une variabilité limitée de nos jeux de données, les modèles pour lesquels le R^2_{adj} est supérieur à 0,50, vont être commentés ainsi que celui de la pentane-2,3-dione en raison de son caractère déterminant pour la signature organoleptique.

L'équation pour la pentanedione, dont le contrôle de la réponse a déjà été exploré précédemment, révèle une participation significative de trois facteurs à son titre final à 14 J : la température, et les pourcentages en matière grasse et en matière sèche, qui expliquent 39 % de la variabilité. L'effet négatif de la température sur le taux de pentanedione contribue à résumer près de 47 % de la variance expliquée modélisée. La matière sèche interagit à la fois avec la matière grasse et la température, avec des effets contraires sur la production de pentanedione. Pour préciser ces effets, les surfaces de réponses sont présentées Figure IV. 9. L'observation des surfaces permet de déterminer que, face à la température, ni la matière grasse ni la matière sèche n'ont une influence suffisante pour moduler la réponse linéaire entre la température et la quantité de pentanedione (maximale quand la température diminue, voir Figure IV. 9 A et B). Cette donnée recoupe les observations empiriques réalisées par Yoplait pour définir la température optimale de production aromatique de la souche. Cependant, la surface de réponse entre la matière grasse et la matière sèche révèle qu'un point de fonctionnement (plateau de stabilité de la réponse) est identifiable au centre du domaine de chacun de ces facteurs (Figure IV. 9 C). Au-delà de cette composition du lait, la réponse, bien que pouvant être plus importante, est instable, ce qui ne permet pas de garantir la reproductibilité de résultats pour de faibles variations de ces deux facteurs. Là encore, cette observation peut expliquer le manque de reproductibilité constaté lors de formulations divergentes du lait. Les coordonnées de cette zone où la réponse pentanedione est stable et égale à $0,0321 \pm 0,003$ ont été déterminées (MS : $0 \pm 0,25$, MG : $0 \pm 0,25$ et TP=-1 en valeurs codées).

Tableau IV. 12 : Caractérisation des modèles déterminés pour chaque réponse individuelle de *L. lactis ssp. lactis ND5F* à partir des quatre facteurs étudiés. Cas des réponses liées à des COVs. R : coefficient de corrélation, R² coefficient de détermination, R²_{adj} coefficient R²ajusté en fonction du nombre de coefficients pris en compte et Q², coefficient prédictif. En gris, les modèles jugés peu descriptif de la réponse étudiée (R_{adj}²>0,35). ε: erreur du modèle. Certaines réponses ont été transformées au préalable pour établir le modèle afin de satisfaire les hypothèses mathématiques nécessaires à la résolution des moindres carrés. A l'exception des deux acides aminés, les réponses sont en aires relatives obtenues par GC-MS.

Réponse	Equation	Pourcentage de chaque effet			R	R ²	R _{adj} ²	Q ²
		Effet I	Effet II	Interaction				
Glu _{T14} (μM)	$[Glu_{T14}]^2 = 1,54 \cdot 10^5 + 4,59 \cdot 10^5 \times X_{MS} - 5,70 \cdot 10^5 \times X_{MS} \times X_{TP} + \varepsilon$	66,03	-	33,97	0,65	0,43	0,39	0,29
Gly _{T14} (μM)	$\frac{1}{\sqrt{[Gly_{T14}]}} = 0,147 + 0,020 \times X_{TP} - 0,262 \times X_{MS} \times X_{MG} + \varepsilon$	66,10	-	36,90	0,54	0,29	0,24	0,15
Pentane-2,3-dione	$[Pentanedione] = 0,0247 - 0,0064 \times X_{TP} - 0,0083 \times X_{MS} \times X_{MG} + 0,0005 \times X_{MS} \times X_{TP} + \varepsilon$	47,65	-	52,35	0,67	0,45	0,39	0,25
Méthanthiol	$[Methanthiol] = 0,00643 - 0,00127 \times X_{TP} + 0,00256 \times X_{TP} \times X_{MS} - 0,0032 \times X_{TP} \times X_{O_2} + \varepsilon$	22,15	-	77,85	0,65	0,42	0,36	0,23
DMDS	$[\sqrt{DMDS}] = 0,1246 + 0,0329 \times X_{MS} - 0,051 \times X_{TP} + 0,0467 \times X_{MS} \times X_{TP} + \varepsilon$	83,59	-	16,41	0,75	0,57	0,52	0,38
Furanméthanol	$Ln([Furanméthanol]) = -2,689 + 1,114 \times X_{TP} - 1,328 \times X_{MG}^2 + \varepsilon$	53,45	46,55	-	0,60	0,36	0,31	0,21
Acétate d'éthyle	$Log_{10}([Acétate d'éthyle]) = -2,057 - 0,161 \times X_{TP} - 0,269 \times X_{MG}^2 - 0,220 \times X_{TP}^2 + \varepsilon$	53,67	46,33	-	0,58	0,33	0,25	0,09
Méthylbutanal	$Log_{10}([Méthylbutanal]) = 1,756 + 0,377 \times X_{MS} - 0,190 \times X_{TP} - 0,287 \times X_{MG}^2 + 0,368 \times X_{MS} \times X_{O_2} + \varepsilon$	65,13	18,38	16,49	0,79	0,63	0,57	0,42

[THESE CONFIDENTIELLE]

Diacétyl	$\frac{1}{\sqrt{[Diacétyl]}} = 1,202 - 0,184 \times X_{MS} - 0,160 \times X_{TP}^2 - 0,227 \times X_{MS} \times X_{TP} + \varepsilon$	50,89	23,55	23,56	0,62	0,39	0,32	0,21
Acétaldéhyde	$\ln([Acétaldéhyde]) = -3,582 - 0,297 \times X_{O_2} + 0,558 \times X_{MS} \times X_{TP} + 0,569 \times X_{MG} \times X_{O_2} + \varepsilon$	29,45	-	70,55	0,68	0,46	0,40	0,26
Acide hexanoïque	$[Acide\ hexanoïque] = 0,799 + 0,258 \times X_{MG} + \varepsilon$	100	-	-	0,57	0,32	0,30	0,18

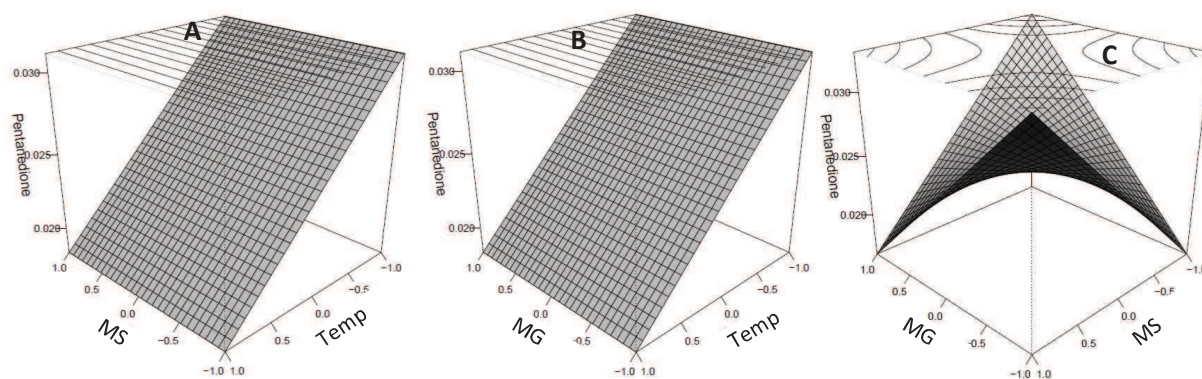


Figure IV. 9 : Surfaces de réponses (perspectives et graphiques de contours) de la vitesse de production de pentane-2,3-dione par *L. lactis* ssp. *lactis* ND5F en lait en fonction des trois facteurs sélectionnés : température (Temp), matière sèche (MS) et matière grasse (MG). A) Pentane-2,3-dione en fonction de la température et de la matière sèche, B) Pentane-2,3-dione en fonction de la matière grasse et de la température et C) Pentane-2,3-dione en fonction de la matière grasse et de la matière sèche. Le troisième facteur est assigné par défaut à la valeur au centre de son domaine.

L'absence d'influence significative de l'oxygène sur le taux de pentanedione peut paraître surprenante au vu de la précédente investigation. Une explication possible est que la régulation initiale de la PO_2 est un phénomène transitoire qui survient lorsque la concentration en biomasse est encore faible (1.10^8 UFC.mL⁻¹), ce qui diminuerait son impact potentiel.

Le diméthylsulfide (DMDS) et le méthylbutanal sont des composés faiblement présents dans le cadre de la spécificité organoleptique de ND5F. Modéliser l'impact des facteurs environnementaux testés permet de décrire l'évolution de leur réponse à respectivement 52 et 57 % afin éventuellement de cibler les conditions opératoires permettant de garantir un niveau bas de ces molécules. Comme attendu, le DMDS, produit du catabolisme de la méthionine, augmente avec le pourcentage de matière sèche, source protéique, et diminue à forte température comme la majorité des produits du métabolisme. Ces deux coefficients linéaires représentent à eux seuls 83 % de la variance expliquée par le modèle. Ces deux facteurs sont de nouveau identifiés comme ayant les mêmes effets leviers sur le taux de méthylbutanal, produit du catabolisme de la leucine. De plus, une contribution de la matière grasse minoritaire (18 %) est observée pour moduler cette synthèse. L'oxygène et la matière sèche interagissent ensemble, comme le révèle l'interaction au coefficient positif entre ces deux termes.

II.2.2 Implication des différents facteurs à l'échelle de la signature globale de la souche

Dans un but d'obtenir une vision globale de l'impact des différents facteurs sur la signature de ND5F, et non plus réponse par réponse, il a été analysé, pour chaque facteur, la proportion de coefficients significatifs identifiés dans les vingt réponses observées (résultats non pondérés par le nombre ou le type d'effet du coefficient dans une même équation). Il est ainsi observé que les quatre facteurs sont impliqués au minimum dans 48 % des modèles établis (Figure IV. 10 A). Ce résultat confirme ainsi la pertinence de ces facteurs pour la majorité des réponses sélectionnées. Il peut être également souligné que la température atteint une proportion de 76 % et peut donc être considérée comme le levier prépondérant pour moduler ces réponses.

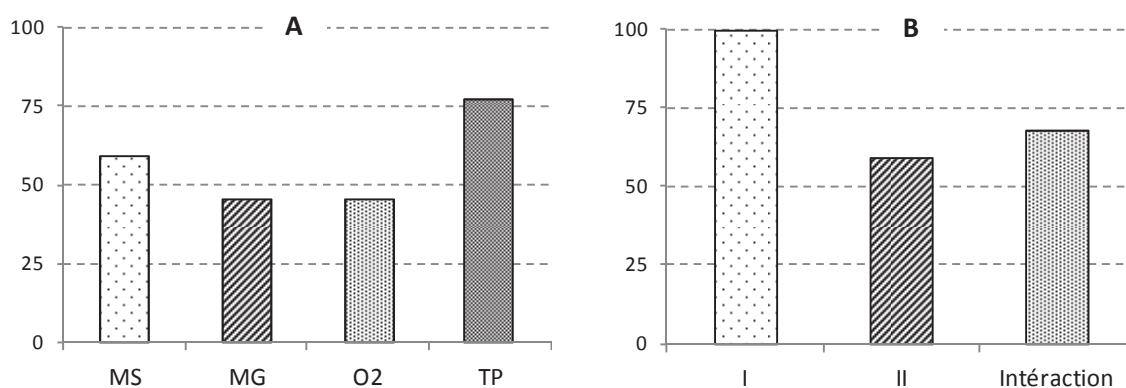


Figure IV. 10 : Implication des quatre facteurs, matière sèche (MS), matière grasse (MG), PO₂ initiale (O₂) et température (TP), à l'échelle de la signature globale de *L. lactis ssp. lactis* ND5F en lait. A) Analyse du poids relatif de chaque facteur et B) Analyse du type d'effet majoritaire : effet linéaire (I), quadratique (II) ou d'interaction (Interaction).

Un autre point concerne le type d'effet majoritairement retrouvé dans ces modèles (Figure IV. 10 B). Il est ainsi découvert que chaque modèle comprend à minima un effet linéaire à partir duquel une première orientation de la réponse est facilitée. Comme déjà observé, seulement 60 % des équations présentent des termes quadratiques. La présence dans près de 70 % des équations d'interactions révèle l'interdépendance des effets des facteurs les uns par rapport aux autres, et confirme la nécessité de s'appuyer sur l'équation pour prédire avec exactitude l'impact de ces modifications.

II.3 Recherche d'un point de fonctionnement permettant l'optimisation de la réponse globale

Après avoir formalisé indépendamment l'effet de quatre paramètres sélectionnés sur des éléments identifiés comme fondateurs de la signature de ND5F, il est désormais nécessaire de mettre en place une analyse du comportement global de la souche. Seules les réponses pour lesquelles le coefficient de détermination R^2_{adj} et le coefficient prédictif Q^2 étaient respectivement supérieurs à 0,25 et 0,10 ont été sélectionnées pour cette optimisation, soit 18 variables.

Dans un premier temps, les critères d'optimisation pour chaque réponse ont été définis (minimisation, maximisation ou atteinte d'une valeur cible) en adéquation avec la précédente analyse de spécificité de ND5F. En effet, l'objectif est d'assigner à chacune des réponses des valeurs permettant d'augmenter le caractère de la signature « crémeuse » de *L. lactis ssp. lactis* ND5F en lait. Pour cela, dans le cas d'un choix d'une valeur cible, la moyenne des six réplicats biologiques de ND5F lors de l'étude de typicité phénotypique de la souche a servi de point de départ (chapitre II), et a été ajustée pour maximiser la contribution « crème » ou « base neutre » en fonction des réponses. Les valeurs assignées pour les différentes bornes (inférieures ou supérieures) sont choisies en se basant sur l'amplitude des réponses mesurées lors de la réalisation du plan d'expériences (Tableau IV. 13).

Tableau IV. 13 : Critères d'optimisation définis pour chacune des dix-huit réponses sélectionnées dans le cadre de la signature globale de *L. lactis ssp. lactis* ND5F en lait. Les valeurs des bornes supérieures, inférieures et cibles sont indiquées pour les réponses transformées, et entre parenthèses pour les valeurs non transformées, afin de faciliter l'interprétation. Les réponses en gris sont les réponses exclues du deuxième

Réponse	Action	Borne inférieure	Cible	Borne supérieure
Lactose _{T14}	Cibler	140	160	180
Lactate _{T14}	Cibler	60	85	100
Acétate _{T14}	Minimiser	-	1	3
V _{max} ^a	Cible	-10	-8,5	-6,5
T _{max} ^a	Minimiser	-	5	7

modèle d'optimisation avec un filtre plus strict sur la qualité des modèles ($R^2_{adj} > 0,50$ et $Q^2 > 0,25$).

$V_{\max}^{\text{red}}$	Cible	-12	-10	-8
D_{TF}	Minimiser	-	8,5	11,5
$\ln(B_{\text{T14}})$	Minimiser	-	13 ($4,4 \cdot 10^5$)	15 ($4,4 \cdot 10^6$)
$(\text{Glu}_{\text{T14}})^2$	Maximiser	$1 \cdot 10^6$ ($1 \cdot 10^3$)	$2 \cdot 10^6$ ($1,41 \cdot 10^3$)	-
$\frac{1}{\sqrt{\text{Glu}_{\text{T14}}}}$	Minimiser		0,15	0,35
Pentanedione	Maximiser	0,01	0,02	-
Méthanthiol	Maximiser	0,002	0,003	-
$\sqrt{\text{DMDS}}$	Minimiser	-	0,25 (0,0625)	0,35 (0,1225)
$\ln(\text{Furanméthanol})$	Minimiser	-	-1 (0,37)	0,50 (1,65)
$\log(\text{Méthylbutanal})$	Minimiser	-	-1,70 (0,020)	-2,00 (0,010)
$\frac{1}{\sqrt{\text{Diacétyle}}}$	Cibler	2 (0,25)	1,20 (0,913)	1 (1)
$\ln(\text{Acétaldéhyde})$	Cibler	-4,50 (0,011)	-3,50 (0,031)	-2,50 (0,082)
Acide hexanoïque	Minimiser	-	0,80	1

Cette démarche d'optimisation multi-réponses a été également menée sur les équations dont le R^2_{adj} et le Q^2 étaient supérieurs ou égaux à respectivement 0,50 et 0,25. Seulement dix des dix-huit réponses ont satisfait à ce nouveau critère (Tableau IV. 13). Par l'application de ce filtre plus strict sur

Matière sèche	Matière grasse	PO_{zinitale}	Température

la qualité des modèles, il est souhaité analyser l'impact des réponses moins fiables précédemment prises en compte sur la prédictibilité de l'optimisation, et vérifier si ces réponses ne concourent pas à définir de nouvelles coordonnées pour le point de fonctionnement des réponses communes.

Dans un second temps, une approche s'appuyant sur la transformation de chaque optimum local en fonction de la désirabilité [Harrington, 1980] a été utilisée pour modéliser simultanément les multiples réponses. Cette approche consiste à transformer chaque modèle en une fonction dans un domaine commun [0,1], de les combiner en réalisant une moyenne géométrique et d'optimiser la fonction finale résultante. La capacité de cette fonction à satisfaire les optimums prédéfinis sera évaluée par la comparaison entre la réponse prédite de cette fonction et les valeurs cibles souhaitées.

II.3.1 Optimisation sans contrainte d'assignation pour des facteurs

Dans un premier temps, le choix a été fait de ne mettre en œuvre aucune pondération afin que chaque variable ait le même poids dans l'optimisation de la signature de ND5F. Par rapport à la somme des dix-huit optimums considérés, le point de fonctionnement identifié est présenté (Tableau IV. 14).

Tableau IV. 14 : Point de fonctionnement identifié pour les quatre facteurs avec optimisation multiple des dix-huit réponses sélectionnées dans le cadre de la signature globale de *L. lactis* ssp. *lactis* ND5F en lait.

	(MS)	(MG)	(O ₂)	(TP)
Valeur codée	0,0565	-0,4520	0,4331	-0,3013
Valeur réelle	44 g.kg ⁻¹	11 g.kg ⁻¹	72 %	32 °C
Réponse	Action	Cible	Valeur prédite	Désirabilité
Les Lactose _{T14}	Cibler	160	165,51	0,86
Lactate _{T14}	Cibler	85	85,80	0,94
Acétate _{T14}	Minimiser	1	0,35	1

valeurs de trois des facteurs de l'optimum de fonctionnement trouvé sont cohérentes avec les valeurs sélectionnées empiriquement pour le procédé, en terme de température (32°C), matière sèche (44 g.kg⁻¹) ou encore la PO₂ initiale (le seuil de 70 % est proche d'un lait sans prétraitement, qui est d'environ 60 %). Ce résultat est une première indication de la pertinence de la simulation proposée. Par contre, le faible taux de matière grasse est plus surprenant au vu des formulations utilisées par Yoplait SA.

Pour juger de l'adéquation de ce point de fonctionnement à l'optimisation désirée, les réponses prédites par les modèles à ce point de fonctionnement vont être comparées aux gammes de réponses attendues (Tableau IV. 15). Pour faciliter l'analyse des résultats, le paramètre de désirabilité retranscrit pour chaque réponse la proximité de la réponse prédite à la gamme de réponses souhaitées. Seul l'objectif désiré de biomasse à 14 J semble ne pas pouvoir être satisfait, avec une désirabilité individuelle inférieure à 50 % (à 29 %).

$V_{\max}^a$	Maximiser	8,5	8,36	0,93
$T_{\max}^a$	Minimiser	5	4,57	1
$V_{\max}^{\text{red}}$	Maximiser	10	13,64	1
D_{TF}	Minimiser	8,5	9,14	0,78
$\ln(B_{\text{T14}})$	Minimiser	13	14,41	0,29
$(\text{Glu}_{\text{T14}})^2$	Maximiser	$2 \cdot 10^6$	$1,6 \cdot 10^6$	0,57
$\frac{1}{\sqrt{\text{Gly}_{\text{T14}}}}$	Minimiser	0,15	0,13	1
Pentanedione	Maximiser	0,02	0,0254	1
Méthanthiol	Maximiser	0,003	0,0071	1
$\sqrt{\text{DMDS}}$	Minimiser	0,25	0,014	1
$\ln(\text{Furanméthanol})$	Minimiser	-1	-3,22	1
$\log(\text{Méthylbutanal})$	Minimiser	-1,70	-1,67	0,95
$\frac{1}{\sqrt{\text{Diacytyle}}}$	Cibler	1,20	1,21	0,99
$\ln(\text{Acétaldéhyde})$	Cibler	-3,50	-3,92	0,58
Acide hexanoïque	Minimiser	0,80	0,69	1

Tableau IV. 15 : Récapitulatif des valeurs prédites pour chaque réponse à partir du point de fonctionnement optimal présumé. En gris, la seule réponse au seuil de désirabilité inférieur à 50 %.

Lamême démarche a été appliquée sur les 10 réponses à la qualité de modélisation supérieure. Dans un premier temps, il a été démontré, pour ce même point de fonctionnement proposé, que la désirabilité de chaque réponse individuelle était similaire à celle précédemment observée sur les 18 réponses. Dans un second temps, à partir de ces 10 réponses, un nouveau point de fonctionnement maximisant l'optimisation des réponses a été proposé, qui permet d'améliorer de 5 à 10 % la proximité de la réponse prédite à la valeur cible. Pour cela, le pourcentage de matière sèche devrait

être diminué à $41,7 \text{ g.kg}^{-1}$ et celui de la matière grasse à $10,6 \text{ g.kg}^{-1}$. Comme précédemment observé, la réponse biomasse à 14 J présente une désirabilité individuelle faible sur ce nouveau procédé proposé, preuve que ces deux points de fonctionnement ne correspondent pas à l'optimisation désirée pour cette réponse.

Dans toute démarche de modélisation, il est nécessaire, pour valider les modèles obtenus, de valider expérimentalement l'adéquation entre la prédiction (Tableau IV. 15) et les nouveaux résultats. Pour cela, cette validation va être réalisée sur le point de fonctionnement du procédé obtenu avec les 18 réponses pour disposer d'un maximum d'informations et intégrer les COVs à notre analyse. (voir II.3.3).

II.3.2 Optimisation avec un taux de matière grasse nulle du lait par défaut

En réponse à la demande des consommateurs et à l'émergence du marché des produits « light », les industriels doivent développer de nouveaux produits à teneur réduite en matière grasse, dits « allégés ». Être capable de garantir la typicité d'un produit tout en respectant ces demandes nutritionnelles représente un nouveau défi. L'utilisation d'équations pour modéliser chaque réponse intervenant dans la typicité de la souche permet de prévoir les effets d'un taux nul de matière grasse. Ainsi, la modification du pourcentage de matière grasse interagira sur un peu plus de la moitié des variables (13 sur 22 variables) étudiées. Les domaines concernés sont aussi variés que la cinétique d'acidification (V_{max}^a , D_{Tf}), la croissance cellulaire (biomasse finale et viabilité) ainsi que certains COVs, avec en priorité ceux issus de la lipolyse (acide hexanoïque, acétate d'éthyle) ou en interaction avec les voies de protéolyse/glycolyse (acétaldéhyde, furanméthanol, acétate). Il est à noter que l'annulation du pourcentage de matière grasse à partir des paramétrages de l'ancien procédé mène à une désirabilité globale nulle du fait des très faibles désirabilités locales des réponses liées à la matière grasse. Par conséquent, la même démarche d'optimisation conjointe des 18 réponses a été effectuée de nouveau en imposant une concentration nulle en matière grasse. Un niveau de désirabilité de la réponse globale de 80 % a pu être obtenu avec les paramètres suivants : température : 30°C , oxygène : 50% et matière sèche de 41 g.kg^{-1} .

Les paramètres obtenus sont, une fois de plus, tout à fait acceptables pour être utilisés dans un procédé. Toutefois, il est à noter que certaines des réponses liées au paramètre matière grasse présentent des taux de désirabilité de deux à trois fois plus faible dans ces conditions par rapport à la précédente optimisation (V_{max}^a , t_{max}^a ou encore la biomasse finale à 14 J). Cependant, il est envisageable de mettre en place une pondération afin de favoriser pour ces réponses l'obtention des niveaux attendus.

II.3.3 Adéquation des pistes quantitatives trouvées avec l'analyse qualitative organoleptique mise en œuvre.

Les 32 laits fermentés produits au cours de ce plan d'expérience ont été soumis à une dégustation à l'aveugle à 14 J par quatre juges experts qui ont évalué qualitativement les laits selon une liste de sept descripteurs sélectionnés : « crème », « acide », « faisselle », « sérum », « petit lait », « amertume » et « butyrate ». Sur ces 32 laits fermentés produits, seulement sept ont été qualifiés de « crème » avec des intensités diverses (Tableau IV. 16). Les différentes conditions ont été

Expérience	Intensité crème	Matière sèche	Matière grasse	Oxygène	Température
2	+	53	0	50	34
5	+++	43	20	0	28
10	+	53	20	50	28
16	++	43	40	100	34
20	+	53	20	100	34
21	+++	43	0	50	28
22	+	43	40	50	28

comparées afin d'identifier si des recoupements de conditions expérimentales étaient possibles, pouvant indiquer un lien direct entre cette typicité et un facteur du procédé. Cependant, il semble qu'aucune convergence autour d'un facteur ou d'une combinaison précise de facteurs n'ait pu être mise en évidence pour expliquer un tel comportement (trois taux de matière grasse et d'oxygénation, deux pourcentages de matière sèche et deux températures de fermentation sur ces sept conditions) (Tableau IV. 16). A noter cependant que les deux conditions à la typicité crème maximale sont celles qui partagent le même pourcentage de matière sèche à 43% et une température de 28°C.

Tableau IV. 16: Récapitulatif des conditions expérimentales des laits fermentés du plan appartenant à la classe dite « crème » après dégustation à 14 J. Echelle de mesure : 0 (absence) et intensité croissante (+, ++,+++).

Afin d'étudier la pertinence de cette classe organoleptique « crème », les signatures métaboliques de ces laits fermentés ont été étudiées par analyse PLS-DA sur les réponses précédemment utilisées pour modéliser la typicité de ND5F. Il est à noter que le partitionnement de l'échantillon en deux

groupes, un premier constitué uniquement par les sept échantillons « crème », et le second composé de l'intégralité des autres laits fermentés (soit 25 expériences), peut, du fait du déséquilibre en taille des deux groupes, présenter un taux maximal d'erreur lié au hasard de 34 %. C'est donc en comparaison à cette valeur de référence qu'il faut interpréter le taux d'erreur obtenu de 9,7 %, soit une erreur relative de 28 %. A l'image de ce pourcentage, les expériences dites « crème » se répartissent dans un même quart du graphique (corrélation négative avec les composantes 1 et 2) dans lequel aucune autre expérience n'est positionnée (Figure IV. 11). Cependant, il est constaté de fortes divergences le long de l'axe 2 entre les expériences fortement négatives (N°5) et celles nulles ou positives (N°2). Cette constatation est vérifiée par l'augmentation d'un facteur deux du taux d'erreur lors de la prise en compte de dimension supérieure. L'analyse des corrélations entre les variables et les composantes met en évidence que le positionnement négatif par rapport à la composante 1 est lié à la présence spécifique de pentane-2,3-dione et à l'absence de furanméthanol au sein des différents laits. Il est ainsi de nouveau confirmé que la présence de la pentane-2,3-dione est directement liée à la perception de la note crème. Ainsi, même si la qualité de modélisation de cette réponse par notre modèle est faible, les facteurs testés ont quand même eu une influence significative d'un point de vue du niveau de production. Cette étude permet également d'expliquer la modulation de l'intensité de la flaveur crème le long de la dimension deux. En effet, les essais les plus intenses corrélaient négativement avec la dimension deux, révélant de plus faibles concentrations en acétate, lactate, DMDS et méthylbutanal. Une fois de plus, la contribution d'un terrain organoleptique neutre au développement de la flaveur crème est mise en évidence.

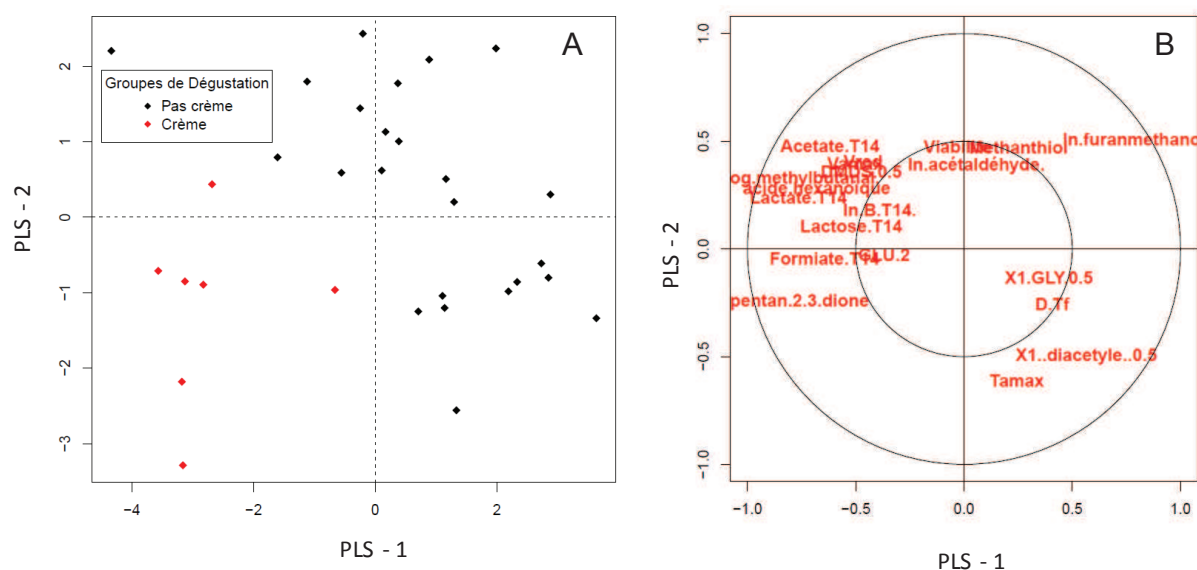


Figure IV. 11 : Analyse PLS-DA de la classification en deux groupes « crème » et « pas crème » à partir des 32 essais du plan d'expériences et des vingt réponses mesurées. A : Représentations des essais et B des réponses sur les composantes 1 et 2 de la PLS-DA. Le cercle interne sur la représentation des variables matérialise un coefficient de corrélation de 0,50 avec les composantes PLS.

II.4 Transfert et prédictibilité du point de fonctionnement du modèle à l'échelle pilote

Afin de vérifier la prédictibilité du modèle et la pertinence du point de fonctionnement déduit de l'optimisation multi-réponses menée sans contrainte, trois triplicats biologiques ont été réalisés avec *L. lactis ssp. lactis ND5F* dans les conditions précédemment définies (matière grasse, sèche et température) par Yoplait sur leur installation pilote (Tableau IV. 14). Il a été impossible de mettre en place un prétraitement initial du lait sur les installations pour fixer la PO_2 à 70 %. Cependant, cette valeur étant très proche de celle du lait sans traitement, il a été décidé de réaliser l'expérience sans ce paramètre. Pour cette validation, seules les réponses lactose, acides organiques (lactate, acétate), glutamate et biomasse à 14 J ont été mesurées, ainsi que le suivi de la V_{max}^a , du T_{max}^a et de D_{Tf} pendant la fermentation.

Tableau IV. 17 : Comparatif de la signature globale de *L. lactis* ssp. *lactis* ND5F en lait entre les valeurs prédites par le modèle et celles expérimentales mesurées pour le point de fonctionnement optimal défini.

Réponse	Cible souhaitée	Valeur prédite par le modèle	Désirabilité	Valeur mesurée expérimentalement
Lactose _{T14}	160	165,51	0,86	164 ± 7,2
Lactate _{T14}	85	85,80	0,94	84 ± 2
Acétate _{T14}	1	0,35	1	0,50 ± 0,051
V _{max} ^a	8,5	8,36	0,93	10,00*
T _{max} ^a	5	4,57	1	4,3*
D _{TF}	8,5	9,14	0,78	9,10 ± 0,11
Ln(B _{T14})	13	14,41	0,29	18,11 ± 0,04
(Glu _{T14}) ²	2.10 ⁶	1,6.10 ⁶	0,57	1,8.10 ⁶ ± 1,1.10 ⁵
$\frac{1}{\sqrt{Glu_{T14}}}$	0,15	0,13	1	0,13 ± 0,01

Expériences réalisées en triplicat. *Impossibilité de disposer pour ces mesures d'un écart-type.

Il est observé qu'à l'exception du dénombrement à 14 J, l'intégralité des autres réponses mesurées incluent dans leur distribution la valeur prédite par le modèle, et cela avec des écart-types de l'ordre de 10 %. Ainsi, il n'existe pas de différences significatives entre les valeurs mesurées et celles prédites. Sur la base de ce contrôle, il peut être déduit que les modèles proposés sont validés pour la gamme d'étude. Les divergences pour la réponse biomasse entre la prédiction et la valeur obtenue peuvent être attribuées à une qualité du modèle insuffisante (or le R^2_{adj} et le Q^2 étaient égaux à 0,83 et 0,63, ce qui est convenable), ou plus probablement, à des modifications liées au changement d'échelle. Il serait nécessaire de mettre en place un suivi pendant la fermentation et le stockage à 4°C de la biomasse pour comprendre l'origine de ces différences et construire un nouveau modèle.

Afin de compléter cette démarche de validation des modèles, deux autres points au sein de l'espace expérimental ont été réalisés. Ces deux essais ont été décidés afin de valider la consigne du facteur matière grasse dont la valeur était apparue surprenante sur le point de fonctionnement testé par rapport aux valeurs usuelles utilisées par Yoplait. Pour cela, par rapport à la précédente condition, la concentration en matière grasse a été modifiée (MG : 20 et 38,5 g.kg⁻¹), les autres facteurs du procédé restant inchangés. Les valeurs prédites ont été comparées à celles expérimentalement mesurées, et là encore, à l'exception de la biomasse viable à 14 J, les réponses obtenues sont similaires à celles qui avaient été prédites (résultats non présentés).

Cependant, il reste à déterminer pour ces différents essais le comportement des COVs et à évaluer la prédictibilité de ces modèles sur l'intégralité des réponses.

Pour juger de l'impact de ces deux nouveaux procédés sur la signature des produits de *ND5F*, une dégustation comparative de ces trois essais à 14 J a été organisée. Trois juges experts et 20 juges non experts en qualité de consommateurs potentiels ont participé à l'évaluation. En raison de différences notables de la texture des produits, seul le caractère aromatique a été comparé. Dans un premier temps, il a pu être identifié par les 3 juges experts que les 3 laits fermentés issus des trois conditions étaient bien différents. Sur les 20 consommateurs testés, 18 ont identifié la condition optimale comme étant différente des deux autres lors d'un tri-test. Lorsqu'il a été demandé aux personnes de décrire l'origine de ces différences, les 3 juges experts ainsi que 17 consommateurs ont utilisé les termes « crémeux », « crème », « gras ». Ce résultat, bien qu'effectué sur un panel réduit, démontre le succès de l'optimisation de la signature aromatique de *ND5F* dans un procédé industriel. Il témoigne également de la pertinence des méthodes mises en œuvre pour sélectionner ces déterminants et les utiliser dans un cadre applicatif.

CONCLUSION

L'analyse conjointe par bioinformatique et analyse structurale des séquences des gènes impliqués dans la voie de synthèse de la pentane-2,3-dione a permis de postuler sur la fonctionnalité des enzymes correspondantes. L'analyse protéomique a permis de confirmer certaines hypothèses et d'appréhender la complexité de ces régulations.

Nos résultats ont mis en évidence la possibilité d'influencer le titre final en pentane-2,3-dione à partir de modifications de paramètres nutritionnels et physico-chimiques. En particulier, il a pu être démontré l'important effet levier de la concentration en thréonine pour augmenter de plus d'un facteur 4 le titre final en métabolite d'intérêt. L'analyse protéomique a révélé une sous-expression de l'AHAIR pouvant également contribuer au phénotype observé. La diminution des BCAAs, bien qu'ayant entraîné une activation des voies de biosynthèse des BCAAs, comme observé au niveau de la sur-expression de BcaT, a contribué à diminuer le titre final. L'absence de détection d'un grand nombre de protéines des voies de biosynthèse des BCAAs ne permet pas de conclure quant à l'origine mécanistique précise de ce résultat. Cependant, l'hypothèse d'une redistribution du flux du pyruvate vers les voies de biosynthèse de la valine et de la leucine au détriment de celle de l'isoleucine peut être proposée. Ainsi, la concentration en BCAAs dans le milieu apparaît comme un facteur prépondérant dans le contrôle de la production de pentane-2,3-dione. L'aérobiose provoque d'importants changements au niveau du métabolisme central avec un shift vers un métabolisme mixte. Ainsi, près de 60 % du carbone du lactose est converti en acétoïne/diacétyl. Ces profondes modifications sont le résultat de la diminution de la quantité de la LDH, de la PFL et de l'augmentation de l'ALS en aval du pyruvate. Au niveau de la pentane-2,3-dione, le titre final a été augmenté de 28 % dans cette condition, ce qui est comparativement peu par rapport à l'augmentation de diacétyl/acétoïne. D'autres enzymes ont été impactées par ces conditions d'aérobiose, qui nécessiteront d'être étudiées en détail pour en comprendre les régulations.

L'étude globale de l'évolution de la signature de *L. lactis* ND5F a été analysée en milieu industriel. Une démarche de plans d'expériences a permis de formaliser, pour 20 réponses sélectionnées

comme discriminantes pour la typicité de *ND5F*, une équation modélisant l'impact de facteurs environnementaux. Quatre facteurs, principaux éléments du procédé de production de lait fermenté, ont été étudiés : la température, la matière grasse et sèche et l'oxygénation. Il a été observé que les modèles établis pour les réponses liées à des indicateurs physiologiques, technologiques et à des métabolites du métabolisme central, présentaient un seuil de variabilité expliquée de plus de 70 %. Ce taux atteste de la pertinence du contrôle de ces facteurs pour décrire les variations observées. Les réponses mesurées sur les COVs ont été modélisées avec une qualité inférieure. Plusieurs explications sont possibles, dont un caractère souche dépendant pour ces variables qui expliquerait les taux inférieurs de variance expliquée obtenus. Afin de mettre en place une analyse du comportement global de la souche et proposer un nouveau procédé renforçant la note « crème » des produits, un point de fonctionnement issu d'une optimisation multi réponses a été testé. Sur la majorité des réponses mesurées, la qualité prédictive des modèles précédemment établis a pu être confirmée. Les résultats de dégustation ont validé la possibilité d'influencer la signature organoleptique de *ND5F* à partir de modifications ciblées du procédé de production des laits fermentés.

CONCLUSION & PERSPECTIVES

Conclusion et discussion

La bactérie lactique *Lactococcus lactis* est, de manière traditionnelle, largement utilisée dans l'industrie laitière. Des études récentes ont démontré la grande biodiversité au sein de cette espèce. En effet, des souches d'origines diverses (laitières vs. environnementales) présentent des propriétés fonctionnelles [Ayad et al. 1999; Cachon et al. 2002; Fernández et al. 2011; Bayjanov et al. 2013], ainsi que des caractéristiques génétiques et génomiques, différentes [Kelly et al. 2010; Passerini et al. 2010]. Une telle diversité offre un large spectre de phénotypes envisageables. Le phénotype est, la plupart du temps, caractérisé de façon réductrice, relativement à un ou quelques paramètre(s) d'intérêt, comme la vitesse d'acidification, la production de certains arômes, d'exopolysaccharides, ou de bactériocines. Mais expliquer le phénotype global d'une souche nécessite d'aborder toute la complexité du métabolisme et de la physiologie cellulaire. Etablir les liens fonctionnels entre le génotype d'une souche et les caractéristiques globales de son phénotype reste encore aujourd'hui un vrai challenge. Il était donc crucial d'obtenir une meilleure compréhension des déterminants de cette diversité pour améliorer la sélection de souches aux caractéristiques ciblées.

Au cours de ce travail, nous avons entrepris l'étude de la singularité fonctionnelle de levains laitiers de *Lactococcus lactis*. Une démarche comparative a été menée afin de positionner notre souche d'intérêt (*L. lactis* ssp. *lactis* ND5F) face à la diversité fonctionnelle de la sous espèce *lactis*. Pour répondre à la problématique scientifique d'une réelle compréhension du lien entre génotype et phénotype, une approche transversale et intégrative multi-échelles a été conduite. Celle-ci repose sur l'intégration de différentes données biologiques obtenues à l'échelle du génotype, à partir d'analyses génétiques et génomiques, mais également au niveau du phénotype, avec des études métaboliques, physiologiques et protéomiques, extrêmement larges et complètes. Une telle approche fait appel à de nombreuses méthodes analytiques complexes, mais aussi à une analyse statistique dédiée rigoureuse. Cette démarche a d'abord été appliquée à l'étude de la diversité globale d'un panel de levains laitiers, avant d'être employée pour l'analyse de la spécificité de la souche ND5F. Afin d'ancrer l'analyse phénotypique au plus près des réalités applicatives, nous nous sommes placés dans les conditions du procédé modèle de fabrication de lait fermenté.

L'analyse de la diversité fonctionnelle nous a permis d'enrichir nos connaissances dans de multiples domaines, aussi bien importants sur le plan cognitif qu'utiles pour la recherche appliquée : évaluation intégrée du génotype des souches, caractérisation et organisation de la diversité fonctionnelle de la sous-population des levains, détermination de l'état protéique de cellules en

phase de post-acidification sur lait ou encore optimisation d'outils statistiques pour la détermination de la spécificité de souches. Cet apport cognitif original est résumé et discuté ci-dessous, en mettant l'accent sur : la classification fonctionnelle des souches établie en lait et le haut degré de spécificité observé pour ces souches génétiquement proches ; les différents marqueurs proposés pour le criblage de souches aux propriétés technologiques ciblées et le poids relatif des niveaux d'analyse pour l'interprétation des classifications intégratives établies.

▪ **Une analyse exhaustive de la diversité phénotypique des levains laitiers**

Afin d'analyser de façon pertinente la diversité phénotypique des levains de la sous-espèce *lactis*, une caractérisation génotypique a été menée afin de sélectionner et valider un panel de souches représentatif de la diversité génétique. En particulier, après vérification de la présence des gènes requis pour une croissance en lait, premier critère technologique, le positionnement phylogénétique par MLST a permis de choisir neuf souches de la classe « domestiquées » pour réaliser cette caractérisation phénotypique de la sous-espèce *lactis*. La diversité fonctionnelle de ces souches de *L. lactis* ssp. *lactis* a été caractérisée à partir d'un éventail de réponses métaboliques (concentration en acides organiques, acides aminés et composés organiques volatiles, COVs), physiologiques (croissance, survie) et technologiques (vitesse d'acidification et de réduction, durée de fermentation), afin d'obtenir une image complète de l'activité de ferment des souches. L'Analyse en Composantes Principales (ACP) de ces 82 marqueurs a mis en évidence la forte diversité phénotypique au sein de l'échantillon, puisque neuf signatures ont été observées, soit une signature par souche (ellipse à 95 % de confiance). Par conséquent, ce résultat confirme la pertinence de ce jeu de variables pour établir des signatures laitières des souches et prédire leur spécificité. De plus, ce haut degré de divergence entre les souches autorise une identification précise des souches en lait fermenté (bootstrap > 0,94). Ainsi, malgré une variabilité génétique limitée d'après les études de phylogénie, une importante diversité phénotypique est présente au sein des souches domestiquées, permettant d'envisager des applications spécifiques pour les différentes souches : i) production intense d'acide lactique pour assurer la biopréservation des produits, ii) synthèse du neurotransmetteur acide - γ -aminobutyrique pour un produit à la fonction santé, ou encore iii) biosynthèse de certains COVs à la saveur particulière.

▪ **Des marqueurs prédictifs pour la sélection ciblée de souches**

Dans le cadre de cette étude, il a pu être mis en évidence pour chaque niveau d'expression investigué, un groupe de marqueurs prédictifs de la diversité fonctionnelle des levains. Ainsi, au niveau génétique, une détection des gènes *lacE* et *prtP*, codant respectivement la lactose perméase

du complexe phosphotransférase (PTS) et la protéase de paroi, permet un criblage rapide garantissant l'aptitude des souches à développer un phénotype laitier (habilité à croître en lait en culture pure et à coaguler efficacement le lait en 24 h). De plus, l'analyse par MLST permet de distinguer rapidement les levains laitiers par leur appartenance au statut « domestiqué », et de préciser leurs relations phylogénétiques par rapport aux autres souches de cet écotype. Cependant, cette étude a démontré la faible fiabilité du gène *citP*, codant une citrate perméase, pour conditionner la proximité phénotypique des souches du biovar *diacetylactis*. En effet, comme déjà exploré au sein des isolats environnementaux [Passerini et al. 2010], les concentrations finales en acétoïne/diacétyl peuvent varier d'un facteur quatre pour des souches CitP⁺. De plus, de nombreux autres COVs présentent des comportements singuliers pour ces souches.

Pour réduire le temps et le coût de ces analyses phénotypiques fastidieuses, une sélection des marqueurs les plus appropriés pour discriminer les neuf souches a été menée (25 % des variables les plus contributives). Vingt variables ont été identifiées, toutes appartenant aux composés organiques volatiles (COVs). Ainsi, ces 20 COVs concentrent une variabilité inter-souches plus importante que les indicateurs technologiques, physiologiques ou encore d'autres métabolites. Les neuf souches laitières se structurent alors en quatre groupes robustes de souches aux orientations métaboliques et aromatiques divergentes : lipolyse, glycolyse, protéolyse et « non exprimée ». Trois groupes de COVs correspondent aux trois voies métaboliques majeures de la production d'arômes : glycolyse (4 variables), protéolyse (5 variables) et lipolyse (11 variables). Alors que les deux premières ont été étudiées en détails [Smit et al. 2005a; Smit et al. 2008; Cheng 2010], la lipolyse a généralement été sous-estimée par les travaux réalisés sur matrice en lait écrémé, or elle concentre la majorité des marqueurs discriminants des souches. Ce résultat confirme par ailleurs la nécessité de conduire des analyses au plus près de la matrice finale pour tenir compte de cette aptitude des souches particulièrement importante lors de la production des crèmes maturées ou de fromages. Combinée aux avancées de la GC-MS en lait [Smit et al. 2004; Bachmann et al. 2009b], cette liste de 20 COVs ouvre la voie au criblage haut débit et prédictif des potentialités aromatiques de collections de souches afin d'améliorer la connaissance et la sélection des ferments. Par ailleurs, l'analyse de ces variables pourra également renseigner indirectement sur le comportement global des souches du fait des corrélations existantes entre ces COVs et les indicateurs technologiques et physiologiques (durée de fermentation, vitesse d'acidification, production de lactate).

De nouveau, il est important de noter que les souches du biovar *diacetylactis* ne sont pas regroupées dans un même sous-groupe. Cette observation atteste que le niveau de production de diacétyl/acétoïne n'est pas pertinent pour discriminer la signature de ces souches, et qu'il existe une diversité métabolique intragroupe plus importante, et largement sous-estimée.

Pour la première fois, il a été mis en évidence sur neuf levains de *L. lactis* ssp. *lactis* une diversité protéique lors de la phase de post-acidification en lait industriel (matière protéique et matière grasse non nulles). Les enzymes de la glycolyse, de la nutrition azotée (peptidases, transporteurs d'oligopeptides), ou encore du métabolisme des nucléotides, ont été identifiées comme les protéines clefs de l'adaptation à la croissance en lait (protéines communes les plus abondantes). En parallèle, quarante-sept protéines ont pu être sélectionnées comme les principaux marqueurs de la spécificité d'expression, qui discriminent au mieux chacune des neuf souches. L'ajout de ND5F (2 variants) à cette analyse sélectionne 30 protéines discriminantes, identifiées comme robustes vis-à-vis de l'échantillon de souches. Cette sélection comprend, en particulier, un tiers de protéines du métabolisme énergétique, dont quatre enzymes en aval du pyruvate (ALS, ButA, PDH et Mae) et trois sous-unités de la citrate lyase (*CitCDF*). Le second groupe le plus abondant au sein de ces trente marqueurs discriminants concerne les protéines de la traduction (protéines ribosomales, ARNt synthases), des peptidases, et des enzymes impliquées dans la réponse aux stress (oxydatif, acide et osmotique). Cet ensemble de protéines structure la diversité protéique des levains en trois groupes (A, B et C). A noter qu'une nouvelle fois, l'intégralité des souches du biovar. *diacetylactis* ne sont pas réunies au sein d'un sous-groupe unique, confirmant la variabilité d'expression de ces souches. Les souches du sous-groupe C se singularisent notamment par des concentrations plus élevées en enzymes de réponses aux stress (RexA, SodA, GaD..). Ainsi, il se pourrait que ces souches puissent lutter plus efficacement contre la toxicité de l'oxygène, ou encore l'acidification du milieu. Ces observations devront être complétées par une analyse phénotypique ciblée pour valider les potentialités fonctionnelles de ces souches. Des recoupements entre marqueurs protéiques et signatures phénotypiques ont pu néanmoins être réalisés pour les deux souches *DIA-A* et *S87*.

Certaines de ces protéines peuvent également représenter des marqueurs potentiels pour la discrimination de souches de la sous-espèce *lactis*, lors de la réalisation d'empreintes protéiques en matrice solide complexe. En effet, grâce au développement des récentes techniques de protéomique par MALDI-TOF-MS (Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight), notamment en lait [Welker 2011; Šalplachta et al. 2012; Barreiro et al. 2012], il est possible d'envisager un criblage haut-débit d'expression protéique des souches. Cela permettrait de tester la robustesse des trente protéines identifiées, et d'étudier la diversité protéique à plus grande échelle pour découvrir des souches aux profils atypiques. Cependant, un travail conséquent reste à effectuer pour standardiser les conditions d'analyses et vérifier l'évolution de ces protéines en fonction des conditions environnementales et des différents états physiologiques.

▪ **Une classification intégrative du génotype au phénotype pour comprendre la diversité fonctionnelle des levains**

Bien qu'il ait pu être observé une singularité des souches sur chacun des niveaux d'expression étudiés, avec la possibilité de définir en particulier un ensemble de variables discriminantes, la classification des souches diffère selon le type de variables considérées. Ces différences de relations de proximité au sein des classifications confirment, comme précédemment démontré, le haut degré de régulation de l'expression du génome ainsi que du protéome [Bachmann et al. 2009b; Tan-a-ram et al. 2011; Kelly et al. 2010]. Pour proposer une classification unifiée reflétant la prise en compte de trois niveaux d'expression de la diversité des levains étudiés (génomique-protéique-phénotypique), l'impact de l'addition des différents jeux de données a été analysée. Ainsi, l'inclusion des variables génomiques à l'analyse phénotypique a permis de tenir compte des potentialités génétiques. Il a été observé que l'addition du génotype conduit à cinq sous-groupes au lieu de quatre, et permet donc une classification plus résolutive de la diversité des souches. En particulier, le précédent groupe de souches non dédiées à une voie métabolique particulière (sous-groupe IV de souches au caractère aromatique « non exprimé ») a été profondément modifié avec la prise en compte de ces données génomiques, puisque la seule souche du complexe clonal 2 (CC2) a été singularisée. Ce résultat met en lumière les divergences génétiques de cette souche, et permet de considérer les potentielles modifications évolutives de cette souche phylogénétiquement plus éloignée. Cependant, l'absence de variables génomiques sélectionnées parmi les variables les plus contributives sur ce nouveau set de données relativise l'importance du génotype par rapport aux larges différences phénotypiques observées sur les souches domestiquées.

Cette approche intégrée du génotype au phénotype a été également complétée en y intégrant le niveau intermédiaire de l'expression protéique, afin d'explorer les liens fonctionnels entre le génotype, le protéome et les caractéristiques phénotypiques d'une souche. Cependant, en fonction de l'objectif poursuivi, soit exploration des signatures des levains, soit identification et regroupements des souches en fonction de leur typicité aromatique, plusieurs classifications intégratives fonctionnelles peuvent-être envisagées. Ainsi, l'exploration de la singularité des souches sur les 1157 variables génomiques, phénotypiques et protéiques identifiées a permis d'apprécier, pour chaque levain, les potentialités exprimées (à l'échelle protéique ou phénotypique) en accord avec leur singularité génomique. Afin de faciliter cette analyse, il a pu être démontré la pertinence des marqueurs sélectionnés lors de l'analyse indépendante de chaque niveau de diversité (les 47

protéines, les 20 COVs et l'intégralité des distances génétiques) pour prédire la diversité du génotype au phénotype des 9 souches.

Par ailleurs, il a été également proposé, à l'échelle intégrative, une classification en adéquation avec la typicité aromatique des souches et non plus avec la singularité des souches. Pour cela, une sélection des variables génotypiques, protéiques et phénotypiques selon la supervision des 4 groupes phénotypiques précédemment observés, a permis d'isoler 81 variables majoritairement phénotypiques et protéiques comme prédictives de cette diversité fonctionnelle ciblée. Cette dernière classification intégrée constitue un outil d'aide à la sélection mais également à la compréhension de la typicité des levains.

▪ **Les déterminants de la spécificité organoleptique de la souche industrielle *L. lactis* ssp. *lactis* ND5F**

La définition de la singularité de notre souche d'intérêt nécessite une approche conjointe d'évaluation basée, à la fois, sur son adaptation au procédé laitier, et sur son positionnement au sein de la diversité précédemment caractérisée.

Le diagnostic multicritères (croissance, acidification, réduction, consommation d'oxygène) mené au sein du procédé modèle, a permis d'identifier les principaux leviers et verrous de la croissance de *L. lactis* ND5F. Ainsi, les conséquences de l'auto-acidification ont été traitées en identifiant les responsabilités des deux facteurs inhibiteurs, l'acide lactique non dissocié et la diminution du pH du milieu, sur les fonctions anaboliques et cataboliques. Néanmoins, les cibles cellulaires, pH intracellulaire, activités enzymatiques et concentrations de métabolites intracellulaires, resteraient à étudier pour préciser les aspects mécanistiques du mode d'action du stress acide. Un effet de la présence initiale d'oxygène sur l'activité acidifiante a été identifié, mais également, plus largement, sur le métabolisme, grâce à l'utilisation combinée de milieu semi-synthétique sous différentes atmosphères gazeuses. Les différents modes de préparation du lait, et notamment la thermisation lors de l'étape de stérilisation, sont apparus comme un levier pour augmenter la fermentescibilité de la matière première, en jouant sur la composition en acides aminés libres. L'analyse des exigences nutritionnelles de ND5F en lait, a permis d'identifier une limitation de la croissance par les acides aminés (en particulier les acides aminés branchés) que l'ajout exogène permet de rétablir, ouvrant la possibilité d'association avec des souches protéolytiques.

La caractérisation génotypique de ND5F a permis de classer et de positionner phylogénétiquement la souche, et d'attester de sa singularité génomique (empreintes PFGE). Ainsi, il est établi que ND5F appartient à la sous espèce *L. lactis* ssp. *lactis*, possède un typage allélique ST15 à partir du schéma MLST et fait partie des souches dites « domestiquées » de faible variabilité génétique. De plus, le

séquençage de la souche *ND5F* a identifié 2814 gènes (chromosomiques et plasmidiques) sur le génome de *ND5F*, dont seuls 48 sont spécifiques par comparaison aux génomes séquencés de l'espèce *L. lactis*. La majorité de ces gènes spécifiques code des activités liées aux prophages ou à des fonctions inconnues. En particulier, aucun gène impliqué dans un métabolisme spécifique n'a été identifié, empêchant toute prédiction ou orientation génétique de la spécificité aromatique de la souche.

Le positionnement phénotypique de *ND5F* au sein de l'échantillon de souches étudiées dans ce travail de thèse a révélé que les deux phénomènes majoritaires de la singularité de cette souche sont la faible activité lipolytique et glycolytique de la souche, et un comportement médian sur la majorité des autres variables mesurées. C'est dans ce contexte organoleptique neutre que la pentane-2,3-dione, molécule aux notes « crème », « beurre » et « gras » [Acree & Arn 2012] a été identifiée comme l'un des déterminants métaboliques de la typicité organoleptique de *ND5F*. En effet, il existe une parfaite adéquation entre les notes aromatiques de ce composé et le descripteur global utilisé pour identifier les laits fermentés par cette souche. Sur la base des connaissances phénotypiques de la singularité de *ND5F*, une liste de 14 marqueurs discriminants a pu être établie (4 COVs, 4 indicateurs physiologiques et technologiques, 5 sucres et métabolites en aval du pyruvate, ainsi qu'un acide aminé). L'identification de ces 14 déterminants et des relations fonctionnelles avec la signature de *ND5F* simplifie la problématique de caractérisation de la typicité de *ND5F*.

L'expression protéique a permis de mettre en exergue le haut niveau de singularité de la souche, puisqu'elle est isolée parmi un unique sous-groupe. Quatorze protéines spécifiques ont été identifiées après introduction de *ND5F* dans le modèle de diversité protéique construit. En complément, l'analyse différentielle a isolé 8 protéines significativement surexprimées sur le protéome de *ND5F* par rapport aux autres souches. Il a notamment été identifié le régulateur transcriptionnel LlrE de l'un des systèmes à deux composants, ou encore une protéine ribosomale. Cependant, la fonctionnalité de ces 8 protéines nécessite d'être étudiée plus en détails pour appréhender les conséquences physiologiques de cette surexpression sur le phénotype observé.

▪ **Les leviers pour contrôler la signature aromatique dans les produits laitiers**

Il a été recherché la possibilité d'influencer le titre final en pentane-2,3-dione à partir de modifications de paramètres nutritionnels et physico-chimiques du procédé. En particulier, il a pu être démontré l'important effet levier de la concentration en thréonine pour augmenter de plus de 4 fois le titre final en métabolite d'intérêt. A l'échelle protéique, il a été possible d'identifier dans cette condition une sous-expression de l'acétohydroxyacide isomérase (AHAIR) pouvant contribuer au phénotype observé. La diminution des acides aminés branchés, bien qu'ayant entraîné une activation des voies de biosynthèse des BCAAs, comme observé au niveau de la sur-expression de BcaT, a contribué à diminuer le titre final. Il est fait l'hypothèse d'une redistribution du flux de pyruvate vers les voies de biosynthèse de la valine et de la leucine au détriment de celle de l'isoleucine pour expliquer ce phénotype inattendu. Ainsi, la concentration en BCAAs dans le milieu apparaît comme un facteur prépondérant dans le contrôle de la production de pentane-2,3-dione et un facteur à prendre en compte lors de l'association à des souches aux activités protéolytiques connues pour libérer des BCAAs dans le milieu. En comparaison, l'aérobiose provoque d'importants changements au niveau du métabolisme central avec un shift vers un métabolisme mixte pour produire de l'acétoïne/diacétyle. Au niveau de la pentane-2,3-dione, le titre final a été seulement augmenté de 28 % dans cette condition, ce qui est comparativement peu par rapport à l'augmentation de diacétyle/acétoïne. Ce résultat semblerait indiquer que la limitation de la production est située en amont de la décarboxylation oxydative de l' α -acéto-hydroxybutanoate, au niveau de l'accumulation de ce précurseur de la pentane-2,3-dione.

L'étude globale de l'évolution de la signature de *L. lactis* ND5F a été analysée en milieu industriel. Une démarche de plans d'expériences a permis de formaliser, pour 20 réponses identifiées comme discriminantes pour la typicité de ND5F (14 marqueurs de typicité de ND5F plus 6 variables jugées importantes dans la production de lait fermenté), une équation modélisant l'impact de facteurs environnementaux sur cette réponse. Quatre facteurs, principaux éléments du procédé de production de lait fermenté, ont été étudiés : la température, les matières grasses et sèches, et l'oxygénation. Les modèles établis pour les réponses liées à des indicateurs physiologiques, technologiques et à des métabolites du métabolisme central, présentaient un coefficient de détermination R^2 de plus de 0,7, attestant de la pertinence du contrôle de ces facteurs pour décrire les variations observées. Les réponses mesurées sur les COVs ont été modélisées avec une qualité inférieure, peut-être à cause d'un caractère souche-dépendant de ces variables. Cependant, l'influence de ces différents facteurs sur la production de pentane-2,3-dione a été analysée. Une zone

de stabilité au sein d'un domaine a pu être modélisée, garantissant une stabilité et une robustesse de la réponse déterminée (Matière sèche : $43 \pm 2,5 \text{ g.kg}^{-1}$; Matière grasse : $20 \pm 5 \text{ g.kg}^{-1}$; Température : $28 \text{ }^{\circ}\text{C}$).

Un point de fonctionnement issu d'une optimisation multi réponses a été testé afin de proposer un nouveau procédé renforçant la note « crème » des produits. Ce procédé proposé a été expérimentalement validé, pour la majorité des paramètres, par la qualité prédictive des réponses mesurées par rapport aux cibles d'optimisation. De plus, les résultats de dégustation ont validé la possibilité d'influencer la signature organoleptique de ND5F à partir de modifications ciblées du procédé de production des laits fermentés.

PERSPECTIVES

- Elargir l'étude de la diversité à l'espèce *L. lactis*

Nous avons montré qu'une grande diversité phénotypique existe au sein des souches testées, et que chacune d'entre elles présente un profil singulier. La sélection de souches intéressantes nécessite donc d'étendre le nombre de souches analysées pour pouvoir répondre au mieux aux demandes des industriels et aux attentes des consommateurs. Ces nouvelles souches permettraient de tester la robustesse des quatre sous-groupes identifiés pour décrire la diversité phénotypique des levains de *L. lactis* ssp. *lactis*. En effet, les résultats des analyses exploratoires sont dépendants de la variabilité contenue dans le jeu de données, et peuvent donc fluctuer si certaines souches très singulières sont ajoutées à l'échantillon testé.

Devant l'importante diversité des souches environnementales, il serait intéressant d'évaluer la diversité fonctionnelle d'isolats au sein d'environnements naturels tels que les laits crus mais aussi les fourrages. En effet, ces réservoirs naturels pourraient abriter des souches originales aux propriétés intéressantes dans le cadre d'une utilisation en culture mixte avec les levains (production de bactériocines, d'exopolysaccharides, ou d'arômes). Un criblage en lait additionné de glucose et/ou de lysozyme pourrait permettre d'explorer la diversité phénotypique de cette sous-population très diverse sur le plan phylogénétique.

En plus d'élargir le nombre de souches analysées pour la sous-espèce *lactis*, il serait intéressant d'entreprendre la même étude pour la sous espèce *cremoris*, d'intérêt laitier également. Les outils que nous avons proposés pour l'étude de la diversité de *L. lactis* ssp. *Lactis* ont tous applicables à la sous-espèce *cremoris*. De plus, les analyses effectuées jusqu'à présent sur des souches de la sous-espèce *cremoris* ont montré que des souches identifiées génétiquement à la sous-espèce

cremoris par analyse de l'ARNr 16S (divergence sur la totalité des séquences codantes de 15% [Wegmann et al. 2007]) peuvent avoir un phénotype *lactis*, comme la souche MG1363 [Kelly & Ward 2002; Rademaker et al. 2007; Kütahya et al. 2011]. La distribution des phénotypes sur l'arbre phylogénétique de l'espèce permettra de mieux comprendre également les processus évolutifs de l'espèce et d'évaluer la diversité fonctionnelle de chacune de ces deux sous-espèces.

▪ **Explorer la fonctionnalité des protéines discriminantes**

La diversité protéique mise en évidence sur la trentaine de protéines discriminantes communes au panel de souches comprend, tout comme pour les protéines spécifiques de ND5F, de nombreuses protéines inconnues ou à la fonctionnalité seulement parcellaire. Ce manque d'information freine l'établissement de liens forts entre les protéines (présence, abondance) et le phénotype. En particulier, dans le cadre du régulateur transcriptionnel LlrE mis en évidence au sein des 8 protéines différentiellement surexprimées par ND5F, le régulon reste encore à l'heure actuelle inconnu. Du fait de l'importance de l'expression de ce gène dans la réponse cœur au stress de *L. lactis*, et notamment à la carence en isoleucine [Dressaire et al. 2008], il apparaît intéressant d'étudier plus en détail les cibles régulatrices de ce système à deux composants. Il pourrait-être ainsi étudié l'effet de la concentration de ce régulateur sur l'adaptation et la survie à différents stress : acidification, carence nutritionnelle, oxydatif, thermique ou encore osmotique, afin d'identifier d'éventuelles conséquences physiologiques pour ces modifications. Cette étude nécessiterait d'être couplée à une approche de transcriptomique comparative pour plus facilement visualiser les gènes cibles de ce régulateur, et éventuellement mettre en évidence des déterminants communs (séquences, biais de codon, famille fonctionnelle de gènes). Pour réaliser cette étude, un travail préliminaire de construction de souches génétiquement modifiées pour présenter le gène *llrE* sous la dépendance de promoteurs de différences forces et inductibles est à accomplir. De telles constructions ont été notamment initiées au sein de l'équipe sur la souche modèle *L. lactis* ssp. *lactis* IL 1403 ; il sera nécessaire de valider par dosage de l'ARNm en RTq-PCR ainsi que par dosage protéique de LlrE, l'efficacité d'expression de ce régulateur et les gammes de concentrations étudiées.

▪ **Enrichir l'approche intégrative de modélisation de la diversité pour prédire la fonctionnalité**

Les divergences d'organisation de la diversité mises en lumière au sein des différents niveaux d'observation du phénotype (indicateurs physiologiques, métabolites et protéines), mais également au niveau intégratif, proviennent probablement de la complexité du processus d'expression génétique. En effet, la mesure de l'expression génique est définie comme la quantification des

produits d'expression de l'ADN. Se pose alors la question du choix de l'élément qui représenterait le mieux les régulations de ces produits d'expression. Le produit direct mais intermédiaire du transfert de l'information du gène, l'ARNm, est souvent mesuré pour rendre compte de la modulation de l'expression d'un gène. Dans cette étude, il a été préféré la mesure d'indicateurs phénotypiques et des niveaux de protéines afin d'obtenir un reflet plus précis de la diversité fonctionnelle, en excluant ainsi du modèle les étapes de régulations (post-)transcriptionnelles et (post-)traductionnelles. Dans l'objectif de tendre vers une analyse mécanistique fine des différents processus cellulaires à l'origine de la singularité de chaque souche, il apparaît nécessaire d'enrichir le modèle construit. L'un des premiers axes de complémentarité serait de développer un dosage d'activités des enzymes discriminantes identifiées pour moduler par l'activité le paramètre concentration, uniquement pris en compte à ce jour dans les données protéiques. De plus, ces dosages d'activité pourraient être complétés par l'intégration des enzymes du métabolisme central et de la nutrition azotée afin de construire une image globale du comportement d'adaptation de la croissance des souches en lait. Une analyse de fluxomique pourrait être intéressante pour analyser si des flux métaboliques sont différents d'une souche à l'autre concernant les principales voies de production d'arômes. En particulier la voie de synthèse de la pentanedione pourrait-être étudiée pour tenter d'explicitier les liens entre les niveaux de protéines discriminantes observées et la production de ce métabolite.

A côté de ces nouveaux aspects expérimentaux, une réflexion sur la modélisation de l'intégration des jeux de données les uns par rapport aux autres peut être également conduite. En effet, dans cette analyse, chaque niveau d'étude a été considéré indépendamment. Il nous apparaît intéressant de travailler aux développements d'outils mathématiques afin d'inclure une strate supplémentaire de régulation hiérarchisée de l'expression génique omise jusqu'à présent. A l'heure actuelle, aucune méthode statistique ne permet de mettre en place cette analyse pour étudier la diversité hiérarchisée d'un sous-ensemble d'individus. Seule l'analyse canonique [Combes et al. 2008; Lê Cao et al. 2009] permet de comparer et interpréter des relations d'interdépendances multi-variables entre plusieurs individus, mais cette analyse est limitée à deux jeux de données sans application d'un filtre hiérarchique.

▪ **Adapter l'approche intégrative à la rationalisation de la sélection de cultures mixtes**

La complémentarité des approches d'exploration de la diversité des levains de *L. lactis* et d'identification de la typicité de notre souche d'intérêt, nous a permis de proposer un procédé de production permettant d'exploiter les caractéristiques organoleptiques singulières de notre souche. Cependant, l'utilisation industrielle de la souche en co-culture avec des ferments thermophiles

notamment, nous amène à nous interroger sur l'évolution de cette signature spécifique dans ce procédé industriel. Des essais d'application à l'échelle pilote en co-culture des modifications du procédé proposées dans ce travail ont été réalisés, démontrant un renforcement des notes crémeuses du produit par rapport à l'ancienne formulation. Cependant, une diminution de l'intensité « crème » est perceptible par comparaison à l'essai en culture pure. L'étude approfondie de *ND5F* dans cet écosystème laitier permettrait donc d'enrichir nos connaissances sur l'optimisation d'un procédé en culture mixte pour exploiter la singularité de la souche. La dimension interactive des échanges trophiques avec les autres souches, telle que la disponibilité en acides aminés et plus précisément en acides aminés branchés (pour lesquels *ND5F* est auxotrophe), serait à évaluer sur la croissance de la souche, mais également sur l'évolution du titre de la pentane-2,3dione. Une modélisation de cette nouvelle signature phénotypique permettrait également de proposer un mécanisme d'adaptation des souches dans un écosystème laitier. La souche *ND5F* caractérisée génétiquement et fonctionnellement dans ce travail de thèse, constitue un outil intéressant pour développer cette étude.

Il pourrait également être envisagé d'appliquer la même démarche de caractérisation fonctionnelle, à partir des marqueurs phénotypiques isolés, aux différents ferments utilisés en co-culture avec *ND5F*, afin de modéliser leur signature et évaluer l'impact de ces souches sur l'aromaticité finale du produit.

[THESE CONFIDENTIELLE]

REFERENCES BIBLIOGRAPHIQUES

A

- Abbe, K., Takahashi, S. & Yamada, T., 1982.** Involvement of oxygen-sensitive pyruvate formate-lyase in mixed-acid fermentation by *Streptococcus mutans* under strictly anaerobic conditions. *Journal of Bacteriology*, 152(1), p.175-182.
- Abee, T., Krockel, L. & Hill, C., 1995.** Bacteriocins: modes of action and potentials in food preservation and control of food poisoning. *International Journal of Food Microbiology*, 28(2), p.169-185.
- Abramsky, T., Rowland, L.P. & Shemin, D., 1965.** The formation of isoleucine from beta-methylaspartic acid in *Escherichia coli* W. *The Journal of biological chemistry*, 237, p.PC265-PC266.
- Acharjee, A., Kloosterman, B., de Vos, R.C.H., Werij, J.S., Bachem, C.W.B., Visser, R.G.F. & Maliepaard, C., 2011.** Data integration and network reconstruction with ~omics data using Random Forest regression in potato. *Analytica chimica acta*, 705(1-2), p.56-63.
- Acree & Arn, 2012.** Flavornet: Gas chromatography - olfactometry (GCO) of natural product. Available at: <http://www.flavornet.org/index.html> [Consulté le janvier 9, 2012].
- Ainsworth, S., Zomer, A., de Jager, V., Bottacini, F., van Hijum, S.A.F.T., Mahony, J. & van Sinderen, D., 2013.** Complete Genome of *Lactococcus lactis* subsp. *cremoris* UC509.9, Host for a Model *Lactococcal* P335 Bacteriophage. *Genome announcements*, 1(1).
- Allen, G.I. & Maletić-Savatić, M., 2011.** Sparse non-negative generalized PCA with applications to metabolomics. *Bioinformatics (Oxford, England)*, 27(21), p.3029-3035.
- Alonso, L. & Fraga, M.J., 2001.** Simple and rapid analysis for quantitation of the most important volatile flavor compounds in yogurt by headspace gas chromatography-mass spectrometry. *Journal of chromatographic science*, 39(7), p.297-300.
- Alting, A.C., Engels, W., van Schalkwijk, S. & Exterkate, F.A., 1995.** Purification and Characterization of Cystathionine (beta)-Lyase from *Lactococcus lactis* subsp. *cremoris* B78 and Its Possible Role in Flavor Development in Cheese. *Applied and environmental microbiology*, 61(11), p.4037-4042.
- Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schaffer, A.A. & Yu, Y.-K., 2005.** Protein Database Searches Using Compositionally Adjusted Substitution Matrices. *The FEBS journal*, 272(20), p.5101-5109.
- Aly, S., Flourey, J., Piot, M., Lortal, S. & Jeanson, S., 2012.** The efficacy of nisin can drastically vary when produced in situ in model cheeses. *Food Microbiology*, 32(1), p.185-190.
- Andersen, L., Kilstrup, M. & Neuhaard, J., 1989.** Pyrimidine, purine and nitrogen control of cytosine deaminase synthesis in *Escherichia coli* K12. Involvement of the *glnLG* and *purR* genes in the regulation of *codA* expression. *Archives of Microbiology*, 152(2), p.115-118.
- Anderson, T.W., 1963.** Asymptotic Theory for Principal Component Analysis. *The Annals of Mathematical Statistics*, 34(1), p.122-148.
- Andersson, R.E., Hedlund, C.B. & Jonsson, U., 1979.** Thermal Inactivation of a Heat-Resistant Lipase Produced by the Psychrotrophic Bacterium *Pseudomonas fluorescens*. *Journal of Dairy Science*, 62(3), p.361-367.
- Antoniadis, A., Lambert-Lacroix, S. & Leblanc, F., 2003.** Effective dimension reduction methods for tumor classification using gene expression data. *Bioinformatics*, 19(5), p.563-570.
- Arvanitoyannis, I., Katsota, M., Psarra, E., Soufleros, E. & Kallithraka, S., 1999.** Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). *Trends in Food Science & Technology*, 10(10), p.321-336.
- Ayad, E.H., Verheul, A., de Jong, C., Wouters, J.T. & Smit, G., 1999.** Flavour forming abilities and amino acid requirements of *Lactococcus lactis* strains isolated from artisanal and non-dairy origin. *International Dairy Journal*, 9(10), p.725-735.
- Ayad, E.H.E., Verheul, A., Engels, W.J.M., Wouters, J.T.M. & Smit, G., 2001.** Enhanced flavour formation by combination of selected lactococci from industrial and artisanal origin with

focus on completion of a metabolic pathway. *Journal of Applied Microbiology*, 90(1), p.59–67.

B

- Bachmann, H., Kruijswijk, Z., Molenaar, D., Kleerebezem, M. & van Hylckama Vlieg, J.E.T., 2009.** A high-throughput cheese manufacturing model for effective cheese starter culture screening. *Journal of Dairy Science*, 92(12), p.5868–5882.
- Bachmann, H., Starrenburg, M.J.C., Dijkstra, A., Molenaar, D., Kleerebezem, M., Rademaker, J.L.W. & van Hylckama Vlieg, J.E.T., 2009.** Regulatory Phenotyping Reveals Important Diversity within the Species *Lactococcus lactis*. *Applied and Environmental Microbiology*, 75(17), p.5687–5694.
- Bachmann, H., Starrenburg, M.J.C., Molenaar, D., Kleerebezem, M. & Van Hylckama Vlieg, J.E.T., 2012.** Microbial Domestication Signatures of *Lactococcus lactis* Can Be Reproduced by Experimental Evolution. *Genome Research*, 22(1), p.115–124.
- Baker, C.R., Booth, L.N., Sorrells, T.R. & Johnson, A.D., 2012.** Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell*, 151(1), p.80–95.
- Bantscheff, M., Lemeer, S., Savitski, M.M. & Kuster, B., 2012.** Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry*, 404(4), p.939–965.
- Baracchini, E. & Bremer, H., 1988.** Stringent and growth control of rRNA synthesis in *Escherichia coli* are both mediated by ppGpp. *The Journal of biological chemistry*, 263(6), p.2597–2602.
- Bardowski, J., Ehrlich, S.D. & Chopin, A., 1992.** Tryptophan biosynthesis genes in *Lactococcus lactis* subsp. *lactis*. *Journal of bacteriology*, 174(20), p.6563–6570.
- Barker, M.M., Gaal, T., Josaitis, C.A. & Gourse, R.L., 2001.** Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *Journal of molecular biology*, 305(4), p.673–688.
- Barrangou, R., Fremaux, C., Horvath, P., Romero, D. & Boyaval, P., 2012.** Methods to generate bacteriophage resistant bacterial strains and produce bacteriophage CRISPR loci phage mutants.
- Barreiro, J.R., Braga, P.A.C., Ferreira, C.R., Kostrzewa, M., Maier, T., Wegemann, B., Böettcher, V., Eberlin, M.N. & dos Santos, M.V., 2012.** Nonculture-based identification of bacteria in milk by protein fingerprinting. *PROTEOMICS*, 12(17), p.2739–2745.
- Bassit, N., Boquien, C.-Y., Picque, D. & Corrieu, G., 1993.** Effect of Initial Oxygen Concentration on Diacetyl and Acetoin Production by *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis*. *Applied and Environmental Microbiology*, 59(6), p.1893.
- Bayjanov, J.R., Molenaar, D., Tzeneva, V., Siezen, R.J. & Hijum, S.A.F.T. van, 2012.** PhenoLink - a web-tool for linking phenotype to omics data for bacteria: application to gene-trait matching for *Lactobacillus plantarum* strains. *BMC Genomics*, 13(1), p.1–12.
- Bayjanov, J.R., Starrenburg, M.J., van der Sijde, M.R., Siezen, R.J. & van Hijum, S.A., 2013.** Genotype-phenotype matching analysis of 38 *Lactococcus lactis* strains using random forest methods. *BMC microbiology*, 13(1), p.68.
- Béal, C. & Corrieu, G., 1994.** Viability and Acidification Activity of Pure and Mixed Starters of *Streptococcus salivarius* ssp. *thermophilus* 404 and *Lactobacillus delbrueckii* ssp. *bulgaricus* 398 at the Different Steps of Their Production. *LWT - Food Science and Technology*, 27(1), p.86–92.
- Beal, C., Skokanova, J., Latrille, E., Martin, N. & Corrieu, G., 1999.** Combined Effects of Culture Conditions and Storage Time on Acidification and Viscosity of Stirred Yogurt. *Journal of Dairy Science*, 82(4), p.673–681.
- Becker, T. & Puhán, Z., 1989.** Effect of different processes to increase the milk solids non-fat content on the rheological properties of yoghurt. *Milchwissenschaft*, 44(10), p.626–629.

- Belfares, L., Altaba, S., Pons, M.N., Engasser, J.M. & Fick, M., 1993.** Contribution of inorganic and organic ions to conductivity changes during lactic acid fermentation. *Bioprocess Engineering*, 9(5), p.197-204.
- Bell, S.C. & Turner, J.M., 1977.** Bacterial catabolism of threonine. Threonine degradation initiated by l-threonine hydrolyase (deaminating) in a species of *Corynebacterium*. *The Biochemical journal*, 164(3), p.579-587.
- Benjamini, Y. & Hochberg, Y., 1995.** Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), p.289-300.
- Benson, K.H., Godon, J.-J., Renault, P., Griffin, H.G. & Gasson, M.J., 1996.** Effect of *ilvBN*-encoded α -acetolactate synthase expression on diacetyl production in *Lactococcus lactis*. *Applied Microbiology and Biotechnology*, 45(1-2), p.107-111.
- Berg, R.A. van den, Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K. & Werf, M.J. van der, 2006.** Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7(1), p.1-15.
- Beyer, N.H., Roepstorff, P., Hammer, K. & Kilstrup, M., 2003.** Proteome analysis of the purine stimulon from *Lactococcus lactis*. *PROTEOMICS*, 3(5), p.786-797.
- Biliaderis, C.G., Khan, M.M. & Blank, G., 1992.** Rheological and sensory properties of yogurt from skim milk and ultrafiltered retentates. *International Dairy Journal*, 2(5), p.311-323.
- Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. & Shao, Y., 1997.** The complete genome sequence of *Escherichia coli* K-12. *Science (New York, N.Y.)*, 277(5331), p.1453-1462.
- Blow, N., 2009.** Transcriptomics: The digital generation. *Nature*, 458(7235), p.239-242.
- De Bok, F.A.M., Janssen, P.W.M., Bayjanov, J.R., Sieuwerts, S., Lommen, A., van Hylckama Vlieg, J.E.T. & Molenaar, D., 2011.** Volatile Compound Fingerprinting of Mixed-Culture Fermentations. *Applied and Environmental Microbiology*, 77(17), p.6233-6239.
- Bok, F.A.M. de, Janssen, P.W.M., Bayjanov, J.R., Sieuwerts, S., Lommen, A., Vlieg, J.E.T. van H. & Molenaar, D., 2011.** Volatile Compound Fingerprinting of Mixed-Culture Fermentations. *Applied and Environmental Microbiology*, 77(17), p.6233-6239.
- Bolón-Canedo, V., Sánchez-Marroño, N. & Cerviño-Rabuñal, J., 2013.** Scaling Up Feature Selection: A Distributed Filter Approach. In C. Bielza et al., éd. *Advances in Artificial Intelligence*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, p. 121-130.
- Bolotin, A., Mauger, S., Malarme, K., Ehrlich, S.D. & Sorokin, A., 1999.** Low-redundancy sequencing of the entire *Lactococcus lactis* IL1403 genome. *Antonie van Leeuwenhoek*, 76(1-4), p.27-76.
- Bolotin, A., Quinquis, B., Ehrlich, S.D. & Sorokin, A., 2012.** Complete genome sequence of *Lactococcus lactis* subsp. *cremoris* A76. *Journal of bacteriology*, 194(5), p.1241-1242.
- Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S.D. & Sorokin, A., 2001.** The Complete Genome Sequence of the Lactic Acid Bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Research*, 11(5), p.731-753.
- Boucher, B., Brothersen, C. & Broadbent, J.R., 2006.** Influence of starter and nonstarter lactic acid bacteria on medium redox. *Australian Journal of Dairy Technology*, 61(2), p.116-118.
- Boulesteix, A.-L., 2004.** PLS dimension reduction for classification with microarray data. *Statistical applications in genetics and molecular biology*, 3, p.Article33.
- Boulesteix, A.-L., Bender, A., Bermejo, J.L. & Strobl, C., 2012.** Random forest Gini importance favours SNPs with large minor allele frequency: impact, sources and recommendations. *Briefings in Bioinformatics*, 13(3), p.292-304.
- Le Bourgeois, P., Lautier, M., van den Berghe, L., Gasson, M.J. & Ritzenthaler, P., 1995.** Physical and genetic map of the *Lactococcus lactis* subsp. *cremoris* MG1363 chromosome: comparison with that of *Lactococcus lactis* subsp. *lactis* IL 1403 reveals a large genome inversion. *Journal of bacteriology*, 177(10), p.2840-2850.

- Boutrou, R., Sepulchre, A., Gripon, J.C. & Monnet, V., 1998.** Simple Tests for Predicting the Lytic Behavior and Proteolytic Activity of Lactococcal Strains in Cheese. *Journal of Dairy Science*, 81(9), p.2321-2328.
- Brandsma, J., Floris, E., Dijkstra, A.R., Rijnen, L., Wouters, J. & Meijer, W., 2008.** Natural diversity of aminotransferases and dehydrogenase activity in a large collection of *Lactococcus lactis* strains. *International Dairy Journal*, 18, p.1103-1108.
- Branger, A., Richer, M.-M. & Roustel, S., 2007a.** Critères de choix des ferments en rapport avec la filière. In *Microbiochimie et alimentation*. Educagri Editions, p. 184-200.
- Branger, A., Richer, M.-M. & Roustel, S., 2007b.** Utilisation du vivant. In *Microbiochimie et alimentation*. Educagri Editions, p. 167-183.
- Branny, P., De La Torre, F. & Garel, J.R., 1993.** Cloning, sequencing, and expression in *Escherichia coli* of the gene coding for phosphofructokinase in *Lactobacillus bulgaricus*. *Journal of Bacteriology*, 175(17), p.5344-5349.
- Branny, P., De La Torre, F. & Garel, J.R., 1996.** The genes for phosphofructokinase and pyruvate kinase of *Lactobacillus delbrueckii* subsp. *bulgaricus* constitute an operon. *Journal of Bacteriology*, 178(15), p.4727-4730.
- Le Bras, G. & Garel, J.R., 1991.** Properties of D-lactate dehydrogenase from *Lactobacillus bulgaricus*: a possible different evolutionary origin for the D- and L-lactate dehydrogenases. *FEMS microbiology letters*, 63(1), p.89-93.
- Le Bras, G. & Garel, J.R., 1993.** Pyruvate kinase from *Lactobacillus bulgaricus*: possible regulation by competition between strong and weak effectors. *Biochimie*, 75(9), p.797-802.
- Brasca, M., Morandi, S., Lodi, R. & Tamburini, A., 2007.** Redox potential to discriminate among species of lactic acid bacteria. *Journal of Applied Microbiology*, 103(5), p.1516-1524.
- Breiman, L., 1984.** *Classification and regression trees*, Wadsworth International Group.
- Breiman, L., 2001.** Random Forests. *Machine Learning*, 45(1), p.5-32.
- Breukink, E., van Kraaij, C., van Dalen, A., Demel, R.A., Siezen, R.J., de Kruijff, B. & Kuipers, O.P., 1998.** The orientation of nisin in membranes. *Biochemistry*, 37(22), p.8153-8162.
- Breukink, E., van Kraaij, C., Demel, R.A., Siezen, R.J., Kuipers, O.P. & de Kruijff, B., 1997.** The C-terminal region of nisin is responsible for the initial interaction of nisin with the target membrane. *Biochemistry*, 36(23), p.6968-6976.
- Brisabois, A., Lafargue, V., Brouillaud, A., de Buyser, M.-L., Collette, C., Garin-Bastuji, B. & Thorel, M.-F., 1997.** Les germes pathogènes dans le lait et les produits laitiers : situation en France et en Europe. *Rev. Sci. tech. Off. int. Epiz*, 16(1), p.452-471.
- Broadbent, J.R., Barnes, M., Brennand, C., Strickland, M., Houck, K., Johnson, M.E. & Steele, J.L., 2002.** Contribution of *Lactococcus lactis* Cell Envelope Proteinase Specificity to Peptide Accumulation and Bitterness in Reduced-Fat Cheddar Cheese. *Applied and Environmental Microbiology*, 68(4), p.1778-1785.
- Broadbent, J.R., McMahon, D.J., Welker, D.L., Oberg, C.J. & Moineau, S., 2003.** Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *Journal of dairy science*, 86(2), p.407-423.
- Broadbent, J.R., Rodríguez, B.T., Joseph, P., Smith, E.A. & Steele, J.L., 2006.** Conversion of *Lactococcus lactis* cell envelope proteinase specificity by partial allele exchange. *Journal of applied microbiology*, 100(6), p.1307-1317.
- Bruinenberg, P.G., De Roo, G. & Limsowtin, G., 1997.** Purification and Characterization of Cystathionine (gamma)-Lyase from *Lactococcus lactis* subsp. *cremoris* SK11: Possible Role in Flavor Compound Formation during Cheese Maturation. *Applied and Environmental Microbiology*, 63(2), p.561-566.
- Budin-Verneuil, A., Pichereau, V., Auffray, Y., Ehrlich, D. & Maguin, E., 2007.** Proteome phenotyping of acid stress-resistant mutants of *Lactococcus lactis* MG1363. *Proteomics*, 7(12), p.2038-2046.

- Buist, G., Kok, J., Leenhouts, K.J., Dabrowska, M., Venema, G. & Haandrikman, A.J., 1995.** Molecular cloning and nucleotide sequence of the gene encoding the major peptidoglycan hydrolase of *Lactococcus lactis*, a muramidase needed for cell separation. *Journal of bacteriology*, 177(6), p.1554-1563.
- Bunthof, C.J., van den Braak, S., Breeuwer, P., Rombouts, F.M. & Abee, T., 1999.** Rapid fluorescence assessment of the viability of stressed *Lactococcus lactis*. *Applied and environmental microbiology*, 65(8), p.3681-3689.

C

- Cachon, Jeanson, S., Aldarf, M. & Divies, C., 2002.** Characterisation of lactic starters based on acidification and reduction activities. *Le Lait*, 82(3), p.281-288.
- Cachon, R., Debleau, C., Feron, G., Ibarra, D. & Ledon, H., 2004.** Process For Modifying The Sensory Characteristics Of A Fermented Milk Product And Its Maturation During Conservation Of Said Product.
- Cai, Y., Yang, J., Pang, H. & Kitahara, M., 2011.** *Lactococcus fujiensis* sp. nov., a lactic acid bacterium isolated from vegetable matter. *International Journal of Systematic and Evolutionary Microbiology*, 61(7), p.1590-1594.
- Lê Cao, K.-A., Gonçalves, O., Besse, P. & Gadat, S., 2007.** Selection of biologically relevant genes with a wrapper stochastic algorithm. *Statistical applications in genetics and molecular biology*, 6, p.Article29.
- Lê Cao, K.-A., González, I. & Déjean, S., 2009a.** integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics*, 25(21), p.2855-2856.
- Lê Cao, K.-A., Martin, P.G., Robert-Granié, C. & Besse, P., 2009.** Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics*, 10(1), p.34.
- Casalta, E., Vassal, Y., Desmazeaud, M.J. & Casabianca, F., 1995.** Comparaison de l'activité acidifiante de souches de *Lactococcus lactis* isolées de lait et de fromage de corse. *LWT - Food Science and Technology*, 28(3), p.291-299.
- Cashel, M., 1969.** The Control of Ribonucleic Acid Synthesis in *Escherichia coli* IV. Relevance of unusual phosphorylated compounds from amino acid starved stringent strains. *Journal of Biological Chemistry*, 244(12), p.3133-3141.
- Cashel, M., Gentry, D., Hernandez, V. & Vinella, D., 1997.** The stringent response. In *Escherichia coli and Salmonella: cellular and molecular biology*. Washington, D.C.: American Society for Microbiology: Neidhardt F C, Curtiss III R, Ingraham J L, Lin E C C, Low K B, Magasanik B, Reznikoff W S, Riley M, Schaechter M, Umberger H E, editors; Neidhardt F C, Curtiss III R, Ingraham J L, Lin E C C, Low K B, Magasanik B, Reznikoff W S, Riley M, Schaechter M, Umberger H E,, p. 1456-1496.
- Van Casteren, W.H., Dijkema, C., Schols, H., Beldman, G. & Voragen, A.G., 1998.** Characterisation and modification of the exopolysaccharide produced by *Lactococcus lactis* subsp. *cremoris*B40. *Carbohydrate Polymers*, 37(2), p.123-130.
- Cerning, J., 1995.** Production of exopolysaccharides by lactic acid bacteria and dairy propionibacteria. *Le Lait*, 75(4-5), p.463-472.
- Cesselin, B., Ali, D., Gratadoux, J.-J., Gaudu, P., Duwat, P., Gruss, A. & Karoui, M.E., 2009.** Inactivation of the *Lactococcus lactis* high-affinity phosphate transporter confers oxygen and thiol resistance and alters metal homeostasis. *Microbiology*, 155(7), p.2274-2281.
- Chambellon, E. & Yvon, M., 2003.** CodY-Regulated Amino transferases AraT and BcaT Play a Major Role in the Growth of *Lactococcus lactis* in Milk by Regulating the Intracellular Pool of Amino Acids. *Applied and Environmental Microbiology*, 69(6), p.3061-3068.
- Chang, D.-E., Smalley, D.J. & Conway, T., 2002.** Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Molecular microbiology*, 45(2), p.289-306.

- Changeux, J.-P., 1961.** The Feedback Control Mechanism of Biosynthetic L-Threonine Deaminase by L-Isoleucine. *Cold Spring Harbor Symposia on Quantitative Biology*, 26, p.313-318.
- Chapot-Chartier, M.-P., Deniel, C., Rousseau, M., Vassal, L. & Gripon, J.-C., 1994.** Autolysis of two strains of *Lactococcus lactis* during cheese ripening. *International Dairy Journal*, 4(3), p.251-269.
- Chatterji, D., Fujita, N. & Ishihama, A., 1998.** The mediator for stringent control, ppGpp, binds to the beta-subunit of Escherichia coli RNA polymerase. *Genes to cells: devoted to molecular & cellular mechanisms*, 3(5), p.279-287.
- Chen, L., 2000.** *Thermophilic Enzymes and Their Impact on Milk Powder During Storage*. University of Waikato.
- Chen, L., Daniel, R.M. & Coolbear, T., 2003.** Detection and impact of protease and lipase activities in milk and milk powders. *International Dairy Journal*, 13(4), p.255-275.
- Chen, T., Cao, Y., Zhang, Y., Liu, J., Bao, Y., Wang, C., Jia, W. & Zhao, A., 2013.** Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. *Evidence-based complementary and alternative medicine: eCAM*, 2013, p.298183.
- Cheng, H., 2010.** Volatile flavor compounds in yogurt: a review. *Critical Reviews in Food Science and Nutrition*, 50(10), p.938-950.
- Cho, S.L., Nam, S.-W., Yoon, J.-H., Lee, J.S., Sukhoom, A. & Kim, W., 2008.** *Lactococcus chungangensis* sp. nov., a lactic acid bacterium isolated from activated sludge foam. *International journal of systematic and evolutionary microbiology*, 58(Pt 8), p.1844-1849.
- Chopin, A., 1993.** Organization and regulation of genes for amino acid biosynthesis in lactic acid bacteria. *FEMS Microbiology Reviews*, 12(1-3), p.21-37.
- Cocaign-Bousquet, M., Even, S., Lindley, N.D. & Loubière, P., 2002.** Anaerobic sugar catabolism in *Lactococcus lactis*: genetic regulation and enzyme control over pathway flux. *Applied microbiology and biotechnology*, 60(1-2), p.24-32.
- Cocaign-Bousquet, M., Garrigues, C., Novak, L., Lindley, N.D. & Loubiere, P., 1995.** Rational development of a simple synthetic medium for the sustained growth of *Lactococcus lactis*. *Journal of Applied Microbiology*, 79(1), p.108-116.
- Cogan, J. F., Walsh, D. & Condon, S., 1989.** Impact of aeration on the metabolic end-products formed from glucose and galactose by *Streptococcus lactis*. *Journal of Applied Microbiology*, 66(1), p.77-84.
- Collins, L.B. & Thomas, T.D., 1974.** Pyruvate kinase of *Streptococcus lactis*. *Journal of bacteriology*, 120(1), p.52-58.
- Collins, Y.F., McSweeney, P.L.H. & Wilkinson, M.G., 2003.** Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *International Dairy Journal*, 13(11), p.841-866.
- Collins, Y.F., McSweeney, P.L.H. & Wilkinson, M.G., 2003.** Evidence of a relationship between autolysis of starter bacteria and lipolysis in cheddar cheese during ripening. *The Journal of dairy research*, 70(1), p.105-113.
- Combes, S., González, I., Déjean, S., Baccini, A., Jehl, N., Juin, H., Cauquil, L., Gabinaud, B., Lebas, F. & Larzul, C., 2008.** Relationships between sensory and physicochemical measurements in meat of rabbit from three different breeding systems using canonical correlation analysis. *Meat Science*, 80(3), p.835-841.
- Condon, S., 1987.** Responses of lactic acid bacteria to oxygen. *FEMS Microbiology Letters*, 46(3), p.269-280.
- Cook, G.M. & Russell, D.J.B., 1994.** The effect of extracellular pH and lactic acid on pH homeostasis in *Lactococcus lactis* and *Streptococcus bovis*. *Current Microbiology*, 28(3), p.165-168.

- Corrieu, G., 2008.** Métabolisme et ingénierie métabolique. In *Bactéries lactiques: De la génétique aux ferments*. Tec & Doc, p. 271-509.
- Corrieu, G., Spinnler Eric, Picque, D. & Jomier, Y., 1989.** Method of revealing and monitoring the acidifying activity of fermentation agents in fermentation vats and device for implementing it.
- Costa, N.R., Lourenço, J. & Pereira, Z.L., 2011.** Desirability function approach: A review and performance evaluation in adverse conditions. *Chemometrics and Intelligent Laboratory Systems*, 107(2), p.234-244.
- Cretenet, M., Laroute, V., Ulvé, V., Jeanson, S., Nouaille, S., Even, S., Piot, M., Girbal, L., Le Loir, Y., Loubière, P., Lortal, S. & Coccagn-Bousquet, M., 2011.** Dynamic analysis of the *Lactococcus lactis* transcriptome in cheeses made from milk concentrated by ultrafiltration reveals multiple strategies of adaptation to stresses. *Applied and environmental microbiology*, 77(1), p.247-257.
- Crow, V.L., Coolbear, T., Gopal, P.K., Martley, F.G., McKay, L.L. & Riepe, H., 1995.** The role of autolysis of lactic acid bacteria in the ripening of cheese. *International Dairy Journal*, 5(8), p.855-875.
- Crow, V.L., Holland, R., Pritchard, G.G. & Coolbear, T., 1994.** The diversity of potential cheese ripening characteristics of lactic acid starter bacteria: 2. The levels and subcellular distributions of peptidase and esterase activities. *International Dairy Journal*, 4(8), p.723-742.
- Crow, V.L. Vaughan L., Martley, F.G., Coolbear, T. & Roundhill, S.J., 1995.** The influence of phage-assisted lysis of *Lactococcus lactis* subsp. *lactis* ML8 on cheddar cheese ripening. *International Dairy Journal*, 5(5), p.451-472.
- Cutler, F. original by L.B. and A. & Wiener, R. port by A.L. and M., 2012.** *randomForest: Breiman and Cutler's random forests for classification and regression*, Available at: <http://cran.r-project.org/web/packages/randomForest/index.html> [Consulté le août 15, 2013].
- Czerny, M., Christlbauer, Martin, Christlbauer, Monika, Fischer, A., Granvogel, M., Hammer, M., Hartl, C., Hernandez, N.M. & Schieberle, P., 2008.** Re-investigation on odour thresholds of key food aroma compounds and development of an aroma language based on odour qualities of defined aqueous odorant solutions. *European Food Research and Technology*, 228(2), p.265-273.

D

- Dabour, N. & LaPointe, G., 2005.** Identification and Molecular Characterization of the Chromosomal Exopolysaccharide Biosynthesis Gene Cluster from *Lactococcus lactis* subsp. *cremoris* SMQ-461. *Applied and Environmental Microbiology*, 71(11), p.7414-7425.
- Das, D. & Goyal, A., 2012.** Lactic Acid Bacteria in Food Industry. In T. Satyanarayana & B. N. Johri, éd. *Microorganisms in Sustainable Agriculture and Biotechnology*. Springer Netherlands, p. 757-772.
- Dash, M. & Liu, H., 1997.** Feature selection for classification. *Intelligent Data Analysis*, p.131-156.
- Deeth, H.C. & Touch, V., 2000.** Methods for detecting lipase activity in milk and milk products. *Australian Journal of Dairy Technology*, 55(3), p.153-168.
- Degeest, B., Vaniengelgem, F. & De Vuyst, L., 2001.** Microbial physiology, fermentation kinetics, and process engineering of heteropolysaccharide production by lactic acid bacteria. *International Dairy Journal*, 11, p.747-757.
- Degeest, B. & Vuyst, L.D., 1999.** Indication that the Nitrogen Source Influences Both Amount and Size of Exopolysaccharides Produced by *Streptococcus thermophilus* LY03 and Modelling of the Bacterial Growth and Exopolysaccharide Production in a Complex Medium. *Applied and Environmental Microbiology*, 65(7), p.2863.

- Dejean, S., Gonzalez, I., Monget, K.-A.L.C. with contributions from P., Coquery, J., Yao, F. & Liquet, B., 2013. *mixOmics: Omics Data Integration Project*, Available at: <http://cran.r-project.org/web/packages/mixOmics/index.html> [Consulté le août 15, 2013].
- Delgado, S. & Mayo, B., 2004. Phenotypic and genetic diversity of *Lactococcus lactis* and *Enterococcus* spp. strains isolated from Northern Spain starter-free farmhouse cheeses. *International journal of food microbiology*, 90(3), p.309-319.
- Delgado, S., Rachid, C.T.C.C., Fernández, E., Rychlik, T., Alegría, A., Peixoto, R.S. & Mayo, B., 2013. Diversity of thermophilic bacteria in raw, pasteurized and selectively-cultured milk, as assessed by culturing, PCR-DGGE and pyrosequencing. *Food microbiology*, 36(1), p.103-111.
- Delorme, C., Ehrlich, S.D. & Renault, P., 1992. Histidine biosynthesis genes in *Lactococcus lactis* subsp. *lactis*. *Journal of bacteriology*, 174(20), p.6571-6579.
- Delorme, C., Godon, J.J., Ehrlich, S.D. & Renault, P., 1993. Gene inactivation in *Lactococcus lactis*: histidine biosynthesis. *Journal of bacteriology*, 175(14), p.4391-4399.
- Dennis, P.P., Ehrenberg, M. & Bremer, H., 2004. Control of rRNA synthesis in *Escherichia coli*: a systems biology approach. *Microbiology and molecular biology reviews: MMBR*, 68(4), p.639-668.
- Derringer, G.S., 1980. Simultaneous Optimization of Several Response Variables. *Journal of Quality Technology*, 12, p.214-219.
- Derzelle, S., Hallet, B., Ferain, T., Delcour, J. & Hols, P., 2003. Improved adaptation to cold-shock, stationary-phase, and freezing stresses in *Lactobacillus plantarum* overproducing cold-shock proteins. *Applied and environmental microbiology*, 69(7), p.4285-4290.
- Desfossés-Foucault, E., Lapointe, G. & Roy, D., 2013. Dynamics and rRNA transcriptional activity of lactococci and lactobacilli during Cheddar cheese ripening. *International journal of food microbiology*, 166(1), p.117-124.
- Dhaisne, A., Guellerin, M., Laroute, V., Laguerre, S., Coccagn-Bousquet, M., Bourgeois, P.L. & Loubiere, P., 2013. Genotypic and Phenotypic Analysis of Dairy *Lactococcus lactis* Biodiversity in Milk: Volatile Organic Compounds as Discriminating Markers. *Applied and Environmental Microbiology*, 79(15), p.4643-4652.
- Diamant, S., Rosenthal, D., Azem, A., Eliahu, N., Ben-Zvi, A.P. & Goloubinoff, P., 2003. Dicarboxylic amino acids and glycine-betaine regulate chaperone-mediated protein-disaggregation under stress. *Molecular microbiology*, 49(2), p.401-410.
- Díaz-Uriarte, R. & Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7, p.3.
- Dietterich, T.G., 2000a. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, 40(2), p.139-157.
- Dietterich, T.G., 2000b. Ensemble Methods in Machine Learning. In *Multiple Classifier Systems*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, p. 1-15.
- Dodge, Y. & Rousson, V., 2004. *Analyse de régression appliquée*, Dunod.
- Doeven, Mark K, Kok, J. & Poolman, B., 2005. Specificity and selectivity determinants of peptide transport in *Lactococcus lactis* and other microorganisms. *Molecular microbiology*, 57(3), p.640-649.
- Donot, F., Fontana, A., Baccou, J.C. & Schorr-Galindo, S., 2012. Microbial exopolysaccharides: Main examples of synthesis, excretion, genetics and extraction. *Carbohydrate Polymers*, 87(2), p.951-962.
- Dressaire, C., Gitton, C., Loubière, P., Monnet, V., Queinnec, I. & Coccagn-Bousquet, M., 2009. Transcriptome and proteome exploration to model translation efficiency and protein stability in *Lactococcus lactis*. *PLoS computational biology*, 5(12), p.e1000606.
- Dressaire, C., Laurent, B., Loubière, P., Besse, P. & Coccagn-Bousquet, M., 2010. Linear covariance models to examine the determinants of protein levels in *Lactococcus lactis*. *Molecular bioSystems*, 6(7), p.1255-1264.

- Dressaire, C., Redon, E., Gitton, C., Loubière, P., Monnet, V. & Coccagn-Bousquet, M., 2011.** Investigation of the adaptation of *Lactococcus lactis* to isoleucine starvation integrating dynamic transcriptome and proteome information. *Microbial cell factories*, 10 Suppl 1, p.S18.
- Dressaire, C., Redon, E., Milhem, H., Besse, P., Loubière, P. & Coccagn-Bousquet, M., 2008.** Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses. *BMC Genomics*, 9(1), p.343.
- Drewe, O., Reil, G., Parlar, H. & Görg, A., 2004.** Setting up standards and a reference map for the alkaline proteome of the Gram-positive bacterium *Lactococcus lactis*. *PROTEOMICS*, 4(5), p.1293-1304.
- Durfee, T., Hansen, A.-M., Zhi, H., Blattner, F.R. & Jin, D.J., 2008.** Transcription Profiling of the Stringent Response in *Escherichia coli*. *Journal of Bacteriology*, 190(3), p.1084-1096.
- Duwat, P., Ehrlich, S.D. & Gruss, A., 1999.** Effects of metabolic flux on stress response pathways in *Lactococcus lactis*. *Molecular microbiology*, 31(3), p.845-858.

E

- Eckhardt, T.H., Skotnicka, D., Kok, J. & Kuipers, O.P., 2013.** Transcriptional regulation of fatty acid biosynthesis in *Lactococcus lactis*. *Journal of bacteriology*, 195(5), p.1081-1089.
- Ehrenberg, M., Bremer, H. & Dennis, P.P., 2013.** Medium-dependent control of the bacterial growth rate. *Biochimie*, 95(4), p.643-658.
- Ehrenberg, M., Dennis, P.P. & Bremer, H., 2010.** Maximum *rrn* promoter activity in *Escherichia coli* at saturating concentrations of free RNA polymerase. *Biochimie*, 92(1), p.12-20.
- Ellis, D.I., Dunn, W.B., Griffin, J.L., Allwood, J.W. & Goodacre, R., 2007.** Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics*, 8(9), p.1243-1266.
- Ellis, D.I. & Goodacre, R., 2006.** Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *The Analyst*, 131(8), p.875-885.
- Engels, W.J.M., Alting, A.C., Arntz, M.M.T.G., Gruppen, H., Voragen, A.G.J., Smit, G. & Visser, S., 2000.** Partial purification and characterization of two aminotransferases from *Lactococcus lactis* subsp. *cremoris* B78 involved in the catabolism of methionine and branched-chain amino acids. *International Dairy Journal*, 10(7), p.443-452.
- Escamilla-Hurtado, M.L., Tomasini-Campocoso, A., Valdés-Martínez, S. & Soriano-Santos, J., 1996.** Diacetyl formation by lactic bacteria. *Revista latinoamericana de microbiología*, 38(2), p.129-137.
- Even, S., Garrigues, C., Loubiere, P., Lindley, N.D. & Coccagn-Bousquet, M., 1999.** Pyruvate Metabolism in *Lactococcus lactis* Is Dependent upon Glyceraldehyde-3-phosphate Dehydrogenase Activity. *Metabolic Engineering*, 1(3), p.198-205.
- Eymann, C., Homuth, G., Scharf, C. & Hecker, M., 2002.** *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. *Journal of bacteriology*, 184(9), p.2500-2520.

F

- Felipe, F.L. de, Kleerebezem, M., Vos, W.M. de & Hugenholtz, J., 1998.** Cofactor Engineering: a Novel Approach to Metabolic Engineering in *Lactococcus lactis* by Controlled Expression of NADH Oxidase. *Journal of Bacteriology*, 180(15), p.3804.
- Felipe, F.L. de, Magni, C., Mendoza, D. de & López, P., 1995.** Citrate utilization gene cluster of the *Lactococcus lactis* biovar *diacetylactis*: organization and regulation of expression. *Molecular and General Genetics MGG*, 246(5), p.590-599.
- Felis, G.E., Molenaar, D., Dellaglio, F. & van Hylckama Vlieg, J.E.T., 2007.** Dichotomy in post-genomic microbiology. *Nature biotechnology*, 25(8), p.848-849.
- Fernández, Beerthuyzen, M.M., Brown, J., Siezen, R.J., Coolbear, T., Holland, R. & Kuipers, O.P., 2000.** Cloning, Characterization, Controlled Overexpression, and Inactivation of the Major

- Tributylin Esterase Gene of *Lactococcus lactis*. *Applied and Environmental Microbiology*, 66(4), p.1360-1368.
- Fernández de Palencia, P., De la Plaza, M., Amárita, F., Requena, T. & Peláez, C., 2006.** Diversity of amino acid converting enzymes in wild lactic acid bacteria. *Enzyme and Microbial Technology*, 38(1-2), p.88-93.
- Fernández, E., Alegría, A., Delgado, S., Martín, M.C. & Mayo, B., 2011.** Comparative phenotypic and molecular genetic profiling of wild *Lactococcus lactis* subsp. *lactis* strains of the *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* genotypes, isolated from starter-free cheeses made of raw milk. *Applied and environmental microbiology*, 77(15), p.5324-5335.
- Fernández, E., Delgado, S., Martín, M.C. & Mayo, B., 2011.** Comparative Phenotypic and Molecular Genetic Profiling of Wild *Lactococcus Lactis* Subsp. *Lactis* Strains of the *L. Lactis* Subsp. *Lactis* and *L. Lactis* Subsp. *Cremoris* Genotypes, Isolated from Starter-Free Cheeses Made of Raw Milk. *Applied and Environmental Microbiology*, 77(15), p.5324-5335.
- Fernández, L., Beerthuyzen, M.M., Brown, J., Siezen, R.J., Coolbear, T., Holland, R. & Kuipers, O.P., 2000.** Cloning, characterization, controlled overexpression, and inactivation of the major tributyrin esterase gene of *Lactococcus lactis*. *Applied and environmental microbiology*, 66(4), p.1360-1368.
- Fernández, M. & Zúñiga, M., 2006.** Amino acid catabolic pathways of lactic acid bacteria. *Critical reviews in microbiology*, 32(3), p.155-183.
- Fernández-García, E., Carbonell, M. & Nuñez, M., 2002.** Volatile fraction and sensory characteristics of Manchego cheese. 1. Comparison of raw and pasteurized milk cheese. *Journal of Dairy Research*, 69(04), p.579-593.
- Flambard, B., Helinck, S., Richard, J. & Juillard, V., 1998.** The Contribution of Caseins to the Amino Acid Supply for *Lactococcus lactis* Depends on the Type of Cell Envelope Proteinase. *Applied and Environmental Microbiology*, 64(6), p.1991-1996.
- Foda, E.A., Hammond, E.G., Reinbold, G.W. & Hotchkiss, D.K., 1974.** Role of Fat in Flavor of Cheddar Cheese. *Journal of Dairy Science*, 57(10), p.1137-1142.
- Foucaud, C., Kunji, E., Hagting, A., Richard, J., Konings, W., Desmazeaud, M. & Poolman, B., 1995.** Specificity of peptide transport systems in *Lactococcus lactis*: evidence for a third system which transports hydrophobic di- and tripeptides. *J. Bacteriol.*, 177(16), p.4652-4657.
- Fox, P.F., 2000.** *Fundamentals of Cheese Science*, Springer.
- Fox, P.F. & McSweeney, P.L.H., 2004.** Cheese: An Overview. In P. L. H. M. Patrick F. Fox, éd. *Cheese: Chemistry, Physics and Microbiology*. Academic Press, p. 1-18.
- Frees, D., Vogensen, F.K. & Ingmer, H., 2003.** Identification of proteins induced at low pH in *Lactococcus lactis*. *International journal of food microbiology*, 87(3), p.293-300.
- Friedrich, J.E. & Acree, T.E., 1998.** Gas chromatography olfactometry (GC/O) of dairy products. *International dairy journal*, 8(3), p.235-241.

G

- Galinier, A., Deutscher, J. & Martin-Verstraete, I., 1999.** Phosphorylation of either crh or HPr mediates binding of CcpA to the *Bacillus subtilis* xyn cre and catabolite repression of the xyn operon. *Journal of molecular biology*, 286(2), p.307-314.
- Gao, S. & Steele, J.L., 1998.** Purification and Characterization of Oligomeric Species of an Aromatic Amino Acid Aminotransferase from *Lactococcus Lactis* Subsp. *Lactis* S3. *Journal of Food Biochemistry*, 22(3), p.197-211.
- Gao, Y., Lu, Y., Teng, K.-L., Chen, M.-L., Zheng, H.-J., Zhu, Y.-Q. & Zhong, J., 2011.** Complete genome sequence of *Lactococcus lactis* subsp. *lactis* CV56, a probiotic strain isolated from the vaginas of healthy women. *Journal of bacteriology*, 193(11), p.2886-2887.
- Garault, P., Letort, C., Juillard, V. & Monnet, V., 2000.** Branched-chain amino acid biosynthesis is essential for optimal growth of *Streptococcus thermophilus* in milk. *Applied and environmental microbiology*, 66(12), p.5128-5133.

- Gardini, F., Lanciotti, R., Guerzoni, M.E. & Torriani, S., 1999.** Evaluation of aroma production and survival of *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactobacillus acidophilus* in fermented milks. *International dairy journal*, 9(2), p.125-134.
- Garrigues, C., Loubiere, P., Lindley, N.D. & Coccagn-Bousquet, M., 1997.** Control of the shift from homolactic acid to mixed-acid fermentation in *Lactococcus lactis*: predominant role of the NADH/NAD⁺ ratio. *Journal of Bacteriology*, 179(17), p.5282-5287.
- Gasson, M.J., 1996.** Lytic systems in lactic acid bacteria and their bacteriophages. *Antonie van Leeuwenhoek*, 70(2-4), p.147-159.
- Gastaldi, E., Lagaude, A., Marchesseau, S. & Tarodo De La Fuente, B., 1997.** Acid milk gel formation as affected by total solids content. *Journal of food science*, 62(4), p.671-675.
- Genovese, A., Gambuti, A., Piombino, P. & Moio, L., 2007.** Sensory properties and aroma compounds of sweet Fiano wine. *Food Chemistry*, 103(4), p.1228-1236.
- Gevaert, K., Van Damme, P., Ghesquière, B., Impens, F., Martens, L., Helsens, K. & Vandekerckhove, J., 2007.** A la carte proteomics with an emphasis on gel-free techniques. *Proteomics*, 7(16), p.2698-2718.
- Gitton, C., Meyrand, M., Wang, J., Caron, C., Trubuil, A., Guillot, A. & Mistou, M.-Y., 2005.** Proteomic Signature of *Lactococcus lactis* NCDO763 Cultivated in Milk. *Applied and Environmental Microbiology*, 71(11), p.7152-7163.
- Godon, J J, Chopin, M.C. & Ehrlich, S.D., 1992.** Branched-chain amino acid biosynthesis genes in *Lactococcus lactis* subsp. *lactis*. *Journal of bacteriology*, 174(20), p.6580-6589.
- Godon, J J, Delorme, C., Bardowski, J., Chopin, M.C., Ehrlich, S.D. & Renault, P., 1993.** Gene inactivation in *Lactococcus lactis*: branched-chain amino acid biosynthesis. *Journal of bacteriology*, 175(14), p.4383-4390.
- Goupil, N., Corthier, G., Ehrlich, S.D. & Renault, P., 1996.** Imbalance of leucine flux in *Lactococcus lactis* and its use for the isolation of diacetyl-overproducing strains. *Applied and Environmental Microbiology*, 62(7), p.2636.
- Goupil-Feuillerat, N., Coccagn-Bousquet, M., Godon, J.J., Ehrlich, S.D. & Renault, P., 1997.** Dual role of alpha-acetolactate decarboxylase in *Lactococcus lactis* subsp. *lactis*. *Journal of bacteriology*, 179(20), p.6285-6293.
- Goupil-Feuillerat, N., Corthier, G., Godon, J.J., Ehrlich, S.D. & Renault, P., 2000.** Transcriptional and translational regulation of alpha-acetolactate decarboxylase of *Lactococcus lactis* subsp. *lactis*. *Journal of bacteriology*, 182(19), p.5399-5408.
- Goupy, S., Rochut, N., Robins, R.J. & Gentil, E., 2000.** Evaluation of solid-phase microextraction for the isotopic analysis of volatile compounds produced during fermentation by lactic acid bacteria. *Journal of agricultural and food chemistry*, 48(6), p.2222-2227.
- Goupy, J., 2001.** *Plans d'expériences : les mélanges* - Dunod/Industries et technologies., France: Dunod.
- Govindasamy-Lucey, S., Gopal, P.K., Sullivan, P.A. & Pillidge, C.J., 2000.** Varying influence of the autolysin, N-acetyl muramidase, and the cell envelope proteinase on the rate of autolysis of six commercial *Lactococcus lactis* cheese starter bacteria grown in milk. *Journal of Dairy Research*, 67(04), p.585-596.
- Grosch, W., 1993.** Detection of potent odorants in foods by aroma extract dilution analysis. *Trends in Food Science & Technology*, 4(3), p.68-73.
- Guédon, E., Jamet, E. & Renault, P., 2002.** Gene regulation in *Lactococcus lactis*: the gap between predicted and characterized regulators. *Antonie van Leeuwenhoek*, 82(1-4), p.93-112.
- Guédon, E., Martin, C., Gobert, F.-X., Ehrlich, S.D., Renault, P. & Delorme, C., 2001.** Réseau de régulation de la transcription des gènes du système protéolytique de *Lactococcus lactis*. *Le Lait*, 81(1-2), p.65-74.
- Guédon, E., Renault, P., Ehrlich, S.D. & Delorme, C., 2001.** Transcriptional pattern of genes coding for the proteolytic system of *Lactococcus lactis* and evidence for coordinated regulation of key enzymes by peptide supply. *Journal of bacteriology*, 183(12), p.3614-3622.

- Guédon, E., Sperandio, B., Pons, N., Ehrlich, S.D. & Renault, P., 2005.** Overall control of nitrogen metabolism in *Lactococcus lactis* by CodY, and possible models for CodY regulation in Firmicutes. *Microbiology*, 151(12), p.3895-3909.
- Guillot, A., Gitton, C., Anglade, P. & Mistou, M.-Y., 2003.** Proteomic analysis of *Lactococcus lactis*, a lactic acid bacterium. *Proteomics*, 3(3), p.337-354.
- Güler, Z., 2007.** Changes in salted yoghurt during storage. *International Journal of Food Science & Technology*, 42(2), p.235-245.
- Gunstone, F.D., Harwood, J.L. & Dijkstra, A.J., 2010.** *The Lipid Handbook with CD-ROM, Third Edition*, Taylor & Francis.
- Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. & Aebersold, R., 2000.** Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences of the United States of America*, 97(17), p.9390-9395.
- Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. & Aebersold, R., 1999.** Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology*, 17(10), p.994-999.

H

- Habib, F., Johnson, A.D., Bundschuh, R. & Janies, D., 2007.** Large scale genotype-phenotype correlation analysis based on phylogenetic trees. *Bioinformatics*, 23(7), p.785-788.
- Hagting, A., Kunji, E.R., Leenhouts, K.J., Poolman, B. & Konings, W.N., 1994.** The di- and tripeptide transport protein of *Lactococcus lactis*. A new type of bacterial peptide transporter. *The Journal of biological chemistry*, 269(15), p.11391-11399.
- Hand, D.J., 2012.** Assessing the Performance of Classification Methods. *International Statistical Review*, 80(3), p.400-414.
- Hansen, E.B., 2002.** Commercial bacterial starter cultures for fermented foods of the future. *International journal of food microbiology*, 78(1-2), p.119-131.
- Harington, J., 1965.** The Desirability Function. *Industrial Quality Control*, 21, p.494-498.
- Harms, N., Ras, J., Reijnders, W.N., Spanning, R.J. van & Stouthamer, A.H., 1996.** S-formylglutathione hydrolase of *Paracoccus denitrificans* is homologous to human esterase D: a universal pathway for formaldehyde detoxification? *Journal of Bacteriology*, 178(21), p.6296-6299.
- Harper, W.J., Carmona de Catril, A. & Chen, J.L., 1980.** Esterases of lactic streptococci and their stability in cheese slurry systems. *Milchwissenschaft*, 35(3), p.129-132.
- Heinemeyer, E.A. & Richter, D., 1978.** Characterization of the guanosine 5'-triphosphate 3'-diphosphate and guanosine 5'-diphosphate 3'-diphosphate degradation reaction catalyzed by a specific pyrophosphorylase from *Escherichia coli*. *Biochemistry*, 17(25), p.5368-5372.
- Helinck, S., Le Bars, D., Moreau, D. & Yvon, M., 2004.** Ability of Thermophilic Lactic Acid Bacteria To Produce Aroma Compounds from Amino Acids. *Applied and Environmental Microbiology*, 70(7), p.3855-3861.
- Den Hengst, C.D., van Hijum, S.A.F.T., Geurts, J.M.W., Nauta, A., Kok, J. & Kuipers, O.P., 2005.** The *Lactococcus lactis* CodY regulon: identification of a conserved cis-regulatory element. *The Journal of biological chemistry*, 280(40), p.34332-34342.
- Hernandez, V.J. & Bremer, H., 1991.** *Escherichia coli* ppGpp synthetase II activity requires spoT. *The Journal of biological chemistry*, 266(9), p.5991-5999.
- Herrero, M., Simó, C., García-Cañas, V., Ibáñez, E. & Cifuentes, A., 2012.** Foodomics: MS-based strategies in modern food science and nutrition. *Mass Spectrometry Reviews*, 31(1), p.49-69.
- Higuchi, T., Hayashi, H. & Abe, K., 1997.** Exchange of glutamate and gamma-aminobutyrate in a *Lactobacillus* strain. *Journal of Bacteriology*, 179(10), p.3362-3364.
- Van Hijum, S.A.F.T., Medema, M.H. & Kuipers, O.P., 2009.** Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiology and molecular biology reviews: MMBR*, 73(3), p.481-509, Table of Contents.

- Holland, R. & Coolbear, T., 1996.** Purification of tributyrin esterase from *Lactococcus lactis* subsp. *cremoris* E8. *Journal of Dairy Research*, 63(01), p.131-140.
- Holland, R., Liu, S.-Q., Crow, V.L., Delabre, M.-L., Lubbers, M., Bennett, M. & Norris, G., 2005.** Esterases of lactic acid bacteria and cheese flavour: Milk fat hydrolysis, alcoholysis and esterification. *International Dairy Journal*, 15(6-9), p.711-718.
- Howell, K.S., Cozzolino, D., Bartowsky, E.J., Fleet, G.H. & Henschke, P.A., 2006.** Metabolic profiling as a tool for revealing *Saccharomyces* interactions during wine fermentation. *FEMS Yeast Research*, 6(1), p.91-101.
- Hughenoltz, J., 1993.** Citrate metabolism in lactic acid bacteria. *FEMS Microbiology Reviews*, 12(1-3), p.165-178.
- Hughenoltz, J., Perdon, L. & Abee, T., 1993.** Growth and Energy Generation by *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* during Citrate Metabolism. *Applied and Environmental Microbiology*, 59(12), p.4216-4222.
- Hui, Y.H., Meunier-Goddik, L., Josephsen, J., Nip, W.-K. & Stanfield, P.S., 2003.** *Handbook of Food and Beverage Fermentation Technology*, Taylor & Francis.
- Hui, Y.H., Nollet, L.M.L., Guerrero-Legarreta, I., Culbertson, J., Duncan, S., Li-Chan, E.C.Y., Ma, C.-Y., Manley, C.H., Marshall, D., McMeekin, T.A., Rahman, M.S., Toldra, F. & Xiong, Y.L., 2005.** *Handbook of Food Science, Technology, and Engineering, Volume Four*, Taylor & Francis.
- Husson, F., Josse, J., Le, S. & Mazet, J., 2013.** *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R*, Available at: <http://cran.r-project.org/web/packages/FactoMineR/index.html> [Consulté le août 15, 2013].
- Van Hylckama Vlieg, J.E.T., Rademaker, J.L.W., Bachmann, H., Molenaar, D., Kelly, W.J. & Siezen, R.J., 2006.** Natural diversity and adaptive responses of *Lactococcus lactis*. *Current opinion in biotechnology*, 17(2), p.183-190.

I & J

- Imhof, R. & Bosset, J.O., 1994.** Relationships between micro-organisms and formation of aroma compounds in fermented dairy products. *Zeitschrift für Lebensmittel-Untersuchung und Forschung*, 198(4), p.267-276.
- Imhof, R., Glättli, H. & Bosset, J.O., 1995.** Volatile organic compounds produced by thermophilic and mesophilic single strain dairy starter cultures. *LWT - Food Science and Technology*, 28(1), p.78-86.
- Inagaki, N., Iguchi, A., Yokoyama, T., Yokoi, K., Ono, Y., Yamakawa, A., Taketo, A. & Kodaira, K.-I., 2009.** Molecular properties of the glucosaminidase AcmA from *Lactococcus lactis* MG1363: mutational and biochemical analyses. *Gene*, 447(2), p.61-71.
- Ishihama, A., 2000.** Functional modulation of *Escherichia coli* RNA polymerase. *Annual review of microbiology*, 54, p.499-518.
- Jacob, H., 1970.** Redox potential. In *Methods in microbiology*, vol 2. London, England: Academic Press, p. 91-123.
- Jaeger, K.E., Ransac, S., Dijkstra, B.W., Colson, C., van Heuvel, M. & Misset, O., 1994.** Bacterial lipases. *FEMS microbiology reviews*, 15(1), p.29-63.
- Jeanson, S., Hilgert, N., Coquillard, M.-O., Seukpanya, C., Faiveley, M., Neveu, P., Abraham, C., Georgescu, V., Fourcassié, P. & Beuvier, E., 2009.** Milk acidification by *Lactococcus lactis* is improved by decreasing the level of dissolved oxygen rather than decreasing redox potential in the milk prior to inoculation. *International Journal of Food Microbiology*, 131(1), p.75-81.
- Jendresen, C.B., Martinussen, J. & Kilstrup, M., 2012.** The PurR regulon in *Lactococcus lactis* - transcriptional regulation of the purine nucleotide metabolism and translational machinery. *Microbiology*, 158(Pt_8), p.2026-2038.
- Jensen, P.R., 1993.** Minimal Requirements for Exponential Growth of *Lactococcus lactis*. *Appl. Environ. Microbiol.*, 59(12), p.4363-4366.

- Jensen, R.G., 1983.** Detection and determination of lipase (acylglycerol hydrolase) activity from various sources. *Lipids*, 18(9), p.650-657.
- Jensen, R.G., DeJong, F.A. & Clark, R.M., 1983.** Determination of lipase specificity. *Lipids*, 18(3), p.239-252.
- Jensen, R.G., Galluzzo, D.R. & Bush, V.J., 1990.** Selectivity is an Important Characteristic of Lipases (Acylglycerol Hydrolases). *Biocatalysis and Biotransformation*, 3(4), p.307-316.
- Jiang, M., Wang, C., Zhang, Y., Feng, Y., Wang, Y. & Zhu, Y., 2013.** Sparse Partial-least-squares Discriminant Analysis for Different Geographical Origins of *Salvia miltiorrhiza* by (1) H-NMR-based Metabolomics. *Phytochemical analysis: PCA*.
- Jin, D.J., Cagliero, C. & Zhou, Y.N., 2012.** Growth rate regulation in *Escherichia coli*. *FEMS Microbiology Reviews*, 36(2), p.269-287.
- Jishage, M., Kvint, K., Shingler, V. & Nystrom, T., 2002.** Regulation of σ factor competition by the alarmone ppGpp. *Genes & Development*, 16(10), p.1260-1270.
- John, G.T., Goelling, D., Klimant, I., Schneider, H. & Heinzle, E., 2003.** pH-Sensing 96-well microtitre plates for the characterization of acid production by dairy starter cultures. *Journal of Dairy Research*, 70(03), p.327-333.
- Jolliffe, I.T., 2002.** *Principal component analysis*, Available at: http://genes.bibli.fr/opac/index.php?lvl=notice_display&id=68622.
- Jolly, L., Vincent, S.J.F., Duboc, P. & Neeser, J.-R., 2002.** Exploiting exopolysaccharides from lactic acid bacteria. *Antonie van Leeuwenhoek*, 82(1-4), p.367-374.
- De Jong, A., Hansen, M.E., Kuipers, O.P., Kilstrup, M. & Kok, J., 2013.** The Transcriptional and Gene Regulatory Network of *Lactococcus lactis* MG1363 during Growth in Milk. *PLoS ONE*, 8(1), p.e53085.
- Jordan, K.N. & Cogan, T.M., 1988.** Production of acetolactate by *Streptococcus diacetylactis* and *Leuconostoc* spp. *Journal of Dairy Research*, 55(02), p.227-238.
- Juillard, V., Le Bars, D., Kunji, E.R., Konings, W.N., Gripon, J.C. & Richard, J., 1995.** Oligopeptides are the main source of nitrogen for *Lactococcus lactis* during growth in milk. *Applied and environmental microbiology*, 61(8), p.3024-3030.
- Juillard, V., Foucaud, C., Desmazeaud, M. & Richard, J., 1996.** Utilization of nitrogen sources during growth of *Lactococcus lactis* in milk. *Le Lait*, 76(1-2), p.12.
- Juillard, V., Furlan, S., Foucaud, C. & Richard, J., 1996.** Mixed Cultures of Proteinase-Positive and Proteinase-Negative Strains of *Lactococcus lactis* in Milk. *Journal of Dairy Science*, 79(6), p.964-970.
- Juillard, V., Guillot, A., Le Bars, D. & Gripon, J.-C., 1998.** Specificity of Milk Peptide Utilization by *Lactococcus lactis*. *Appl. Environ. Microbiol.*, 64(4), p.1230-1236.
- Juillard, V., Laan, H., Kunji, E.R., Jeronimus-Stratingh, C.M., Bruins, A.P. & Konings, W.N., 1995.** The extracellular PI-type proteinase of *Lactococcus lactis* hydrolyzes beta-casein into more than one hundred different oligopeptides. *Journal of bacteriology*, 177(12), p.3472-3478.

K

- Kalia, A. & Gupta, R.P., 2005.** Proteomics: A Paradigm Shift. *Critical Reviews in Biotechnology*, 25(4), p.173-198.
- Kalogridou-Vassiliadou, D., 1984.** Lipolytic activity and heat resistance of extracellular lipases of some Gram-negative bacteria. *Milchwissenschaft*, 39(10), p.601-603.

- Kamaly, K.M. & Marth, E.H., 1989. Enzyme Activities of Lactic Streptococci and Their Role in Maturation of Cheese: A Review. *Journal of Dairy Science*, 72(8), p.1945-1966.
- Kamaly, K.M., Takayama, K. & Marth, E.H., 1990. Acylglycerol Acylhydrolase (Lipase) Activities of *Streptococcus lactis*, *Streptococcus cremoris*, and Their Mutants. *Journal of Dairy Science*, 73(2), p.280-290.
- Kanjee, U., Ogata, K. & Houry, W.A., 2012. Direct binding targets of the stringent response alarmone (p)ppGpp. *Molecular microbiology*, 85(6), p.1029-1043.
- Karpievitch, Y.V., Hill, E.G., Leclerc, A.P., Dabney, A.R. & Almeida, J.S., 2009. An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. *PloS one*, 4(9), p.e7087.
- Kashket, E.R., 1987. Bioenergetics of lactic acid bacteria: cytoplasmic pH and osmotolerance. *FEMS Microbiology Letters*, 46(3), p.233-244.
- Kato, H., Shiwa, Y., Oshima, K., Machii, M., Araya-Kojima, T., Zendo, T., Shimizu-Kadota, M., Hattori, M., Sonomoto, K. & Yoshikawa, H., 2012. Complete Genome Sequence of *Lactococcus Lactis* IO-1, a Lactic Acid Bacterium That Utilizes Xylose and Produces High Levels of L-Lactic Acid. *Journal of Bacteriology*, 194(8), p.2102-2103.
- Katz, M., Medina, R., Gonzalez, S. & Oliver, G., 2002. Esterolytic and lipolytic activities of lactic acid bacteria isolated from ewe's milk and cheese. *Journal of food protection*, 65(12), p.1997-2001.
- Kawano, S., Shimamura, T., Niida, A., Imoto, S., Yamaguchi, R., Nagasaki, M., Yoshida, R., Print, C. & Miyano, S., 2012. Identifying gene pathways associated with cancer characteristics via sparse statistical methods. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, 9(4), p.966-972.
- Kelly, W. & Ward, L., 2002. Genotypic vs. phenotypic biodiversity in *Lactococcus lactis*. *Microbiology*, 148(11), p.3332-3333.
- Kelly, W.J., Ward, L.J.H. & Leahy, S.C., 2010. Chromosomal Diversity in *Lactococcus lactis* and the Origin of Dairy Starter Cultures. *Genome Biology and Evolution*, 2, p.729-744.
- Kempler, G.M. & McKay, L.L., 1980a Improved Medium for Detection of Citrate-Fermenting *Streptococcus lactis* subsp. *diacetylactis*. *Applied and Environmental Microbiology*, 39(4), p.926-927.
- Kersiene, M., Adams, A., Dubra, A., Kimpe, N.D. & Leskauskaitė, D., 2008. Interactions between flavour release and rheological properties in model custard desserts: Effect of starch concentration and milk fat. *Food Chemistry*, 108(4), p.1183-1191.
- Khuri, A.I. & Cornell, J.A., 1996. *Response Surfaces: Designs and Analyses*, Marcel Dekker Incorporated.
- Kieronczyk, A., Cachon, R., Feron, G. & Yvon, M., 2006. Addition of oxidizing or reducing agents to the reaction medium influences amino acid conversion to aroma compounds by *Lactococcus lactis*. *Journal of Applied Microbiology*, 101(5), p.1114-1122.
- Kieronczyk, A., Skeie, S., Langsrud, T., Le Bars, D. & Yvon, M., 2004. The nature of aroma compounds produced in a cheese model by glutamate dehydrogenase positive *Lactobacillus INF15D* depends on its relative aminotransferase activities towards the different amino acids. *International Dairy Journal*, 14(3), p.227-235.
- Kilstrup, M., 2005. Proteomics of *Lactococcus lactis*: Phenotypes for a Domestic Bacterium. In I. Humphery-Smith & M. Hecker, éd. *Microbial Proteomics*. John Wiley & Sons, Inc., p. 149-178.
- Kilstrup, M., Hammer, K., Ruhdal Jensen, P. & Martinussen, J., 2005. Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiology Reviews*, 29(3), p.555-590.
- Kilstrup, M. & Martinussen, J., 1998. A transcriptional activator, homologous to the *Bacillus subtilis* PurR repressor, is required for expression of purine biosynthetic genes in *Lactococcus lactis*. *Journal of bacteriology*, 180(15), p.3907-3916.
- Klaenhammer, T.R., 1988. Bacteriocins of lactic acid bacteria. *Biochimie*, 70(3), p.337-349.

- Klaenhammer, T.R., 1993.** Genetics of bacteriocins produced by lactic acid bacteria. *FEMS microbiology reviews*, 12(1-3), p.39-85.
- Klaenhammer, T.R., Azcarate-Peril, M.A., Altermann, E. & Barrangou, R., 2007.** Influence of the dairy environment on gene expression and substrate utilization in lactic acid bacteria. *The Journal of nutrition*, 137(3 Suppl 2), p.748S-50S.
- Knoshaug, E.P., Ahlgren, J.A. & Trempey, J.E., 2000.** Growth associated exopolysaccharide expression in *Lactococcus lactis* subspecies *cremoris* Ropy352. *Journal of dairy science*, 83(4), p.633-640.
- Kok, J., 1990.** Genetics of the proteolytic system of lactic acid bacteria. *FEMS microbiology reviews*, 7(1-2), p.15-42.
- Kok, J., Buist, G., Zomer, Aldert L, van Hijum, S.A.F.T. & Kuipers, O.P., 2005.** Comparative and functional genomics of lactococci. *FEMS microbiology reviews*, 29(3), p.411-433.
- Konings, W.N., Poolman, B. & Driessen, A.J., 1989.** Bioenergetics and solute transport in lactococci. *Critical reviews in microbiology*, 16(6), p.419-476.
- Kovaleva, G.Y. & Gelfand, M.S., 2007.** Transcriptional regulation of the methionine and cysteine transport and metabolism in streptococci. *FEMS microbiology letters*, 276(2), p.207-215.
- Kowalczyk, M. & Bardowski, J., 2007.** Regulation of Sugar Catabolism in *Lactococcus lactis*. *Critical Reviews in Microbiology*, 33(1), p.1-13.
- Van Kranenburg, R., Boels, I.C., Kleerebezem, M. & de Vos, W.M., 1999.** Genetics and engineering of microbial exopolysaccharides for food: approaches for the production of existing and novel polysaccharides. *Current opinion in biotechnology*, 10(5), p.498-504.
- Kranenburg, R. van, Marugg, J.D., Van Swam, I.I., Willem, N.J. & De Vos, W.M., 1997.** Molecular characterization of the plasmid-encoded eps gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis*. *Molecular Microbiology*, 24(2), p.387-397.
- Kristo, E., Biliaderis, C.G. & Tzanetakis, N., 2003.** Modelling of the acidification process and rheological properties of milk fermented with a yogurt starter culture using response surface methodology. *Food Chemistry*, 83(3), p.437-446.
- Kuhn, M., 2012.** *desirability: Desirability Function Optimization and Ranking*, Available at: <http://cran.r-project.org/web/packages/desirability/index.html> [Consulté le août 16, 2013].
- Kütahya, O.E., Starrenburg, M.J.C., Rademaker, J.L.W., Klaassen, C.H.W., Vlieg, J.E.T. van H., Smid, E.J. & Kleerebezem, M., 2011.** High-Resolution Amplified Fragment Length Polymorphism Typing of *Lactococcus lactis* Strains Enables Identification of Genetic Markers for Subspecies-Related Phenotypes. *Applied and Environmental Microbiology*, 77(15), p.5192-5198.
- Kwak, E.-J. & Lim, S.-I., 2004.** The effect of sugar, amino acid, metal ion, and NaCl on model Maillard reaction under pH control. *Amino Acids*, 27(1), p.85-90.

L

- Lahtvee, P.-J., Adamberg, K., Arike, L., Nahku, R., Aller, K. & Vilu, R., 2011.** Multi-omics approach to study the growth efficiency and amino acid metabolism in *Lactococcus lactis* at various specific growth rates. *Microbial Cell Factories*, 10(1), p.12.
- Landaud, S., Lieben, P. & Picque, D., 1998.** Quantitative analysis of diacetyl, pentanedione and their precursors during beer fermentation by an accurate GC/MS method. *Journal of the Institute of Brewing*, 104(2), p.93-99.
- Lapujade, P., Coccagn-Bousquet, M. & Loubiere, P., 1998.** Glutamate Biosynthesis in *Lactococcus lactis* subsp. *lactis* NCDO 2118. *Applied and Environmental Microbiology*, 64(7), p.2485.
- Larsen, N., Boye, M., Siegmundfeldt, H. & Jakobsen, M., 2006.** Differential Expression of Proteins and Genes in the Lag Phase of *Lactococcus lactis* subsp. *lactis* Grown in Synthetic Medium and Reconstituted Skim Milk. *Applied and Environmental Microbiology*, 72(2), p.1173-1179.
- Larsen, R., Buist, G., Kuipers, O.P. & Kok, J., 2004.** ArgR and AhrC are both required for regulation of arginine metabolism in *Lactococcus lactis*. *Journal of bacteriology*, 186(4), p.1147-1157.

- Larsen, R., Kloosterman, T.G., Kok, J. & Kuipers, O.P., 2006.** GlnR-mediated regulation of nitrogen metabolism in *Lactococcus lactis*. *Journal of bacteriology*, 188(13), p.4978-4982.
- Law, B.A., Sezgin, E. & Sharpe, M.E., 1976.** Amino Acid Nutrition of Some Commercial Cheese Starters in Relation to Their Growth in Peptone-Supplemented Whey Media. *Journal of Dairy Research*, 43(02), p.291-300.
- Law, B.A., Sharpe, M.E. & Reiter, B., 1974.** The release of intracellular dipeptidase from starter streptococci during Cheddar cheese ripening. *Journal of Dairy Research*, 41(01), p.137-146.
- Ledeboer, A.M., Bezemer, S., de Haard, J.J.W., Schaffers, I.M., Verrips, C.T., van Vliet, C., Düsterhöft, E.-M., Zoon, P., Moineau, S. & Frenken, L.G.J., 2002.** Preventing Phage Lysis of *Lactococcus Lactis* in Cheese Production Using A Neutralizing Heavy-Chain Antibody Fragment from Llama. *Journal of Dairy Science*, 85(6), p.1376-1382.
- Lee, S., Epstein, M.P., Duncan, R. & Lin, X., 2012.** Sparse principal component analysis for identifying ancestry-informative markers in genome-wide association studies. *Genetic epidemiology*, 36(4), p.293-302.
- Lenth, R.V., 2013.** *rsm: Response-surface analysis*, Available at: <http://cran.r-project.org/web/packages/rsm/index.html> [Consulté le août 16, 2013].
- Lepeuple, Gemert, E.V. & Chapot-Chartier, M.-P., 1998.** Analysis of the Bacteriolytic Enzymes of the Autolytic *Lactococcus lactis* subsp. *cremoris* Strain AM2 by Renaturing Polyacrylamide Gel Electrophoresis: Identification of a Prophage-Encoded Enzyme. *Applied and Environmental Microbiology*, 64(11), p.4142.
- Lepeuple, Vassal, L., Cesselin, B., Delacroix-Buchet, A., Gripon, J.-C. & Chapot-Chartier, M.-P., 1998.** Involvement of a Prophage in the Lysis of *Lactococcus lactis* subsp. *cremoris* AM2 during Cheese Ripening. *International Dairy Journal*, 8(7), p.667-674.
- Leroy, F. & De Vuyst, L., 2004.** Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends in Food Science & Technology*, 15(2), p.67-78.
- Lesniak, J., Barton, W.A. & Nikolov, D.B., 2003.** Structural and functional features of the *Escherichia coli* hydroperoxide resistance protein OsmC. *Protein science: a publication of the Protein Society*, 12(12), p.2838-2843.
- Levdikov, V.M., Blagova, E., Colledge, V.L., Lebedev, A.A., Williamson, D.C., Sonenshein, A.L. & Wilkinson, A.J., 2009.** Structural rearrangement accompanying ligand binding in the GAF domain of CodY from *Bacillus subtilis*. *Journal of molecular biology*, 390(5), p.1007-1018.
- Levdikov, V.M., Blagova, E., Joseph, P., Sonenshein, A.L. & Wilkinson, A.J., 2006.** The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in gram-positive bacteria. *The Journal of biological chemistry*, 281(16), p.11366-11373.
- Li, R., Khafipour, E., Krause, D.O., Entz, M.H., de Kievit, T.R. & Fernando, W.G.D., 2012.** Pyrosequencing reveals the influence of organic and conventional farming systems on bacterial communities. *PloS one*, 7(12), p.e51897.
- Li, W., Raoult, D. & Fournier, P.-E., 2009.** Bacterial strain typing in the genomic era. *FEMS Microbiology Reviews*, 33(5), p.892-916.
- Ligor, M., Jarmalaviciene, R., Szumski, M., Maruska, A. & Buszewski, B., 2008.** Determination of volatile and non-volatile products of milk fermentation processes using capillary zone electrophoresis and solid phase microextraction coupled to gas chromatography. *Journal of separation science*, 31(14), p.2707-2713.
- Linares, D.M., Kok, J. & Poolman, B., 2010.** Genome sequences of *Lactococcus lactis* MG1363 (revised) and NZ9000 and comparative physiological studies. *Journal of bacteriology*, 192(21), p.5806-5812.
- Lindahl, T. & Nyberg, B., 1972.** Rate of depurination of native deoxyribonucleic acid. *Biochemistry*, 11(19), p.3610-3618.
- Liu, G., Draper, G.C. & Donachie, W.D., 1998.** FtsK is a bifunctional protein involved in cell division and chromosome localization in *Escherichia coli*. *Molecular Microbiology*, 29(3), p.893-903.

- Liu, M., Nauta, A., Francke, C. & Siezen, R.J., 2008.** Comparative Genomics of Enzymes in Flavor-Forming Pathways from Amino Acids in Lactic Acid Bacteria. *Applied and Environmental Microbiology*, 74(15), p.4590-4600.
- Liu, S.-Q., Holland, R. & Crow, V.L., 1998.** Ethyl Butanoate Formation by Dairy Lactic Acid Bacteria. *International Dairy Journal*, 8(7), p.651-657.
- Liu, S.-Q., Holland, R. & Crow, V.L., 2001.** Purification and properties of intracellular esterases from *Streptococcus thermophilus*. *International Dairy Journal*, 11(1-2), p.27-35.
- Llanos, R.M., Harris, C.J., Hillier, A.J. & Davidson, B.E., 1993.** Identification of a novel operon in *Lactococcus lactis* encoding three enzymes for lactic acid synthesis: phosphofructokinase, pyruvate kinase, and lactate dehydrogenase. *Journal of bacteriology*, 175(9), p.2541-2551.
- Looijesteijn, P.J., Boels, I.C., Kleerebezem, M. & Hugenholtz, J., 1999.** Regulation of Exopolysaccharide Production by *Lactococcus lactis* subsp. *cremoris* by the Sugar Source. *Applied and Environmental Microbiology*, 65(11), p.5003.
- Luesink, E.J., van Herpen, R.E., Grossiord, B.P., Kuipers, O.P. & de Vos, W.M., 1998.** Transcriptional activation of the glycolytic *las* operon and catabolite repression of the *gal* operon in *Lactococcus lactis* are mediated by the catabolite control protein CcpA. *Molecular microbiology*, 30(4), p.789-798.

M

- Magni, C., de Mendoza, D., Konings, W.N. & Lolkema, J.S., 1999.** Mechanism of Citrate Metabolism in *Lactococcus lactis*: Resistance against Lactate Toxicity at Low pH. *Journal of Bacteriology*, 181(5), p.1451-1457.
- Magnusson, L.U., Farewell, A. & Nyström, T., 2005.** ppGpp: a global regulator in *Escherichia coli*. *Trends in microbiology*, 13(5), p.236-242.
- Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. & Spratt, B.G., 1998.** Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), p.3140-3145.
- Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar, D., Lucas, S., Huang, K., Goodstein, D.M., Hawkins, T., Plengvidhya, V., Welker, D., Hughes, J., Goh, Y., Benson, A., Baldwin, K., Lee, J.-H., Díaz-Muñiz, I., Dosti, B., Smeianov, V., Wechter, W., Barabote, R., Lorca, G., Altermann, E., Barrangou, R., Ganesan, B., Xie, Y., Rawsthorne, H., Tamir, D., Parker, C., Breidt, F., Broadbent, J., Hutkins, R., O'Sullivan, D., Steele, J., Unlu, G., Saier, M., Klaenhammer, T., Richardson, P., Kozyavkin, S., Weimer, B. & Mills, D., 2006.** Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42), p.15611-15616.
- Maligoy, M., Mercade, M., Coccagn-Bousquet, M. & Loubiere, P., 2008.** Transcriptome analysis of *Lactococcus lactis* in coculture with *Saccharomyces cerevisiae*. *Applied and environmental microbiology*, 74(2), p.485-494.
- Mariaca, R.G., Fernandez-Garcia, E., Mohedano, A.F. & Nufiez, M., 2001.** Volatile Fraction of Ewe's Milk Semi-Hard Cheese Manufactured with and without the Addition of a Cysteine Proteinase. *Food Science and Technology International*, 7(2), p.131-139.
- Marshall, V.M., 1987.** Lactic acid bacteria: starters for flavour. *FEMS Microbiology Letters*, 46(3), p.327-336.
- Marshall, V.M. & Cole, W.M., 1983.** Threonine aldolase and alcohol dehydrogenase activities in *Lactobacillus bulgaricus* and *Lactobacillus acidophilus* and their contribution to flavour production in fermented milks. *Journal of Dairy Research*, 50(03), p.375-379.

- Martinussen, J. & Hammer, K., 1998.** The *carB* gene encoding the large subunit of carbamoylphosphate synthetase from *Lactococcus lactis* is transcribed monocistronically. *Journal of bacteriology*, 180(17), p.4380-4386.
- Martinussen, J., Schallert, J., Andersen, B. & Hammer, K., 2001.** The pyrimidine operon *pyrRPB-carA* from *Lactococcus lactis*. *Journal of bacteriology*, 183(9), p.2785-2794.
- Mazzoli, R., Pessione, E., Dufour, M., Laroute, V., Giuffrida, M.G., Giunta, C., Cocaïgn-Bousquet, M. & Loubière, P., 2010.** Glutamate-induced metabolic changes in *Lactococcus lactis* NCDO 2118 during GABA production: combined transcriptomic and proteomic analysis. *Amino acids*, 39(3), p.727-737.
- Mc Sweeney, P.L.H., 2004.** Biochemistry of cheese ripening. *International Journal of Dairy Technology*, 57(2/3), p.127-144.
- McSweeney, P.L.H. & Sousa, M.J., 2000.** Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. *Le Lait*, 80(3), p.32.
- Mechold, U., Potrykus, K., Murphy, H., Murakami, K.S. & Cashel, M., 2013.** Differential regulation by ppGpp versus pppGpp in *Escherichia coli*. *Nucleic acids research*, 41(12), p.6175-6189.
- Meding, S., Nitsche, U., Balluff, B., Elsner, M., Rauser, S., Schöne, C., Nipp, M., Maak, M., Feith, M., Ebert, M.P., Friess, H., Langer, R., Höfler, H., Zitzelsberger, H., Rosenberg, R. & Walch, A., 2012.** Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging. *Journal of proteome research*, 11(3), p.1996-2003.
- Meghrou, J., Fliss, I., Bouksaim, M. & Lacroix, C., 1999.** Digoxigenin-labeled probe for rapid identification of nisinogenic *Lactococcus lactis* strains. *FEMS microbiology letters*, 171(1), p.43-48.
- Mercade, M., 2002.** *Analyse métabolique et énergétique de Lactobacillus bulgaricus sous contrainte d'extrémophilie industrielle*. Thèse de doctorat. Toulouse, France: Institut national des sciences appliquées.
- Metchnikoff, E. & Mitchell, P.C., 1908.** *The prolongation of life: optimistic studies.*, New York & London: G. P. Putnam's sons. Available at: <http://catalog.hathitrust.org/Record/001587341> [Consulté le septembre 18, 2013].
- Meyrand, M., Boughammoura, A., Courtin, P., Mézange, C., Guillot, A. & Chapot-Chartier, M.-P., 2007.** Peptidoglycan N-acetylglucosamine deacetylation decreases autolysis in *Lactococcus lactis*. *Microbiology (Reading, England)*, 153(Pt 10), p.3275-3285.
- Michels, P.A.M., Michels, J.P.J., Boonstra, J. & Konings, W.N., 1979.** Generation of an electrochemical proton gradient in bacteria by the excretion of metabolic end products. *FEMS Microbiology Letters*, 5(5), p.357-364.
- Mills, O.E. & Thomas, T., 1981.** Nitrogen sources for growth of lactic streptococci in milk. *New Zealand Journal Dairy Science Technology*, 16, p.43-55.
- Mills, S., McAuliffe, O.E., Coffey, A., Fitzgerald, G.F. & Ross, R.P., 2006.** Plasmids of lactococci – genetic accessories or genetic necessities? *FEMS Microbiology Reviews*, 30(2), p.243-273.
- Mills, S., O'sullivan, O., Hill, C., Fitzgerald, G. & Ross, R.P., 2010.** The changing face of dairy starter culture research: From genomics to economics. *International Journal of Dairy Technology*, 63(2), p.149-170.
- Milo, C. & Reineccius, G.A., 1997.** Identification and Quantification of Potent Odorants in Regular-Fat and Low-Fat Mild Cheddar Cheese. *Journal of Agricultural and Food Chemistry*, 45(9), p.3590-3594.
- Mitchell, P., 1974.** Chemiosmotic Coupling in Energy Transduction: A Logical Development of Biochemical Knowledge. In J. Avery, éd. *Membrane Structure and Mechanisms of Biological Energy Transduction*. Springer US, p. 5-24.
- Moens, L., Evans, R.J., Looker, M.J. & Nimlos, M.R., 2004.** A comparison of the Maillard reactivity of proline to other amino acids using pyrolysis-molecular beam mass spectrometry. *Fuel*, 83(11-12), p.1433-1443.

- Mogk, A., Deuerling, E., Vorderwülbecke, S., Vierling, E. & Bukau, B., 2003.** Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Molecular microbiology*, 50(2), p.585-595.
- Molenaar, D., Bosscher, J.S., ten Brink, B., Driessen, A.J. & Konings, W.N., 1993.** Generation of a proton motive force by histidine decarboxylation and electrogenic histidine/histamine antiport in *Lactobacillus buchneri*. *Journal of bacteriology*, 175(10), p.2864-2870.
- Monnet, C., Haddad, S. & Corrieu, G., 1999.** An improved method for screening alpha-acetolactate producing mutants. *Journal of microbiological methods*, 37(2), p.183-185.
- Monnet, C., Nardi, M., Hols, P., Gulea, M., Corrieu, G. & Monnet, V., 2003.** Regulation of branched-chain amino acid biosynthesis by alpha-acetolactate decarboxylase in *Streptococcus thermophilus*. *Letters in applied microbiology*, 36(6), p.399-405.
- Monnet, C., Schmilt, P. & Divies, C., 1994.** Diacetyl Production in Milk by an α -Acetolactic Acid Accumulating Strain of *Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*. *Journal of Dairy Science*, 77(10), p.2916-2924.
- Monnet, C., Schmitt, P. & Divies, C., 1997.** Development and Use of a Screening Procedure for Production of α -Acetolactate by *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* Strains. *Applied and environmental microbiology*, 63(2), p.793-795.
- Monnet, V. & Gripon, J., 1994.** Metabolisme azoté des bactéries lactiques. In *Bactéries lactiques*. Loriga, Uriage, France: de Roissart H., Luquet F.M, p. 331-347.
- Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B. & Worm, B., 2011.** How Many Species Are There on Earth and in the Ocean? *PLoS Biol*, 9(8), p.e1001127.
- Munck, L., Nørgaard, L., Engelsen, S.B., Bro, R. & Andersson, C.A., 1998.** Chemometrics in food science—a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemometrics and Intelligent Laboratory Systems*, 44(1-2), p.31-60.

N

- Nardi, M., Fiez-Vandal, C., Tailliez, P. & Monnet, V., 2002.** The EstA esterase is responsible for the main capacity of *Lactococcus lactis* to synthesize short chain fatty acid esters in vitro. *Journal of Applied Microbiology*, 93(6), p.994-1002.
- Neves, A.R., Pool, Wietske A, Kok, J., Kuipers, O.P. & Santos, H., 2005.** Overview on sugar metabolism and its control in *Lactococcus lactis* - the input from in vivo NMR. *FEMS microbiology reviews*, 29(3), p.531-554.
- Neves, A.R., Ramos, A., Costa, H., van Swam, I.I., Hugenholtz, J., Kleerebezem, M., de Vos, W. & Santos, H., 2002.** Effect of different NADH oxidase levels on glucose metabolism by *Lactococcus lactis*: kinetics of intracellular metabolite pools determined by in vivo nuclear magnetic resonance. *Applied and environmental microbiology*, 68(12), p.6332-6342.
- Nicolaou, N. & Goodacre, R., 2008.** Rapid and quantitative detection of the microbial spoilage in milk using Fourier transform infrared spectroscopy and chemometrics. *The Analyst*, 133(10), p.1424-1431.
- Nilsson, D. & Lauridsen, A.A., 1992.** Isolation of purine auxotrophic mutants of *Lactococcus lactis* and characterization of the gene hpt encoding hypoxanthine guanine phosphoribosyltransferase. *Molecular and General Genetics MGG*, 235(2-3), p.359-364.
- Nomura, M., Kimoto, H., Someya, Y. & Suzuki, I., 1999.** Novel characteristic for distinguishing *Lactococcus lactis* subsp. *lactis* from subsp. *cremoris*. *International Journal of Systematic Bacteriology*, 49(1), p.163-166.
- Nomura, M., Kobayashi, M., Narita, T., Kimoto-Nira, H. & Okamoto, T., 2006.** Phenotypic and molecular characterization of *Lactococcus lactis* from milk and plants. *Journal of applied microbiology*, 101(2), p.396-405.

- Nomura, M., Kobayashi, M. & Okamoto, T., 2002.** Rapid PCR-based method which can determine both phenotype and genotype of *Lactococcus lactis* subspecies. *Applied and environmental microbiology*, 68(5), p.2209-2213.
- Nordkvist, M., Jensen, N.B.S. & Villadsen, J., 2003.** Glucose metabolism in *Lactococcus lactis* MG1363 under different aeration conditions: requirement of acetate to sustain growth under microaerobic conditions. *Applied and environmental microbiology*, 69(6), p.3462-3468.
- Nouaille, S., Even, S., Charlier, C., Le Loir, Y., Coccagn-Bousquet, M. & Loubière, P., 2009.** Transcriptomic response of *Lactococcus lactis* in mixed culture with *Staphylococcus aureus*. *Applied and environmental microbiology*, 75(13), p.4473-4482.
- Novak, L., 1998.** Le réseau métabolique de *Lactococcus lactis* ssp. *lactis* NCDO 2118: interaction entre métabolisme carboné et azoté. Thèse de doctorat-INSA de Toulouse. France.
- Novak, L., Coccagn-Bousquet, M., Lindley, N.D. & Loubiere, P., 1997.** Metabolism and Energetics of *Lactococcus lactis* during Growth in Complex or Synthetic Media. *Applied and Environmental Microbiology*, 63(7), p.2665-2670.
- Novák, L. & Loubiere, P., 2000.** The Metabolic Network of *Lactococcus lactis*: Distribution of ¹⁴C-Labeled Substrates between Catabolic and Anabolic Pathways. *Journal of Bacteriology*, 182(4), p.1136-1143.
- Nyström, T., 2004.** Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? *Molecular microbiology*, 54(4), p.855-862.

O

- O'Connell-Motherway, M., van Sinderen, D., Morel-Deville, F., Fitzgerald, G.F., Ehrlich, S.D. & Morel, P., 2000.** Six putative two-component regulatory systems isolated from *Lactococcus lactis* subsp. *cremoris* MG1363. *Microbiology (Reading, England)*, 146 (Pt 4), p.935-947.
- O'Donovan, C.M., Wilkinson, M.G., Guinee, T.P. & Fox, P.F., 1996.** An investigation of the autolytic properties of three lactococcal strains during cheese ripening. *International Dairy Journal*, 6(11-12), p.1149-1165.
- O'Flaherty, S. & Klaenhammer, T.R., 2011.** The impact of omic technologies on the study of food microbes. *Annual review of food science and technology*, 2, p.353-371.
- O'Hara, R.B. & Sillanpää, M.J., 2009.** A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4(1), p.85-117.
- O'Sullivan, D., Ross, R.P., Fitzgerald, G.F. & Coffey, A., 2000.** Investigation of the Relationship between Lysogeny and Lysis of *Lactococcus lactis* in Cheese Using Prophage-Targeted PCR. *Applied and Environmental Microbiology*, 66(5), p.2192.
- Obis, D., Guillot, A., Gripon, J.C., Renault, P., Bolotin, A. & Mistou, M.Y., 1999.** Genetic and biochemical characterization of a high-affinity betaine uptake system (BusA) in *Lactococcus lactis* reveals a new functional organization within bacterial ABC transporters. *Journal of bacteriology*, 181(20), p.6238-6246.
- Oh, E., Lu, M., Park, C., Park, C., Oh, H.B., Lee, S.Y. & Lee, J., 2011.** Dynamic modeling of lactic acid fermentation metabolism with *Lactococcus lactis*. *Journal of microbiology and biotechnology*, 21(2), p.162-169.
- Oliveira, A.P., Nielsen, J. & Förster, J., 2005.** Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC microbiology*, 5, p.39.
- Ott, A., Fay, L.B. & Chaintreau, A., 1997.** Determination and Origin of the Aroma Impact Compounds of Yogurt Flavor. *Journal of Agricultural and Food Chemistry*, 45(3), p.850-858.

- Ott, A., Germond, J.-E., Baumgartner, M. & Chaintreau, A., 1999.** Aroma Comparisons of Traditional and Mild Yogurts: Headspace Gas Chromatography Quantification of Volatiles and Origin of α -Diketones. *Journal of Agricultural and Food Chemistry*, 47(6), p.2379-2385.
- Ott, A, Germond, J.E. & Chaintreau, A., 2000.** Vicinal diketone formation in yogurt: (^{13}C) precursors and effect of branched-chain amino acids. *Journal of Agricultural and Food Chemistry*, 48(3), p.724-731.
- Ott, A, Germond, J.-E. & Chaintreau, A., 2000.** Vicinal Diketone Formation in Yogurt: ^{13}C Precursors and Effect of Branched-Chain Amino Acids. *Journal of Agricultural and Food Chemistry*, 48(3), p.724-731.
- Ott, A., Hugi, A., Baumgartner, M. & Chaintreau, A., 2000.** Sensory investigation of yogurt flavor perception: mutual influence of volatiles and acidity. *Journal of agricultural and food chemistry*, 48(2), p.441-450.
- Otto, R., Lageveen, R.G., Veldkamp, H. & Konings, W.N., 1982.** Lactate efflux-induced electrical potential in membrane vesicles of *Streptococcus cremoris*. *Journal of Bacteriology*, 149(2), p.733-738.
- Otto, R., Sonnenberg, A.S., Veldkamp, H. & Konings, W.N., 1980.** Generation of an electrochemical proton gradient in *Streptococcus cremoris* by lactate efflux. *Proceedings of the National Academy of Sciences of the United States of America*, 77(9), p.5502-5506.

P

- Palmfeldt, J., Levander, F., Hahn-Hägerdal, B. & James, P., 2004.** Acidic proteome of growing and resting *Lactococcus lactis* metabolizing maltose. *PROTEOMICS*, 4(12), p.3881-3898.
- Parvin, C.A., 2004.** An Introduction to Multivariate Statistical Analysis, 3rd ed. T.W. Anderson. Hoboken, NJ: John Wiley & Sons, 2003, 742 pp., \$99.95, hardcover. ISBN 0-471-36091-0. *Clinical Chemistry*, 50(5), p.981-982.
- Passerini, D., 2011.** *Diversité génétique, génomique et fonctionnelle de Lactococcus lactis*. LISBP & LMGM: Université de Toulouse.
- Passerini, D., Beltramo, C., Coddeville, M., Quentin, Y., Ritzenthaler, P., Daveran-Mingot, M.-L. & Le Bourgeois, P., 2010.** Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. *PloS one*, 5(12), p.e15306.
- Passerini, D., Laroute, V., Coddeville, M., Le Bourgeois, P., Loubière, P., Ritzenthaler, P., Cocaigh-Bousquet, M. & Daveran-Mingot, M.-L., 2013.** New insights into *Lactococcus lactis* diacetyl- and acetoin-producing strains isolated from diverse origins. *International journal of food microbiology*, 160(3), p.329-336.
- Patek, M., 2007.** Branched-Chain Amino Acids. In *Amino acid biosynthesis pathways regulation and metabolic engineering*. Microbiology Monographs. Springer-Verlag Berlin Heidelberg, p. 129-162.
- Paul, Brian J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W. & Gourse, R.L., 2004.** DksA: A Critical Component of the Transcription Initiation Machinery that Potentiates the Regulation of rRNA Promoters by ppGpp and the Initiating NTP. *Cell*, 118(3), p.311-322.
- Paul, Brian J, Ross, W., Gaal, T. & Gourse, R.L., 2004.** rRNA transcription in *Escherichia coli*. *Annual review of genetics*, 38, p.749-770.
- Pedersen, M.B., Iversen, S.L., Sørensen, K.I. & Johansen, E., 2005.** The long and winding road from the research laboratory to industrial applications of lactic acid bacteria. *FEMS microbiology reviews*, 29(3), p.611-624.

- Pérez-Enciso, M. & Tenenhaus, M., 2003.** Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Human genetics*, 112(5-6), p.581-592.
- Petranovic, D., Guédon, E., Sperandio, B., Delorme, C., Ehrlich, D. & Renault, P., 2004.** Intracellular effectors regulating the activity of the *Lactococcus lactis* CodY pleiotropic transcription regulator. *Molecular microbiology*, 53(2), p.613-621.
- Phillips, A.T., Nuss, J.I., Moosic, J. & Foshay, C., 1972.** Alternate Pathway for Isoleucine Biosynthesis in *Escherichia coli*. *Journal of Bacteriology*, 109(2), p.714-719.
- Picard, F., Dressaire, C., Girbal, L. & Coccagn-Bousquet, M., 2009.** Examination of post-transcriptional regulations in prokaryotes by integrative biology. *Comptes rendus biologiques*, 332(11), p.958-973.
- Picard, F., Loubière, P., Girbal, L. & Coccagn-Bousquet, M., 2013.** The significance of translation regulation in the stress response. *BMC genomics*, 14(1), p.588.
- Picard, F., Milhem, H., Loubière, P., Laurent, B., Coccagn-Bousquet, M. & Girbal, L., 2012.** Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. *BMC genomics*, 13, p.528.
- Picque, D., Perret, B., Latrille, E. & Corrieu, G., 1992.** Caractérisation et classification de bactéries lactiques à partir de la mesure de leur cinétique d'acidification. *Lebensmittel - Wissenschaft + Technologie*, 25(2), p.181-186.
- Poolman, B., 1993.** Energy transduction in lactic acid bacteria. *FEMS Microbiology Reviews*, 12(1-3), p.125-147.
- Poolman, B. & Konings, W.N., 1993.** Secondary solute transport in bacteria. *Biochimica et biophysica acta*, 1183(1), p.5-39.
- Pope, G.A., MacKenzie, D.A., Defernez, M., Aroso, M.A.M.M., Fuller, L.J., Mellon, F.A., Dunn, W.B., Brown, M., Goodacre, R., Kell, D.B., Marvin, M.E., Louis, E.J. & Roberts, I.N., 2007.** Metabolic footprinting as a tool for discriminating between brewing yeasts. *Yeast (Chichester, England)*, 24(8), p.667-679.
- Presser, K.A., Ratkowsky, D.A. & Ross, T., 1997.** Modelling the growth rate of *Escherichia coli* as a function of pH and lactic acid concentration. *Applied and Environmental Microbiology*, 63(6), p.2355.
- Pudlik, A.M. & Lolkema, J.S., 2013.** Uptake of α -ketoglutarate by citrate transporter CitP drives transamination in *Lactococcus lactis*. *Applied and environmental microbiology*, 79(4), p.1095-1101.

Q & R

- Qiao, M., Zhao, K., Zhu, D., Zhang, Xu, Feng, Z., Xuanyuan, Z., Bai, Y., Xu, H. & Zhang, Xiuming, 2013.** Nisin immune tolerance related gene *irpT* of *Lactococcus lactis* N8 and applications of *irpT*.
- Rademaker, Herbet, H., Starrenburg, M.J.C., Naser, S.M., Gevers, D., Kelly, W.J., Hugenholtz, J., Swings, J. & van Hylckama Vlieg, J.E.T., 2007.** Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel multilocus sequence analysis scheme and (GTG)5-PCR fingerprinting. *Applied and environmental microbiology*, 73(22), p.7128-7137.
- Rademaker, J.L.W., Herbet, H., Starrenburg, M.J.C., Naser, S.M., Gevers, D., Kelly, W.J., Hugenholtz, J., Swings, J. & Vlieg, J.E.T. van H., 2007.** Diversity Analysis of Dairy and Nondairy *Lactococcus lactis* Isolates, Using a Novel Multilocus Sequence Analysis Scheme and (GTG)5-PCR Fingerprinting. *Applied and Environmental Microbiology*, 73(22), p.7128-7137.
- Radke-Mitchell, L.C. & Sandine, W.E., 1986.** Influence of Temperature on Associative Growth of *Streptococcus thermophilus* and *Lactobacillus bulgaricus*. *Journal of Dairy Science*, 69(10), p.2558-2568.
- Rallu, F., Gruss, A. & Maguin, E., 1996.** *Lactococcus lactis* and stress. *Antonie van Leeuwenhoek*, 70(2-4), p.243-251.

- Ray, S., Renu, D., Srivastava, R., Gollapalli, K., Taur, S., Jhaveri, T., Dhali, S., Chennareddy, S., Potla, A., Dikshit, J.B., Srikanth, R., Gogtay, N., Thatte, U., Patankar, S. & Srivastava, S., 2012. Proteomic investigation of falciparum and vivax malaria for identification of surrogate protein markers. *PloS one*, 7(8), p.e41751.
- Raynaud, S., 2006. *Régulation métabolique et transcriptionnelle de l'autoacidification chez Lactococcus lactis*. Thèse. INSA de Toulouse. Université de Toulouse.
- Raynaud, S., Perrin, R., Cocaïgn-Bousquet, M. & Loubiere, P., 2005. Metabolic and transcriptomic adaptation of *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* in response to autoacidification and temperature downshift in skim milk. *Applied and environmental microbiology*, 71(12), p.8016-8023.
- Reineccius, G., 1998. *Source book of flavors*, Springer.
- Reisinger, P., Seidel, H., Tschesche, H. & Hammes, W.P., 1980. The effect of nisin on murein synthesis. *Archives of Microbiology*, 127(3), p.187-193.
- Reizer, J., Novotny, M.J., Panos, C. & Saier, M.H., Jr, 1983. Mechanism of inducer expulsion in *Streptococcus pyogenes*: a two-step process activated by ATP. *Journal of bacteriology*, 156(1), p.354-361.
- Reizer, J. & Saier, M.H., Jr, 1983. Involvement of lactose enzyme II of the phosphotransferase system in rapid expulsion of free galactosides from *Streptococcus pyogenes*. *Journal of bacteriology*, 156(1), p.236-242.
- Riepe, H.R., Pillidge, C.J., Gopal, P.K. & McKay, L.L., 1997. Characterization of the Highly Autolytic *Lactococcus lactis* subsp. *cremoris* Strains CO and 2250. *Applied and Environmental Microbiology*, 63(10), p.3757-3763.
- Rijnen, L., Bonneau, S. & Yvon, M., 1999. Genetic Characterization of the Major Lactococcal Aromatic Aminotransferase and Its Involvement in Conversion of Amino Acids to Aroma Compounds. *Applied and Environmental Microbiology*, 65(11), p.4873-4880.
- Rijnen, L., Courtin, P., Gripon, J.C. & Yvon, M., 2000. Expression of a heterologous glutamate dehydrogenase gene in *Lactococcus lactis* highly improves the conversion of amino acids to aroma compounds. *Applied and environmental microbiology*, 66(4), p.1354-1359.
- Rochat, T., Boudebouze, S., Gratadoux, J.-J., Blugeon, S., Gaudu, P., Langella, P. & Maguin, E., 2012. Proteomic analysis of spontaneous mutants of *Lactococcus lactis*: Involvement of GAPDH and arginine deiminase pathway in H₂O₂ resistance. *PROTEOMICS*, 12(11), p.1792-1805.
- Romeo, Y., Obis, D., Bouvier, J., Guillot, A., Fourçans, A., Bouvier, I., Gutierrez, C. & Mistou, M.-Y., 2003. Osmoregulation in *Lactococcus lactis*: BusR, a transcriptional repressor of the glycine betaine uptake system BusA. *Molecular Microbiology*, 47(4), p.1135-1147.
- Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A. & Pappin, D.J., 2004. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics*, 3(12), p.1154-1169.
- Ross, W., Vrentas, C.E., Sanchez-Vazquez, P., Gaal, T. & Gourse, R.L., 2013. The Magic Spot: A ppGpp Binding Site on *E. coli* RNA Polymerase Responsible for Regulation of Transcription Initiation. *Molecular Cell*, 50(3), p.420-429.
- Ruas-Madiedo, P., Tuinier, R., Kanning, M. & Zoon, P., 2002. Role of exopolysaccharides produced by *Lactococcus lactis* subsp. *cremoris* on the viscosity of fermented milks. *International Dairy Journal*, 12(8), p.689-695.

S

- Saeys, Y., Inza, I. & Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), p.2507-2517.

- Saier, M.H., 1989. Protein phosphorylation and allosteric control of inducer exclusion and catabolite repression by the bacterial phosphoenolpyruvate: sugar phosphotransferase system. *Microbiological Reviews*, 53(1), p.109-120.
- Salama, M., Sandine, W. & Giovannoni, S., 1991. Development and application of oligonucleotide probes for identification of *Lactococcus lactis* subsp. *cremoris*. *Applied and Environmental Microbiology*, 57(5), p.1313-1318.
- Šalplachta, J., Kubesová, A. & Horká, M., 2012. Latest improvements in CIEF: From proteins to microorganisms. *PROTEOMICS*, 12(19-20), p.2927-2936.
- Sanders, J.W., Leenhouts, K., Burghoorn, J., Brands, J.R., Venema, G. & Kok, J., 1998. A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation. *Molecular microbiology*, 27(2), p.299-310.
- Sanders, J.W., Venema, G. & Kok, J., 1999. Environmental stress responses in *Lactococcus lactis*. *FEMS Microbiology Reviews*, 23(4), p.483-501.
- Sanders, J.W., Venema, G., Kok, J. & Leenhouts, K., 1998. Identification of a sodium chloride-regulated promoter in *Lactococcus lactis* by single-copy chromosomal fusion with a reporter gene. *Molecular & general genetics: MGG*, 257(6), p.681-685.
- Savijoki, K., Ingmer, H. & Varmanen, P., 2006. Proteolytic systems of lactic acid bacteria. *Applied microbiology and biotechnology*, 71(4), p.394-406.
- Sawers, G., 1998. The anaerobic degradation of L-serine and L-threonine in enterobacteria: networks of pathways and regulatory signals. *Archives of microbiology*, 171(1), p.1-5.
- Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. & Mann, M., 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proceedings of the National Academy of Sciences of the United States of America*, 93(25), p.14440-14445.
- Shi, T., Seligson, D., Belldegrun, A.S., Palotie, A. & Horvath, S., 2005. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, 18(4), p.547-557.
- Shilov, I.V., Seymour, S.L., Patel, A.A., Loboda, A., Tang, W.H., Keating, S.P., Hunter, C.L., Nuwaysir, L.M. & Schaeffer, D.A., 2007. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Molecular & cellular proteomics: MCP*, 6(9), p.1638-1655.
- Siezen, R.J., Bayjanov, J., Renckens, B., Wels, M., van Hijum, S.A.F.T., Molenaar, D. & van Hylckama Vlieg, J.E.T., 2010. Complete genome sequence of *Lactococcus lactis* subsp. *lactis* KF147, a plant-associated lactic acid bacterium. *Journal of bacteriology*, 192(10), p.2649-2650.
- Siezen, R.J., Bayjanov, J.R., Felis, G.E., van der Sijde, M.R., Starrenburg, M., Molenaar, D., Wels, M., van Hijum, S.A.F.T. & van Hylckama Vlieg, J.E.T., 2011. Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays. *Microbial Biotechnology*, 4(3), p.383-402.
- Siezen, R.J., Starrenburg, M.J.C., Boekhorst, J., Renckens, B., Molenaar, D. & van Hylckama Vlieg, J.E.T., 2008. Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Applied and environmental microbiology*, 74(2), p.424-436.
- Singh, H. & Bennett, R.J., 2005. Milk and milk processing. In *Dairy Microbiology Handbook: Third Edition*. Richard K. Robinson, p. 1-38.
- Smit, B.a., Engels, W. j. m., Bruinsma, J., van Hylckama Vlieg, J. e. t., Wouters, J. t. m. & Smit, G., 2004. Development of a high throughput screening method to test flavour-forming capabilities of anaerobic micro-organisms. *Journal of Applied Microbiology*, 97(2), p.306-313.

- Smit, B A, Engels, W.J.M., Bruinsma, J., van Hylckama Vlieg, J.E.T., Wouters, J.T.M. & Smit, G., 2004.** Development of a high throughput screening method to test flavour-forming capabilities of anaerobic micro-organisms. *Journal of applied microbiology*, 97(2), p.306-313.
- Smit, B.A., Engels, W.J.M. & Smit, G., 2008.** Branched chain aldehydes: production and breakdown pathways and relevance for flavour in foods. *Applied Microbiology and Biotechnology*, 81(6), p.987-999.
- Smit, B. A., Engels, W.J.M., Wouters, J.T.M. & Smit, G., 2004.** Diversity of l-leucine catabolism in various microorganisms involved in dairy fermentations, and identification of the rate-controlling step in the formation of the potent flavour component 3-methylbutanal. *Applied Microbiology and Biotechnology*, 64(3), p.396-402.
- Smit, G., Smit, B.A. & Engels, W.J.M., 2005.** Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiology Reviews*, 29(3), p.591-610.
- Smith, J.S., Hillier, A.J., Lees, G.J. & Jago, G.R., 1975.** The nature of the stimulation of the growth of *Streptococcus lactis* by yeast extract. *Journal of Dairy Research*, 42(01), p.123-138.
- Smyth, G.K., 2004.** Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, 3, p.Article3.
- Snoep, J.L., de Graef, M.R., Westphal, A.H., de Kok, A., Teixeira de Mattos, M.J. & Neijssel, O.M., 1993.** Differences in sensitivity to NADH of purified pyruvate dehydrogenase complexes of *Enterococcus faecalis*, *Lactococcus lactis*, *Azotobacter vinelandii* and *Escherichia coli*: implications for their activity in vivo. *FEMS microbiology letters*, 114(3), p.279-283.
- Sonenshein, A.L., 2005.** CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. *Current opinion in microbiology*, 8(2), p.203-207.
- Sperandio, B., Polard, P., Ehrlich, D.S., Renault, P. & Guédon, E., 2005.** Sulfur amino acid metabolism and its control in *Lactococcus lactis* IL1403. *Journal of bacteriology*, 187(11), p.3762-3778.
- Spinnler, H.E. & Cordieu, G., 1989.** Automatic method to quantify starter activity based on pH measurement. *Journal of dairy research*, 56(5), p.755-764.
- Srivatsan, A. & Wang, J.D., 2008.** Control of bacterial transcription, translation and replication by (p)ppGpp. *Current opinion in microbiology*, 11(2), p.100-105.
- Stanton, C., Ross, R.P., Fitzgerald, G.F. & Van Sinderen, D., 2005.** Fermented functional foods based on probiotics and their biogenic metabolites. *Current opinion in biotechnology*, 16(2), p.198-203.
- Steen, A., Buist, G., Kramer, N.E., Jalving, R., Benus, G.F.J.D., Venema, G., Kuipers, O.P. & Kok, J., 2008.** Reduced lysis upon growth of *Lactococcus lactis* on galactose is a consequence of decreased binding of the autolysin AcmA. *Applied and environmental microbiology*, 74(15), p.4671-4679.
- Steinfath, M., Gärtner, T., Lisec, J., Meyer, R.C., Altmann, T., Willmitzer, L. & Selbig, J., 2010.** Prediction of hybrid biomass in *Arabidopsis thaliana* by selected parental SNP and metabolic markers. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 120(2), p.239-247.
- Stenz, L., Francois, P., Whiteson, K., Wolz, C., Linder, P. & Schrenzel, J., 2011.** The CodY pleiotropic repressor controls virulence in gram-positive pathogens. *FEMS immunology and medical microbiology*, 62(2), p.123-139.
- Stiglic, G., Rodriguez, J.J. & Kokol, P., 2011.** Rotation of Random Forests for Genomic and Proteomic Classification Problems. In H. R. Arabnia & Q.-N. Tran, éd. *Software Tools and Algorithms for Biological Systems*. Advances in Experimental Medicine and Biology. Springer New York, p. 211-221.
- Sybesma, W., Hugenholtz, J., de Vos, W.M. & Smid, E.J., 2006.** Safe use of genetically modified lactic acid bacteria in food: Bridging the gap between consumers, green groups, and industry. *Electronic Journal of Biotechnology*, 9(4), p.0-0.

T

- Tachon, S., Brandsma, J.B. & Yvon, M., 2010.** NoxE NADH oxidase and the electron transport chain are responsible for the ability of *Lactococcus lactis* to decrease the redox potential of milk. *Applied and environmental microbiology*, 76(5), p.1311-1319.
- Tagg, J.R., Dajani, A.S. & Wannamaker, L.W., 1976.** Bacteriocins of gram-positive bacteria. *Bacteriological reviews*, 40(3), p.722-756.
- Taïbi, A., Dabour, N., Lamoureux, M., Roy, D. & LaPointe, G., 2011.** Comparative transcriptome analysis of *Lactococcus lactis* subsp. *cremoris* strains under conditions simulating Cheddar cheese manufacture. *International journal of food microbiology*, 146(3), p.263-275.
- Taïbi, A., Dabour, N., Lamoureux, M., Roy, D. & LaPointe, G., 2010.** Evaluation of the genetic polymorphism among *Lactococcus lactis* subsp. *cremoris* strains using comparative genomic hybridization and multilocus sequence analysis. *International Journal of Food Microbiology*, 144(1), p.20-28.
- Tailliez, P., Tremblay, J., Ehrlich, S.D. & Chopin, A., 1998.** Molecular diversity and relationship within *Lactococcus lactis*, as revealed by randomly amplified polymorphic DNA (RAPD). *Systematic and applied microbiology*, 21(4), p.530-538.
- Takahashi, S., Abbe, K. & Yamada, T., 1982.** Purification of pyruvate formate-lyase from *Streptococcus mutans* and its regulatory properties. *Journal of Bacteriology*, 149(3), p.1034-1040.
- Tamime, A.Y. & Robinson, R.K., 1999.** *Yoghurt: Science and Technology*, Woodhead Publishing.
- Tan-a-ram, P., Cardoso, T., Daveran-Mingot, M.-L., Kanchanatawee, S., Loubière, P., Girbal, L. & Coccagn-Bousquet, M., 2011.** Assessment of the Diversity of Dairy *Lactococcus lactis* subsp. *lactis* Isolates by an Integrated Approach Combining Phenotypic, Genomic, and Transcriptomic Analyses. *Applied and Environmental Microbiology*, 77(3), p.739-748.
- Tanous, C., Chambellon, E., Le Bars, D., Delespaul, G. & Yvon, M., 2006.** Glutamate Dehydrogenase Activity Can Be Transmitted Naturally to *Lactococcus lactis* Strains To Stimulate Amino Acid Conversion to Aroma Compounds. *Applied and Environmental Microbiology*, 72(2), p.1402-1409.
- Tanous, C., Gori, A., Rijnen, L., Chambellon, E. & Yvon, M., 2005.** Pathways for α -ketoglutarate formation by *Lactococcus lactis* and their role in amino acid catabolism. *International Dairy Journal*, 15(6-9), p.759-770.
- Tanous, C., Kieronczyk, A., Helinck, S., Chambellon, E. & Yvon, M., 2002.** Glutamate dehydrogenase activity: a major criterion for the selection of flavour-producing lactic acid bacteria strains. *Antonie van Leeuwenhoek*, 82(1-4), p.271-278.
- Taylor, C.F., Paton, N.W., Lilley, K.S., Binz, P.-A., Julian, R.K., Jr, Jones, A.R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E.W., Dunn, M.J., Heck, A.J.R., Leitner, A., Macht, M., Mann, M., Martens, L., Neubert, T.A., Patterson, S.D., Ping, P., Seymour, S.L., Souda, P., Tsugita, A., Vandekerckhove, J., Vondriska, T.M., Whitelegge, J.P., Wilkins, M.R., Xenarios, I., Yates, J.R., 3rd & Hermjakob, H., 2007.** The minimum information about a proteomics experiment (MIAPE). *Nature biotechnology*, 25(8), p.887-893.
- Ten Brink, B., Otto, R., Hansen, U.P. & Konings, W.N., 1985.** Energy recycling by lactate efflux in growing and nongrowing cells of *Streptococcus cremoris*. *Journal of Bacteriology*, 162(1), p.383-390.
- Tengerdy, R., 1961.** Redox potential changes in the 2-keto-l-gulononic acid fermentation. I correlation between redox potential and dissolved-oxygen concentration†. *Journal of Biochemical and Microbiological Technology and Engineering*, (3), p.241-253.
- Teng-Leary, E. & Kohlhaw, G.B., 1973.** Mechanism of feedback inhibition by leucine. Binding of leucine to wild-type and feedback resistant α -isopropylmalate synthases and its structural consequences. *Biochemistry*, 12(16), p.2980-2986.

- Teusink, B. & Smid, E.J., 2006.** Modelling strategies for the industrial exploitation of lactic acid bacteria. *Nature reviews. Microbiology*, 4(1), p.46-56.
- Thompson, J., 1987.** Sugar transport in the lactic acid bacteria. In *Sugar transport and metabolism in gram-positive bacteria*. Londond: Ellis Horwood, p. 13-38.
- Thompson, J., Turner, K.W. & Thomas, T.D., 1978.** Catabolite Inhibition and Sequential Metabolism of Sugars by *Streptococcus lactis*. *Journal of Bacteriology*, 133(3), p.1163-1174.
- Tibshirani, R., 2011.** Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), p.273-282.
- Torriani, S., Gardini, F., Guerzoni, M.E. & Dellaglio, F., 1996.** Use of response surface methodology to evaluate some variables affecting the growth and acidification characteristics of yoghurt cultures. *International Dairy Journal*, 6(6), p.625-636.
- Travers, A.A., 1988.** Bacterial RNA polymerase--the ultimate metabolic sensor? *BioEssays: news and reviews in molecular, cellular and developmental biology*, 8(6), p.190-193.
- Travers, A.A., 1984.** Conserved features of coordinately regulated *E. coli* promoters. *Nucleic acids research*, 12(6), p.2605-2618.
- Travers, A.A., 1980.** Promoter Sequence for Stringent Control of Bacterial Ribonucleic Acid Synthesis. *Journal of Bacteriology*, 141(2), p.973-976.
- Traxler, M.F., Summers, S.M., Nguyen, H.-T., Zacharia, V.M., Hightower, G.A., Smith, J.T. & Conway, T., 2008.** The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Molecular microbiology*, 68(5), p.1128-1148.
- Tsakalidou, E. & Kalantzopoulos, G., 1992.** Purification and partial characterization of an esterase from *Lactococcus lactis* ssp *lactis* strain ACA-DC 127. *Le Lait*, 72(6), p.533-543.
- Tynkkynen, S., Buist, G., Kunji, E., Kok, J., Poolman, B., Venema, G. & Haandrikman, A., 1993.** Genetic and biochemical characterization of the oligopeptide transport system of *Lactococcus lactis*. *Journal of bacteriology*, 175(23), p.7523-7532.

U & V

- Umbarger, H.E., 1978.** Amino Acid Biosynthesis and its Regulation. *Annual Review of Biochemistry*, 47(1), p.533-606.
- Urbach, G., 1995.** Contribution of lactic acid bacteria to flavour compound formation in dairy products. *International Dairy Journal*, 5(8), p.877-903.
- Urbach, G., 1997.** The flavour of milk and dairy products: II. Cheese: contribution of volatile compounds. *International Journal of Dairy Technology*, 50(3), p.79-89.
- Valence, F., Deutsch, S.M., Richoux, R., Gagnaire, V. & Lortal, S., 2000.** Autolysis and related proteolysis in Swiss cheese for two *Lactobacillus helveticus* strains. *The Journal of dairy research*, 67(2), p.261-271.
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr, Hieter, P., Vogelstein, B. & Kinzler, K.W., 1997.** Characterization of the yeast transcriptome. *Cell*, 88(2), p.243-251.
- Van de Ven, F.J., Van den Hooven, H.W., Konings, R.N. & Hilbers, C.W., 1991.** NMR studies of lantibiotics. The structure of nisin in aqueous solution. *European journal of biochemistry / FEBS*, 202(3), p.1181-1188.
- Viana, R., Monedero, V., Dossonnet, V., Vadeboncoeur, C., Pérez-Martínez, G. & Deutscher, J., 2000.** Enzyme I and HPr from *Lactobacillus casei*: their role in sugar transport, carbon catabolite repression and inducer exclusion. *Molecular Microbiology*, 36(3), p.570-584.
- Viguerie, N., Montastier, E., Maoret, J.-J., Roussel, B., Combes, M., Valle, C., Villa-Vialaneix, N., Iacovoni, J.S., Martinez, J.A., Holst, C., Astrup, A., Vidal, H., Clément, K., Hager, J., Saris, W.H.M. & Langin, D., 2012.** Determinants of human adipose tissue gene expression: impact of diet, sex, metabolic status, and cis genetic regulation. *PLoS genetics*, 8(9), p.e1002959.
- Visser, S., 1993.** Proteolytic Enzymes and Their Relation to Cheese Ripening and Flavor: An Overview. *Journal of Dairy Science*, 76(1), p.329-350.

- Visser, S., Exterkate, F.A., Slangen, C.J. & de Veer, G.J., 1986.** Comparative Study of Action of Cell Wall Proteinases from Various Strains of *Streptococcus cremoris* on Bovine alpha(s1)-, beta-, and kappa-Casein. *Applied and environmental microbiology*, 52(5), p.1162-1166.
- De Vos, W.M., Vos, P., de Haard, H. & Boerrigter, I., 1989.** Cloning and expression of the *Lactococcus lactis* subsp. *cremoris* SK11 gene encoding an extracellular serine proteinase. *Gene*, 85(1), p.169-176.
- De Vuyst, L. & Leroy, F., 2007.** Bacteriocins from lactic acid bacteria: production, purification, and food applications. *Journal of molecular microbiology and biotechnology*, 13(4), p.194-199.
- De Vuyst, L., Vanderveken, F., Van de Ven, S. & Degeest, B., 1998.** Production by and isolation of exopolysaccharides from *Streptococcus thermophilus* grown in a milk medium and evidence for their growth-associated biosynthesis. *Journal of applied microbiology*, 84(6), p.1059-1068.

W

- Wang, L. & Lutkenhaus, J., 1998.** FtsK is an essential cell division protein that is localized to the septum and induced as part of the SOS response. *Molecular Microbiology*, 29(3), p.731-740.
- Ward, J.H., 1963.** Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), p.236-244.
- Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L. & Humphery-Smith, I., 1995.** Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*, 16(7), p.1090-1094.
- Wegmann, U., O'Connell-Motherway, M., Zomer, A., Buist, G., Shearman, C., Canchaya, C., Ventura, M., Goesmann, A., Gasson, M.J., Kuipers, O.P., van Sinderen, D. & Kok, J., 2007.** Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. *Journal of bacteriology*, 189(8), p.3256-3270.
- Welker, M., 2011.** Proteomics for routine identification of microorganisms. *PROTEOMICS*, 11(15), p.3143-3153.
- Welman, A.D. & Maddox, I.S., 2003.** Exopolysaccharides from lactic acid bacteria: perspectives and challenges. *Trends in biotechnology*, 21(6), p.269-274.
- Wendrich, T.M., Blaha, G., Wilson, D.N., Marahiel, M.A. & Nierhaus, K.H., 2002.** Dissection of the mechanism for the stringent factor RelA. *Molecular cell*, 10(4), p.779-788.
- Wendrich, T.M. & Marahiel, M.A., 1997.** Cloning and characterization of a relA/spoT homologue from *Bacillus subtilis*. *Molecular microbiology*, 26(1), p.65-79.
- Whitfield, C. & Valvano, M.A., 1993.** Biosynthesis and expression of cell-surface polysaccharides in Gram-negative bacteria. In *Advances in Microbial Physiology*. San Diego: Academic Press, p. 135-231.
- Whitman, W.B., Coleman, D.C. & Wiebe, W.J., 1998.** Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences*, 95(12), p.6578-6583.
- Wiedemann, I., Benz, R. & Sahl, H.-G., 2004.** Lipid II-mediated pore formation by the peptide antibiotic nisin: a black lipid membrane study. *Journal of bacteriology*, 186(10), p.3259-3261.
- Wijesundera, C. & Drury, L., 1999.** Role of milk fat in production of cheddar cheese flavour using a fat-substituted cheese model. *Australian journal of dairy technology*, 54(1), p.28-35.
- Wolz, C., Geiger, T. & Goerke, C., 2010.** The synthesis and function of the alarmone (p)ppGpp in firmicutes. *International journal of medical microbiology: IJMM*, 300(2-3), p.142-147.
- Wouters, J.A., Frenkiel, H., de Vos, W.M., Kuipers, O.P. & Abee, T., 2001.** Cold shock proteins of *Lactococcus lactis* MG1363 are involved in cryoprotection and in the production of cold-induced proteins. *Applied and environmental microbiology*, 67(11), p.5171-5178.
- Wouters, J.T., Ayad, E.H., Hugenholtz, J. & Smit, G., 2002.** Microbes from raw milk for fermented dairy products. *International Dairy Journal*, 12(2-3), p.91-109.
- Wu, C.C. & MacCoss, M.J., 2002.** Shotgun proteomics: tools for the analysis of complex biological systems. *Current opinion in molecular therapeutics*, 4(3), p.242-250.

X, Y & Z

- Xu, Y. & Kong, J., 2013.** Construction and Potential Application of Controlled Autolytic Systems for *Lactobacillus casei* in Cheese Manufacture. *Journal of food protection*, 76(7), p.1187-1193.
- Yang, X. & Ishiguro, E.E., 2001.** Involvement of the N Terminus of Ribosomal Protein L11 in Regulation of the RelA Protein of *Escherichia coli*. *Journal of Bacteriology*, 183(22), p.6532-6537.
- Yates, J.R., Ruse, C.I. & Nakorchevsky, A., 2009.** Proteomics by mass spectrometry: approaches, advances, and applications. *Annual review of biomedical engineering*, 11, p.49-79.
- Ye, J.J., Reizer, J., Cui, X. & Saier, M.H., Jr, 1994.** Inhibition of the phosphoenolpyruvate:lactose phosphotransferase system and activation of a cytoplasmic sugar-phosphate phosphatase in *Lactococcus lactis* by ATP-dependent metabolite-activated phosphorylation of serine 46 in the phosphocarrier protein HPr. *The Journal of biological chemistry*, 269(16), p.11837-11844.
- Yeung, K.Y. & Ruzzo, W.L., 2001.** Principal component analysis for clustering gene expression data. *Bioinformatics (Oxford, England)*, 17(9), p.763-774.
- Yu, J., Du, X., Wang, W., Zhang, J., Liu, W., Sun, Z., Sun, T. & Zhang, H., 2011.** Phenotypic and genotypic characteristics of lactic acid bacteria isolated from sour congee in Inner Mongolia of China. *The Journal of general and applied microbiology*, 57(4), p.197-206.
- Yvon, M., Berthelot, S. & Gripon, J.C., 1998.** Adding α -Ketoglutarate to Semi-hard Cheese Curd Highly Enhances the Conversion of Amino acids to Aroma Compounds. *International Dairy Journal*, 8(10-11), p.889-898.
- Yvon, M., Chambellon, E. & Gori, A., 2003.** Use of Lactic Acid Bacteria for Producing Alpha-Ketoglutarate.
- Yvon, M., Tanous, C., Chambellon, E. & Nardi, M., 2006.** Plasmid Comprising an Encoding Gene for a Glutamate Deshydrogenase of *Lactococcus Lactis*.
- Yvon, M., Thirouin, S., Rijnen, L., Fromentier, D. & Gripon, J.C., 1997.** An aminotransferase from *Lactococcus lactis* initiates conversion of amino acids to cheese flavor compounds. *Applied and environmental microbiology*, 63(2), p.414-419.
- Zalkin, H. & Dixon, J.E., 1992.** De novo purine nucleotide biosynthesis. *Progress in nucleic acid research and molecular biology*, 42, p.259-287.
- Zhang, Yanhe, Zhang, Yanping, Zhu, Y., Mao, S. & Li, Y., 2010.** Proteomic Analyses To Reveal the Protective Role of Glutathione in Resistance of *Lactococcus lactis* to Osmotic Stress. *Applied and Environmental Microbiology*, 76(10), p.3177-3186.
- Zomer, A.L., Buist, G., Larsen, R., Kok, J. & Kuipers, O.P., 2007.** Time-resolved determination of the CcpA regulon of *Lactococcus lactis* subsp. *cremoris* MG1363. *Journal of bacteriology*, 189(4), p.1366-1381.
- Zou, H., Hastie, T. & Tibshirani, R., 2004.** Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, 15, p.2006.
- Zuo, Y., Wang, Y. & Steitz, T.A., 2013.** The Mechanism of *E. coli* RNA Polymerase Regulation by ppGpp Is Suggested by the Structure of their Complex. *Molecular Cell*, 50(3), p.430-436.

TABLE DES ILLUSTRATIONS

ETUDE BIBLIOGRAPHIQUE

- Figure 1 : Représentation schématique de l'action déstabilisatrice de l'acidification sur le phosphate de calcium micellaire à l'origine de la coagulation du lait [Branger et al. 2007].**..... - 11 -
- Figure 2 : Arbre phylogénétique des principaux genre du groupe des LAB basé sur la concaténation des séquences protéiques ribosomales (bootstrap>75 %) [Makarova et al. 2006].** Les différentes espèces sont regroupées par un code couleur en fonction de leur famille d'appartenance indiquée en gras. - 12 -
- Tableau 1 : Récapitulatif des *Lactococcus lactis* complètement séquencés, disponible sur le site NCBI.**..... - 14 -
- Figure 3 : Méthodologie et applications de l'approche MLST [Passerini 2011].** A) Analyse des séquences de fragments internes de 5 gènes des souches A, B et C, permet de leur attribuer un profil allélique (ou Séquence Type, ST) sous forme d'un code à 5 chiffres. Le typage, regroupement des souches selon leur ST, permet de visualiser l'évolution récente d'une population. La concaténation des 5 STs permet d'analyser la phylogénie, repositionne dans le temps les événements de séparation. B) Représentations d'un typage et de la phylogénie. - 16 -
- Figure 4 : Phylogénie de la sous-espèce *Lactococcus lactis* ssp. *lactis* par l'approche MLST [Passerini 2011].** A) Le positionnement phylogénétique des souches avec respectivement en rouge et en vert, la représentation des deux écotypes « domestiqué » et « environnemental ». B) Analyse eBurst des 54 souches ssp. *lactis* organisées en 4 complexes clonaux (CC1 à CC4), avec indiqué par un cercle coloré l'origine des souches. - 16 -
- Figure 5 : Vue globale des voies métaboliques de transport et de catabolisme des sucres et de l'acide citrique chez *Lactococcus lactis* ssp. *lactis*.** Il est indiqué pour chaque réaction les enzymes qui catalysent ces voies ainsi que les gènes impliqués entre parenthèses. Les symboles ronds grisés représentent des perméases, tandis que les canaux grisés les enzymes membranaires du PTS. Les métabolites encadrés appartiennent à la glycolyse. EMP : Embden-Meyerhoff-Parnas. [Neves et al. 2005; Coccagn-Bousquet et al. 2002]. - 19 -
- Figure 6 : Système Phosphoenolpyruvate-phosphotransférase (PTS) de transport des sucres et positionnement des mécanismes de régulation de type biochimique et génétique.** Cas d'un système protéasique à 4 protéines au sein de la perméase II (A, B, C, D) avec les domaines EII-B et C cytoplasmiques. CREs : catabolite response element..... - 20 -
- Figure 7 : Nutrition carbonée en lait : voie d'entrée des carbohydrates (lactose, galactose, glucose et citrate), glycolyse, et métabolisme du citrate.** Il est indiqué pour chaque réaction le nom des enzymes qui catalysent ces voies, leur abréviation ainsi que les gènes impliqués. Les intermédiaires glycolytiques apparaissent encadrés et le principal produit fermentaire est sous forme d'ellipse. Les intermédiaires glycolytiques en rouge et vert jouent un rôle dans la régulation du métabolisme. - 24 -
- Tableau 2 : Propriétés chimiques et organoleptiques des quatre composés carbonés clefs issus du pyruvate dans les laits fermentés.** ¹: [Corrieu 2008]. ND : Non déterminé..... - 26 -
- Figure 8 : Métabolisme du pyruvate et rôle majeur du ratio NADH/NAD⁺ dans le contrôle du shift homolactique vers un métabolisme mixte pour la production d'arômes.** Il est indiqué pour chaque réaction les enzymes qui catalysent ces voies ainsi que les gènes impliqués. Les carrés représentent les intermédiaires réactionnels, et les ellipses les produits fermentaires. Sur fond violet, les composés responsables d'arômes dans les produits laitiers et en orange les non aromatiques. Les flèches en pointillé matérialisent les actions régulatrices d'intermédiaires glycolytiques (en vert les actions inhibitrices et en

rouge les actions activatrices). DCO : décarboxylation oxydative et DCNO : décarboxylation non oxydative. Les abréviations des enzymes de la glycolyse sont détaillées Figure 7.	- 28 -
Figure 9 : Représentation schématique du système protéolytique en lait et de nutrition azotée chez <i>Lactococcus lactis</i>, et des voies de régulation de l'expression génique, ainsi que du devenir des acides aminés. Adapté de [Guédon et al. 2001; Savijoki et al. 2006]. Les peptidases sont classées en quatre familles sur la base de leur spécificité de clivage : en rouge les endopeptidases, en bleu les aminopeptidases, en orange les proline peptidases, et en cyan les tri/di-peptidases. En bleu, les trois voies principales de devenir des acides aminés : anabolisme (synthèse protéique), formation d'énergie et production de molécules aromatiques. En vert, la régulation de l'expression génique du système protéolytique par CodY un régulateur transcriptionnel.	- 32 -
Tableau 3 : Systèmes de transport d'acides aminés identifiés chez les bactéries lactiques [Corrieu 2008].	- 34 -
Figure 10 : Représentation schématisée du métabolisme de l'arginine chez <i>Lactococcus lactis</i>. Adapté de [Larsen et al. 2004]. Les enzymes ainsi que les noms de gènes (entre parenthèse) sont indiqués pour chaque réaction catalysée.	- 36 -
Tableau 4 : Notes aromatiques et seuils de perception des molécules aromatiques dérivées des acides aminés [Corrieu 2008]. (A. : Acide)	- 38 -
Figure 11 : Principales voies de dégradation des acides aminés chez <i>Lactococcus lactis</i>. Adapté de [Yvon & Rijnen 2001]. Les enzymes ainsi que les noms de gènes (entre parenthèse) sont indiqués en gris. Liste des abréviations : DCO : Décarboxylation oxydative, CBL/CGL : Cystathionine-β-lyases/Cystathionine-γ-lyases, TA : Thréonine aldolase, TIL : Tryptophane indole lyase, TPL : Tyrosine phénol lyase, DMDS : Diméthyl-disulfide, DMTS : Diméthyl-trisulfide.....	- 40 -
Figure 12 : Interconnexion des voies du métabolisme du pyruvate, de production d'acétoïne/butanediol et de la régulation de la biosynthèse des acides aminés branchés. Adapté de [Goupil-Feuillerat et al. 1997]. Sur fond violet, les composés responsables d'arômes dans les produits laitiers, et en orange les non aromatiques. L'opéron <i>leu-ilv</i> est représenté dans la partie basse de la figure. Les flèches en pointillé matérialisent les actions régulatrices transcriptionnelles du régulateur CodY, et les flèches en tirets-points les régulations allostériques (en vert inhibition et en rouge activation).	- 44 -
Figure 13 : Schéma général de la dégradation de la matière grasse du lait et du catabolisme des acides gras libres. Adapté et modifié d'après [McSweeney & Sousa 2000; Cheng 2010].	- 51 -
Tableau 5 : Notes aromatiques et seuils de perception des principales molécules aromatiques dérivées de la dégradation de la matière grasse retrouvées dans les laits fermentés. [Ott et al. 1997; McSweeney & Sousa 2000; Cheng 2010]. (A. : Acide)	- 52 -
Figure 14 : Voies de synthèse <i>de novo</i> des purines (à gauche) et des pyrimidines (à droite).	- 56 -
Figure 15 : Régulation de l'activité de l'acidification chez <i>Lactococcus lactis</i> dans des conditions de culture nutritionnellement non limitantes. [Raynaud 2006].....	- 61 -
Figure 16 : Illustration des principaux facteurs intervenants dans le phénomène d'autolyse des souches, et des conséquences de la libération des constituants cellulaires.	- 66 -
Figure 17 : Représentation de la démarche d'analyse de biologie des systèmes appliquée à la sélection de souches.	- 71 -

Figure 18 : Représentation graphique du processus de traitement et d'analyse des données. A : Vérification de la reproductibilité des données (Nuage de points). B : Analyse de la distribution des données et traitement de normalisation (histogramme). C : Analyse descriptive unidimensionnelle (Boîtes à moustaches). D : Analyse exploratoire multidimensionnelle (Biplot d'une Analyse en Composantes Principales)..... - 76 -

Figure 19 : Diagramme décisionnel pour le choix de la méthode multidimensionnelle à appliquer. Les méthodes précédées d'un carré seront développées dans ce manuscrit..... - 76 -

Figure 20: Principe de l'ACP. Réduction de la dimension du jeu de données de $n=4$ par construction de deux Composantes Principales (CP) maximisant l'inertie, et projection des objets (A) et des variables (B) dans le nouvel espace factoriel défini par les CPs. B. Cercle des corrélations (entre +1 et -1 si données centrées et réduites) qui indique la corrélation entre la variable et la CP..... - 77 -

MATERIELS & METHODES

Tableau 6 : Caractéristiques des souches utilisées dans cette étude..... - 90 -

Tableau 7 : Composition de la poudre de lait écrémée (données Yoplait). - 91 -

Tableau 8 : Composition en matière protéique et grasse des 10 laits employés dans cette étude..... - 91 -

Tableau 9 : Composition du milieu extrait de levure, YE en $g.L^{-1}$ - 92 -

Tableau 10 : Composition des milieux de culture chimiquement définis, en $g.L^{-1}$. Concentrations modifiées pour les cultures à pH libre ou précultures (). En gras les concentrations du MCD de référence modifiées. - 93 -

Tableau 11 : Temps de rétention, ions sélectionnés en SIM et étalon interne pour chaque composé identifié. ^a En gras les ions utilisés pour la quantification, les autres sont employés pour l'identification. - 101 -

Figure 21 : Diagramme méthodologique des différentes étapes d'un plan de surface de réponses..... - 108 -

Tableau 12 : Organisation du plan de Box-Behnken randomisé en valeurs codées pour les 4 facteurs et 3 niveaux sélectionnés et adaptés pour l'étude. Les expériences en gras correspondent au point au centre.- 111 -

Tableau 13 : Tableau d'analyse de variance du modèle (ANOVA)..... - 113 -

RÉSULTATS

Chapitre 1 :

Tableau I. 1 : Conditions de culture pour la culture discontinue de la souche *L. lactis ND5F* au laboratoire. - 123 -

Figure I. 1 : Cinétique de croissance et d'acidification à pH libre de *L. lactis ND5F* sur lait écrémé dans la condition de référence au laboratoire. A) Profil moyen de la concentration cellulaire (UFC.mL⁻¹) obtenu par le lissage de la moyenne des points expérimentaux de 4 réplicats biologiques. Evaluation de la reproductibilité des données par la représentation sur chaque dénombrement moyen (○) de l'écart-type expérimental. Chaque phase exponentielle de croissance est modélisée par régression linéaire (Equation, coefficient de détermination R² et coefficient de variation, CV). B) Comparatif de la cinétique de pH entre le procédé industriel et au laboratoire et cinétique moyenne de production de lactate (lactate+ acide lactique sous forme non dissociée) en mM. Ecarts-types établis à partir des données de 4 réplicats biologiques pour les données du laboratoire..... - 124 -

Tableau I. 2 : Comparaison des descripteurs de l'activité acidifiante de *L. lactis* ND5F à pH libre sur lait en culture discontinue en conditions industrielles et en procédé à l'échelle du laboratoire. Ecarts-types établis à partir des données de 3 réplicats biologiques (données industrielles) et 4 données laboratoires. ND : Nondéterminé..... - 126 -

Figure I. 2 : Profils moyens de décroissance de la pression partielle en oxygène dissout PO_2 (histogramme), du pH, du potentiel d'oxydoréduction Eh_7 et cinétique de production d'acide lactique pendant la culture discontinue de la souche *L. lactis* ND5F sur lait écrémé à pH libre. Ecarts-types établis à partir des données de quatre réplicats biologiques. - 128 -

Figure I. 3 : Cinétiques des concentrations cumulées en substrats (lactose et citrate) ainsi que des métabolites produits (lactate, acétate et arômes : diacétyle et acétoïne) et de la biomasse au cours de la culture discontinue de la souche *L. lactis* ND5F sur lait écrémé à pH libre. Concentrations exprimées en $g.L^{-1}$. Ecarts-types établis à partir des données de quatre réplicats biologiques. - 130 -

Figure I. 4 : Vitesses spécifiques moyennes de production de lactate, $v_{lactate}$, et de consommation de lactose, $v_{lactose}$, au cours de la culture discontinue de la souche *L. lactis* ND5F sur lait écrémé à pH libre. Profil moyen à partir de 4 réplicats biologiques..... - 132 -

Tableau I. 3 : Principaux descripteurs caractérisant le comportement macro-cinétique de croissance, d'acidification, et de production de métabolites de ND5F à pH libre sur lait écrémé. Segmentation de la culture en trois phases : croissance, arrêt de croissance et à l'arrêt de post-acidification à 12 h. Les résultats sont présentés avec leurs écarts-types issus de l'intégration de 4 réplicats biologiques à l'exception (*) des taux de croissance indiqués avec leur erreur sur la pente. ND : Non déterminé..... - 133 -

Tableau I. 4 : Détermination des concentrations moyennes en acides aminés libres dans le lait avant inoculation et lors de la culture de la souche *L. lactis* ND5F à pH libre sur lait écrémé, en $mg.L^{-1}$. Moyenne à partir de 4 réplicats biologiques. (D: composé détecté mais non quantifié ($< 1 mg.L^{-1}$), ... - 134 -

Figure I. 5 : Effet inhibiteur de l'acide lactique (total et non dissocié) sur le taux de croissance maximal en milieu YE à $50 g.L^{-1}$ de lactose et à pH régulé à 6,0. Un modèle linéaire a été établi pour chaque forme et la concentration inhibitrice médiane de l'acide lactique sous forme non dissociée, $IC_{50_{Ac. lactique}}$, a été déterminée. Chaque taux de croissance est le taux moyen issu de duplicats biologiques..... - 137 -

Tableau I. 5 : Comparatif des principaux critères cinétiques de caractérisation de l'activité de croissance et de production de lactate de ND5F en phase exponentielle de croissance entre le milieu YE et lait écrémé sans régulation de pH, en micro-aérobiose. Les données sur milieu YE sont issues de la moyenne de deux réplicats biologiques. Les résultats sont présentés avec leurs écarts-types issus de l'intégration de quatre réplicats biologiques à l'exception (*) des taux de croissances indiqués avec leur erreur sur la pente. Détermination de la significativité des valeurs par t-Test : p-value $< 0,05$ (*), $< 0,01$ (**) et $< 0,001$ (***), la référence est la culture en lait écrémé. - 138 -

Tableau I. 6 : Analyse macro-cinétique du comportement de ND5F à différents pH régulés sur lait écrémé. Les principaux paramètres divergeants sont surlignés en gris. Détermination de la significativité des valeurs par t-Test : p-value $< 0,05$ (*), $< 0,01$ (**) et $< 0,001$ (***), la référence est la culture à pH 6,5. - 140 -

Figure I. 6 : Comparatifs des cinétiques de croissance et d'acidification (A) ainsi que de consommation d'oxygène et d'activité réductrice (B) en fonction du traitement thermique appliqué sur la matrice lait lors de la culture à pH libre de ND5F. Ecart-types issus de 4 réplicats biologique pour le lait classique. - 146 -

Tableau I. 8 : Détermination des concentrations en acides aminés libres dans le lait avant inoculation en fonction du traitement thermique appliqué. D: composé détecté mais non quantifié ($< 1 \text{ mg.L}^{-1}$), ND : Non détecté. - 148 -

Tableau I. 9 : Comparaison de la croissance de *ND5F* en fonction des différentes catégories nutritionnelles complétant le lait écrémé à pH régulé à 6. Le détail de la légende est présenté Tableau I. 3. Détermination de la significativité des valeurs par t-Test : p-value $< 0,05$ (*), $< 0,01$ (**) et $< 0,001$ (***), la référence est la culture en lait classique. - 150 -

Chapitre 2 :

Figure II. 1 : Positionnement de la souche *ND5F* sur l'arbre phylogénétique des souches de *L. lactis* ssp. *lactis* de la collection Genoferment [Passerini et al. 2010]. La flèche noire indique la position de la souche *ND5F*. G1 et G2, les deux groupes déterminés au sein du clade correspondant à la population de souches dites « domestiquées ». Arbre phylogénétique réalisé selon la méthode Neighbor Joining (bootstrap 1000).- 158 -

Figure II. 2 : PFGE de l'ADN de *ND5F* obtenue après digestion par NotI-NotI/ICeuI et ICeuI. Taille en kb. - 1 -

Tableau II. 1 : Estimation de la taille des fragments de *ND5F* obtenus après digestion par NotI-NotI/ICeuI et ICeuI. X* : Taille du fragment indéterminée sur le gel de PFGE. En gras, la taille totale en kb. - 1 -

Figure II. 3 : Recherche de l'opéron chromosomique et du plasmide impliqués dans l'utilisation du citrate chez la souche *ND5F*. A. Schéma des gènes présents dans l'opéron chromosomique. B. Amplification de l'opéron chromosomique (région *citM-citG*). C. Amplification du gène plasmidique *citP*. - 161 -

Tableau II. 2 : Récapitulatif des dix-neuf gènes chromosomiques attribués à des fonctions métaboliques spécifiques à *ND5F* après comparaison avec la souche IL 1403. - 164 -

Tableau II. 3 : Nombre de gènes spécifiques à *ND5F* après comparaison avec les autres souches domestiquées. - 165 -

Figure II. 4 : Position des huit prophages et trois ICEs contenant les gènes spécifiques de *ND5F* sur le chromosome. - 167 -

Figure II. 5 : Classification ascendante hiérarchique (CAH) génotypique intégrative de *ND5F* (flèche noire) et des neuf autres souches domestiquées, à partir des données de phylogénie génétique (MLST) et génomique (PFGE). De gauche à droite : le dendrogramme, le nom des souches, le Séquence Type (ST) et le Complexe Clonal (CC). Les valeurs de bootstrap au dessus de 0,50 sont indiquées pour chaque nœud. La ligne verticale en pointillé matérialise le niveau de coupure de l'arbre pour interprétation. - 172 -

Figure II. 6 : Analyse en composantes principales (ACP) et classification ascendante hiérarchique (CAH) de l'échantillon de neuf souches domestiquées de *L. lactis* ssp. *lactis* ainsi que de la souche *ND5F* à partir des 82 données phénotypiques. Représentation des 10 souches selon les composantes principales (CP) 1 et 2 (A) et 1 et 3 (B). Chaque losange de la même couleur matérialise un réplicat biologique de la même souche (6 au total par souche). Chaque ellipse colorée représente l'intervalle de confiance à 95 % de la distribution de chaque souche. C) Dendrogramme des souches obtenu par classification ascendante hiérarchique (distance euclidienne et critère de Ward). La localisation spécifique de *ND5F* est repérée par un carré rouge pour faciliter la visualisation. - 173 -

- Figure II. 7 : Dendrogramme obtenu après Classification Ascendante Hiérarchique (CAH) de l'échantillon de neuf souches domestiquées de *L. lactis* ssp. *lactis* ainsi que de la souche *ND5F* (flèche noire) à partir des 20 variables les plus contributives, les COVs. Représentation des 4 principaux groupes identifiés avec notification de la valeur de bootstrap calculée pour chaque sous-groupe..... - 175 -**
- Figure II. 8 : Analyse en composantes principales (ACP) des cinq souches du cluster IV dont *ND5F* à partir des 82 données phénotypiques. Représentations respectives des souches sur les dimensions 1 et 2 (A), et 3 et 4 (C). Chaque losange de la même couleur matérialise un réplicat biologique de la même souche. Représentations des variables les plus contributives (seuil à 75 %) pour expliquer cette distribution de souches sur les dimensions 1 et 2 (B) et 3 et 4 (D). Abbréviation : μ : taux de croissance, DMS : Diméthylsulfide, DMDS : Diméthyltrisulfide et DMTS : Diméthyltrisulfide..... - 176 -**
- Figure II. 9 : Analyse PLS-DA de la classification en deux groupes « *ND5F* » et « Autres souches » à partir des 82 variables phénotypiques. A : Evolution du taux d'erreur de classement des souches en fonction du nombre de métabolites sélectionnés pour les deux premières dimensions prises en compte. Représentations des individus (B) et des variables (C) de la PLS-DA sur les composantes 1 et 2 de la PLS. Le cercle interne sur la représentation des variables matérialise un coefficient de corrélation de 0,50 avec les composantes PLS..... - 178 -**
- Figure II. 10 : Observation au microscope à fluorescence de la proportion de cellules lysées de *L. lactis* ssp. *lactis* *ND5F* sur un échantillon de lait fermenté à 14 J. Observation à l'objectif x100 d'un état frais après double coloration DAPI/Iodure de propidium (IP). A : Coloration des cellules totales en bleue par la fluorescence émise par le DAPI après fixation sur les acides nucléiques. B : Coloration des cellules lysées en rouge par la fluorescence émise par l'IP. Après comparaison des deux images, les cellules non lysées sont indiquées sur la figure A par des flèches blanches. La barre blanche indique l'échelle d'observation à 20 μ m..... - 180 -**
- Figure II. 11 : Analyse PLS-DA de la classification en deux groupes « *ND5F* » et les « Autres souches » à partir des 14 variables sélectionnées comme les plus discriminantes. Représentations respectives des individus (A) et des variables (B) sur les composantes 1 et 2. L'ellipse en pointillés matérialise la zone de répartition des souches du groupe « Autre souches ». DMS : Diméthylsulfide. - 181 -**
- Tableau II. 4: Présentation des variables sélectionnées comme discriminantes pour *ND5F* face aux autres souches..... - 181 -**
- Tableau II. 5 : Caractéristiques des déterminants phénotypiques sélectionnés par rapport à la signature de *ND5F*. + : corrélation positive, - : corrélation négative. - 182 -**
- Figure II. 12: Analyse PLS-DA de la typicité de la souche *S87* à partir des données des 14 variables phénotypiques sélectionnées sur la base d'une classification en deux groupes « *S87* » et « Autres souches ». Représentations respectives des individus (A) et des variables (B) sur les composantes 1 et 2. DMS : Diméthylsulfide. - 183 -**
- Tableau II. 6 : Etude comparative des résultats de PLS-DA sur la typicité relative des 10 souches de l'échantillon à partir des 14 variables. - 183 -**
- Figure II. 13 : Analyse en composantes principales (ACP) de l'échantillon de neuf souches domestiquées de *L. lactis* ssp. *lactis* ainsi que de la souche *ND5F* à partir des données génotypiques et phénotypiques. Représentation des 10 souches selon les composantes principales 1 et 2. Chaque losange de la même couleur matérialise un réplicat biologique de la même souche (6 au total par souche). Chaque ellipse colorée représente l'intervalle de confiance à 99 % de la distribution de chaque souche. La localisation spécifique de *ND5F* est repérée par un carré rouge pour faciliter la visualisation..... - 184 -**

Figure II. 14 : Analyse en Composantes Principales (ACP) à partir des données génotypiques et phénotypiques avec apposition des résultats de classification sur les variables les plus contributives (seuil à 75 %).
Analyses sur les neuf souches de l'échantillon (A) ou avec l'ajout de ND5F (B)..... - 185 -

Chapitre 3 :

Figure III. 1: Diagrammes de Venn comparatifs entre les neuf protéomes de *L. lactis* ssp. *lactis* de cette étude, le protéome synthétique de la souche *L. lactis* ssp. *lactis* IL1403 et les autres protéomes complets de *L. lactis* publiés. Comparaisons avec A) le protéome synthétique de *L. lactis* ssp. *lactis* IL1403 (N Accession NCBI NC_002662.1), B) le protéome acide (pH 4-7) de *L. lactis* ssp. *cremoris* NCDO763 en lait [Gitton et al. 2005], C) le protéome acide (pH 4-7) de *L. lactis* ssp. *lactis* IL1403 sur M17-lactose et M17-glucose [Guillot et al. 2003], D) le protéome acide (pH 4.5-5.5 et 5-6) de *L. lactis* ssp. *lactis* IL1403 sur MCD en carence d'isoleucine [Dressaire et al. 2011] et E) le protéome basique (pH 9-12) de *L. lactis* ssp. *lactis* IL1403 sur milieu chimiquement défini SA [Drews et al. 2004]. - 192 -

Tableau III. 1 : Examen des catégories fonctionnelles et sous-catégories significativement sur ou sous-représentées entre les protéines identifiées et le protéome synthétique de *L. lactis* ssp. *lactis* IL1403. Seules les sous-catégories différentiellement exprimées sont présentées. Code p-value : * < 0,01, ** < 0,001 et *** < 1.10⁻⁴. - 196 -

Figure III. 2 : Distribution des quarante-sept protéines discriminantes sélectionnées en fonction de leurs catégories fonctionnelles. Voir Tableau III.1 pour détail des abbréviations des catégories fonctionnelles. P-value d'enrichissement calculée pour chaque catégorie fonctionnelle. Code p-value : * < 0,05, ** < 0,01 et *** < 1.10⁻³. - 201 -

Tableau III. 2 : Analyse fonctionnelle des quarante-sept protéines discriminantes sélectionnées pour les neuf protéomes. Organisation des protéines en quatre classes selon les profils d'expression après classification. Il est indiqué pour chaque protéine le rang relatif minimal, médian et maximal au sein des neuf protéomes..... - 202 -

Figure III. 3 : Dendrogramme des neuf protéomes de *L. lactis* ssp. *lactis* après croissance en lait et prélèvement à pH 4,60. La distance euclidienne et le critère de Ward ont été utilisés pour cette classification sur la base des 47 protéines pré sélectionnées. De gauche à droite, le dendrogramme des souches, le nom des vingt-sept expériences (neuf souches avec trois réplicats de A à C), STs de la MLST, présence du gène *citP* (losange noir), et numéro des quatre groupes phénotypiques (classe I à IV) mis en évidence. Les carrés de A à C correspondent aux trois sous-groupes de souches sur la base des similarités des protéomes. Les valeurs de bootstrap sont indiquées pour chaque nœud (n=1000)..... - 205 -

Figure III. 4 : Analyse en composante principale de la diversité des neuf protéomes de *L. lactis* ssp. *lactis* après croissance en lait et prélèvement à pH 4,60. Représentation des individus sur les dimensions 1 et 2 (A) et 1 et 3 (B), avec mention des trois sous-groupes de souches A-B-C. Les 47 protéines sont représentées par leur nom de gènes selon les dimensions 1 et 2 (C) et 1 et 3 (D), et des ellipses en pointillé matérialisent les différentes classes de variables Protéo I, II-III-IV. Les ellipses de couleur désignent les intervalles de confiance à 95 % autour des différents réplicats des souches. - 207 -

Figure III. 5 : Classification intégrative du génotype au phénotype de la singularité des souches de *L. lactis* ssp. *lactis* après croissance en lait à partir des marqueurs prédictifs identifiés. La distance euclidienne et le critère de Ward ont été utilisés pour cette classification sur la base des 20 distances génétiques (10 distances de phylogénie par MLST et 10 distances génomiques par empreintes génomiques), 20 COVs et 47 protéines pré sélectionnés. De gauche à droite, le dendrogramme des souches, le nom des 54 expériences (neuf souches avec six réplicats de A à F, données protéiques complétées par la moyenne),

STs de la MLST, présence du gène *citP* (losange noir), et numéro des quatre groupes phénotypiques (I à IV) et des trois protéiques (A, B et C). Les valeurs de bootstrap sont indiquées pour chaque nœud (n=1000). La ligne verticale en pointillé décrit le niveau de coupe du dendrogramme pour interprétation. - 209 -

Tableau III. 3 : Liste des trente-neuf variables phénotypiques sélectionnées lors de la recherche des marqueurs prédictifs de la diversité fonctionnelle à l'échelle intégrative. En gras et en italique, les variables déjà sélectionnées lors de l'analyse phénotypique seule. - 212 -

Tableau III. 4 : Liste des trente-quatre protéines sélectionnées lors de la recherche des marqueurs prédictifs de la diversité fonctionnelle à l'échelle intégrative. En gras et en italique, les protéines déjà sélectionnées lors de l'analyse protéique de la diversité. - 213 -

Figure III. 6 : Classification intégrative du génotype au phénotype des souches de *L. lactis* ssp. *lactis* après croissance en lait d'après les quatre groupes phénotypiques observés. La distance euclidienne et le critère de Ward ont été utilisés pour cette classification sur la base des 8 distances génétiques, 39 indicateurs phénotypiques et 34 protéines pré sélectionnés. De gauche à droite, le dendrogramme des souches, le nom des 54 expériences (neuf souches avec six réplicats de A à F, données protéiques complétées par la moyenne), STs de la MLST, présence du gène *citP* (losange noir), et numéro des quatre groupes phénotypiques (I à IV) et des trois groupes protéiques (A, B et C). Les valeurs de bootstrap sont indiquées pour chaque nœud (n=1000). La ligne verticale en pointillé décrit le niveau de coupe du dendrogramme pour interprétation..... - 215 -

Tableau III. 5 : Examen des catégories fonctionnelles différenciellement sur ou sous-représentées pour les comparaisons entre le protéome de la souche *ND5F-Danisco* et les 10 autres souches de l'étude. Pour chaque comparaison, le nombre de protéines différenciellement (sur/sous)-exprimées pour la souche considérée dans chaque catégorie fonctionnelle est mentionné (FDR $\leq 10\%$) (voir Erreur ! Source du renvoi introuvable. détails des catégories). Les catégories significativement enrichies sont désignées en fonction de la *p*-value d'enrichissement selon le code suivant : * $<0,1$ ** $<0,05$ et *** $<1.10^{-2}$ - 218 -

Tableau III. 6 : Protéines significativement sur-exprimées dans le protéome de *ND5F-Danisco* par rapport aux autres souches. - 219 -

Figure III. 7 : Comparaison des différents jeux de protéines discriminantes sélectionnées en fonction du nombre de souches à classer. A) Diagramme de Venn pour comparer les protéines discriminantes identifiées sans et avec ajout de *ND5F*. B) Recherche des corrélations entre les protéines spécifiques à la sélection de *ND5F* et celles sans *ND5F* : exemple de corrélation entre deux protéines (*YwjA* sans *ND5F* et *GmK* avec *ND5F*) . C) Analyse du positionnement de *ND5F* sur les protéines spécifiques de la classification avec *ND5F* : exemple de deux protéines avec et sans *ND5F* D) Analyse des corrélations entre les protéines spécifiques à la sélection sans *ND5F*. Note : les quantités relatives ont été centrées et réduites pour s'affranchir de l'abondance des protéines et comparer les profils d'expression. - 221 -

Tableau III. 7 : Analyse fonctionnelle des vingt-quatre protéines spécifiques de la classification supervisée des onze protéomes dont *ND5F*. Organisation des protéines en quatre classes selon les profils d'expression après classification ascendante hiérarchique. Il est indiqué pour chaque protéine le rang relatif minimal, médian et maximal au sein des 11 protéomes, et le rang médian spécifique pour *ND5F* (moyenne des biovariants *ND5F-Danisco* et *Lab*). Les protéines différentielles surexprimées sont surlignées par une bande grise. - 222 -

Figure III. 8 : Analyse en composantes principales et classification ascendante de la singularité protéique de la souche *L. lactis* ssp. *lactis* *ND5F* par rapport aux neuf autres levains à partir des cinquante-quatre protéines discriminantes sélectionnées. A) La distance euclidienne et le critère de Ward ont été utilisés pour cette classification des souches. Les valeurs de bootstrap sont indiquées pour chaque nœud

(n=1000). B) Représentation des individus sur les dimensions 1 et 2 (A) avec mention des quatre sous-groupes de souches A-B-C-D. Les ellipses de couleur désignent les intervalles de confiance à 95 % autour des différents réplicats des souches. C) Les 54 protéines sont représentées par leur nom de gènes selon les dimensions 1 et 2, les ellipses en pointillé matérialisent les différentes classes Protéo-ND-I, II-III-IV.....

..... - 225 -

Chapitre 4 :

Figure IV. 1 : Interconnexion des voies du métabolisme du pyruvate, de production d'acétoïne/butanediol, de régulation de la biosynthèse des acides aminés branchés et de production de pentane-2,3-dione. Les flèches en pointillés matérialisent les actions régulatrices transcriptionnelles du régulateur CodY, et les flèches tirets-points les régulations allostériques (en vert inhibition et en rouge activation)..... - 232 -

Tableau IV. 1: Comparaison des séquences protéiques des gènes de biosynthèse des acides aminés branchés des souches *NDCO 2118* et *ND5F*. - 234 -

Tableau IV. 2 : Détail des conditions expérimentales (1 à 6) des essais sur MCD de production différentielle de pentane-2,3-dione. (- : Pas de modification) - 239 -

Figure IV. 2 : Evolution de la biomasse et du taux de croissance de *L. lactis* ssp. *lactisND5F* au cours de la culture discontinue à pH libre dans différents MCD. Les données courbes sont obtenues par lissage. Les écart-types sont issus de trois réplicats biologiques. - 240 -

Figure IV. 3 : Evolution du pH au cours de la culture discontinue de la souche *L. lactis* ssp. *lactis ND5F* à pH libre dans différents MCD. - 241 -

Figure IV. 4 : Descriptif du suivi des consignes des valeurs du couple agitation-aération lors de la régulation de la PO_2 à 30 % en MCD...... - 242 -

Figure IV. 5 : Evolution de la biomasse et du taux de croissance de *L. lactis* ssp. *lactisND5F* au cours de la culture discontinue à pH libre dans différents MCD. Les données courbes sont obtenues par lissage. Les écart-types sont issus de trois réplicats biologiques, sauf pour MCD AA 2X en duplicat. - 242 -

Figure IV. 6 : Evolution du pH au cours de la culture discontinue à pH libre de la souche *L. lactis* ssp. *lactis ND5F* dans différents MCD...... - 243 -

Figure IV. 7 : Cinétique de consommation du substrat (lactose à 60 g.L^{-1}) et de formation des produits (lactate, formiate, acétate, éthanol). Cultures en batch de *L. lactis* ssp. *lactisND5F* à pH libre sur MCD, la référence (A), sur MCD THR 2X (B), sur MCD ILE/LEU/VAL 0,2X (C) et MCD O_2 (D). Résultats moyens de trois réplicats. Les productions d'acétoïne et diacétyle ne sont mentionnées que si leur niveau est supérieur à 2 mM. - 245 -

Tableau IV. 3 : Bilans stœchiométriques carbonés et azotés obtenus dans les différentes conditions de culture en MCD de *L. lactis ND5F*. Les valeurs exprimées en millimoles sont ramenées à une concentration de 50 mmol de lactose (=100 mmol de glucose). ND : Non déterminé et - : Absence. Valeurs calculées à partir ^a des profils de consommation des acides aminés, ^b de la biomasse formée, du rapport $1 \text{ uDO} = 0,33 \text{ g.L}^{-1}$, ^c à partir des produits formés et ^[1] de la masse molaire de la biomasse à 129 g.mol^{-1} [Novak et al. 1997]..... - 246 -

Tableau IV. 4 : Bilans carbone (global et catabolique) et rendements de production de lactate obtenus durant la culture de <i>L. lactisND5F</i> dans les différentes conditions. ^[1] Données de [Novak et al. 1997]. ND : Non déterminé. ^a : Bilan carbone global calculé, ^b : Bilan carbone catabolique : quotient des produits formés sur la source de C consommée.....	- 248 -
Tableau IV. 5 : Concentration en pentane-2,3-dione (en $\mu\text{g.L}^{-1}$) au sein des surnageants en fin de culture de <i>L. lactis ND5F</i> en fonction des conditions testées.....	- 249 -
Tableau IV. 6 : Liste des dix-neuf protéines sélectionnées pour l'analyse protéomique comparative avec leur positionnement métabolique et leur gène respectif.	- 253 -
Tableau IV. 7: Réponses protéomiques comparatives de <i>L. lactis ND5F</i> dans différentes conditions de culture face à la référence MCD. Les <i>p</i> -values du test de Student ne sont présentées que si <i>p</i> -value < 0,05. Pour chaque <i>p</i> -value significative, le ratio d'expression entre la condition testée et la condition MCD référence est présenté. Lorsque celui-ci est supérieur à 1,5 ou inférieur à 0,67, une flèche indique s'il y a sur ou sous expression. Abréviations des protéines, voir Tableau IV. 6.....	- 256 -
Tableau IV. 8 : Caractéristiques des vingt réponses sélectionnées pour construire le profil organoleptique de <i>L. lactisND5F</i>. * Biomasse métaboliquement active. Pour précision, les réponses COVs sont analysées en quantification relative.	- 259 -
Tableau IV. 9 : Plage d'étude des quatre facteurs étudiés dans le plan d'expériences de surface de réponses.	- 262 -
Tableau IV. 10 : Caractérisation des modèles déterminés pour chaque réponse individuelle de <i>L. lactis ssp. lactisND5F</i> à partir des quatre facteurs étudiés. Cas des réponses sucre et acides organiques. R : coefficient de corrélation, R^2 coefficient de détermination, R^2_{adj} coefficient R^2 ajusté en fonction du nombre de coefficients pris en compte et Q^2 , coefficient prédictif. ϵ : erreur du modèle.....	- 265 -
Tableau IV. 11 : Caractérisation des modèles déterminés pour chaque réponse individuelle de <i>L. lactis ssp. lactisND5F</i> à partir des quatre facteurs étudiés. Cas des réponses liées à des indicateurs physiologiques et technologiques. R : coefficient de corrélation, R^2 coefficient de détermination, R^2_{adj} coefficient R^2 ajusté en fonction du nombre de coefficients pris en compte et Q^2 , coefficient prédictif. En gris, un modèle jugé peu descriptif de la réponse étudié. ϵ : erreur du modèle.	- 266 -
Figure IV. 8 : Surfaces de réponses (perspectives et graphique de contours) de la vitesse maximale d'acidification, V^a_{max}, de <i>L. lactis ssp. lactisND5F</i> en lait en fonction des trois facteurs sélectionnés : température (Temp), matière sèche (MS) et PO_2 initiale (O_2). A) V^a_{max} en fonction de la température et de la PO_2 initiale, B) V^a_{max} en fonction de la matière sèche et de la PO_2 initiale et C) V^a_{max} en fonction de la température et de la matière sèche. Le troisième facteur est assigné par défaut à la valeur au centre de son domaine pour chaque surface de réponse.	- 268 -
Tableau IV. 12 : Caractérisation des modèles déterminés pour chaque réponse individuelle de <i>L. lactis ssp. lactisND5F</i> à partir des quatre facteurs étudiés. Cas des réponses liées à des COVs. R : coefficient de corrélation, R^2 coefficient de détermination, R^2_{adj} coefficient R^2 ajusté en fonction du nombre de coefficients pris en compte et Q^2 , coefficient prédictif. En gris, les modèles jugés peu descriptif de la réponse étudié. ϵ : erreur du modèle. Certaines réponses ont été transformées au préalable pour établir le modèle afin de satisfaire les hypothèses mathématiques nécessaires à la résolution des moindres carrés. A l'exception des deux acides aminés, les réponses sont en aires relatives obtenues par GC-MS.	- 271 -

Figure IV. 9 : Surfaces de réponses (perspectives et graphiques de contours) de la vitesse de production de pentane-2,3-dione par *L. lactis* ssp. *lactisND5F* en lait en fonction des trois facteurs sélectionnés : température (Temp), matière sèche (MS) et matière grasse (MG). A) Pentane-2,3-dione en fonction de la température et de la matière sèche, B) Pentane-2,3-dione en fonction de la matière grasse et de la température et C) Pentane-2,3-dione en fonction de la matière grasse et de la matière sèche. Le troisième facteur est assigné par défaut à la valeur au centre de son domaine. - 273 -

Figure IV. 10 : Implication des quatre facteurs, matière sèche (MS), matière grasse (MG), PO2 initiale (O₂) et température (TP), à l'échelle de la signature globale de *L. lactis* ssp. *lactisND5F* en lait. A) Analyse du poids relatif de chaque facteur et B) Analyse du type d'effet majoritaire : effet linéaire (I), quadratique (II) ou d'interaction (Interaction). - 274 -

Tableau IV. 13 : Critères d'optimisation définis pour chacune des dix-huit réponses sélectionnées dans le cadre de la signature globale de *L. lactis* ssp. *lactisND5F* en lait. Les valeurs des bornes supérieures, inférieures et cibles sont indiquées pour les réponses transformées, et entre parenthèses pour les valeurs non transformées, afin de faciliter l'interprétation. Les réponses en gris sont les réponses exclues du deuxième modèle d'optimisation avec un filtre plus strict sur la qualité des modèles. - 275 -

Tableau IV. 14 : Point de fonctionnement identifié pour les quatre facteurs avec optimisation multiple des dix-huit réponses sélectionnées dans le cadre de la signature globale de *L. lactis* ssp. *lactisND5F* en lait.. - 276 -

Tableau IV. 15 : Récapitulatif des valeurs prédites pour chaque réponse à partir du point de fonctionnement optimal présumé. En gris, la seule réponse au seuil de désirabilité inférieur à 50 %. - 278 -

Tableau IV. 16: Récapitulatif des conditions expérimentales des laits fermentés du plan appartenant à la classe dite « crème » après dégustation à 14 J. Echelle de mesure : 0 (absence) et intensité croissante (+, ++, +++). - 280 -

Figure IV. 11 : Analyse PLS-DA de la classification en deux groupes « crème » et « pas crème » à partir des 32 essais du plan d'expériences et des vingt réponses mesurées. A : Représentations des essais (B) et des réponses sur les composantes 1 et 2 de la PLS-DA. Le cercle interne sur la représentation des variables matérialise un coefficient de corrélation de 0,50 avec les composantes PLS. - 281 -

Tableau IV. 17: Comparatif de la signature globale de *L. lactis* ssp. *lactisND5F* en lait entre les valeurs prédites par le modèle et celles expérimentales mesurées pour le point de fonctionnement optimal défini. Expériences réalisées en triplicat. * Impossibilité de disposer pour ces mesures d'un écart-type. - 269 -

INDEX DES ABREVIATIONS

INDEX ABREVIATIONS

Composés

α -kg : α -cétoglutarate
 BCAAs Branched Chain amino acid, acide aminés branchés
 EPS : Exopolysaccharides
 F6P : Fructose-6-phosphate,
 FBP :Fructose-1,6-bisphosphate
 DMDS : Diméthyl-disulfure
 DMTS : Diméthyl-trisulfure.
 G6P : Glucose-6-phosphate
 KG : α -cétoglutarate
 OAA : Oxaloacétate
 pp(p)Gpp: Guanosine 3'-diphosphate, 5'-(tri)/(di)phosphate
 PRPP : 5-phosphoribosyl-pyrophosphate,

Enzymes :

ADHE : Alcool déshydrogénase
 ADI : Arginine déiminase
 AHAS : Acétohydroxyacide synthase, ou acétolactate synthase anabolique
 AHAIIR : Acéto-hydroxy acide isomérase
 ALDC : α -acétolactate déshydrogénase
 ALS : α -acétolactate synthase
 AraT : Aminotransférase des acides aminés aromatiques
 BcaT : Aminotransférase des acides aminés branchés
 CBL : Cystathionine- β -lyases
 CGL : Cystathionine- γ -lyases,
 DAR : Diacétyl-acétoïne réductase
 DCNO : Décarboxylation non oxydative
 DCO : Décarboxylation oxydative
 DHAD : Dihydroxy acide déshydratase
 ENO : Enolase
 GAPDH : Glycéraldéhyde-3-phosphate déshydrogénase,
 GLK : Glucokinase
 ILV : Acétolactate synthase anabolique
 HicDH : Hydroxyisocaproïque acides désydrogénases
 LDH : Lactate déshydrogénase
 LeuB : 3-isopropylmalate déshydrogénase
 LeuC : Isopropylmalate déshydratase grande sous-unité
 LeuD : Isopropylmalate déshydratase petitesous-unité
 PDH : Pyruvate déshydrogénase
 PEP : Phosphoénolpyruvate
 PFK : Phosphofructokinase
 PFL : Pyruvate formiate lyase
 PGI : Phosphoglucose isomérase
 PGM : Phosphoglycéromutase
 PK : Pyruvate kinase
 PTS : Système Phosphoénolpyruvate -phosphotransférase
 TA : Thréonine aldolase
 TD : Thréonine déaminase

TIL : Tryptophane indole lyase
 TPI : Trioses phosphate isomérase
 TPL : Tyrosine phénol lyase

Méthodes & Autres

ACP : Analyse en composantes principales
 AFLP : *Amplified Fragment Length Polymorphisme*, polymorphisme de longueur de fragments amplifiés
 CAH : Classification Ascendante Hiérarchique
 CART : *Classification And Regression Trees*, classification et arbre de régression
 CC: *Clonal Complex*, complexe clonal
 CCR : *Carbon catabolite repression*, répression catabolique du carbone
 CcpA : *Catabolite control protein-A*
 CGH : *Comparative Genome Hybridation*, hybridation comparative de génomes
 CINAC : CINétique d'Acidification
 CPs : Composantes Principales
 cre : *Catabolite response element*
 FPM : Force protomotrice
 Eh : Potentiel d'oxydoréduction
 Eh₇ : Potentiel d'oxydoréduction normalisé à pH 7
 LC : *Liquid chromatography*, chromatographie liquide
 MLST : MultiLocus Sequence Typing
 MS : *Mass spectrometry*, spectrométrie de masse
 MS/MS : *Tandem mass spectrometry*, spectrométrie de masse en tandem
 PLS-DA : Partial least squares discriminant analysis , régression catégorielle des moindres carrées
 PFGE : Pulsed-Field Gel Electrophoresis, électrophorèse en champs pulsés
 RF : Random forest, forêts aléatoires
 RFLP : Restriction Fragment Length Polymorphism, polymorphisme de longueur de fragments de restriction
 ST : Sequence Typing, sequence type

Cofacteurs-nucléotides

ATP : Adénosine triphosphate
 NAD(P)H : Nicotinamide adénine dinucléotide (phosphate) réduit

LISTE DES PUBLICATIONS & COMMUNICATIONS

LISTE DES PUBLICATIONS ET COMMUNICATIONS

Publication écrite, soumise à comité de sélection :

A Dhaisne, M Guellerin, V Laroute, S Laguerre, M Ccaign-Bousquet, P Le Bourgeois, and P Loubiere. Genotypic and Phenotypic Analysis of Dairy *Lactococcus lactis* Biodiversity in Milk: Volatile Organic Compounds as Discriminating Markers. *Appl. Environ. Microbiol.* August 2013(79) 4643-4652.

Communications orales:

A Dhaisne, V Laroute, S Laguerre, L Girbal, P Loubière and M Ccaign-Bousquet. Understanding the diversity of dairy *L. lactis* subsp. *lactis* strains in milk applications: Genome, transcriptome and phenotype integration. ICMP 8th: International Conference on Predictive Modelling in Food.Paris (France) 16-20 septembre 2013.

A. Dhaisne, P Loubière. Analytical and statistical approaches to food metabolomics analysis. Science break du LISBP. Toulouse (France) 29 septembre 2011.

Communication affichée :

A. Dhaisne, M. Guellerin, V. Laroute, P. Le Bourgeois, P. Loubière. Approche intégrative de la typicité de *Lactococcuslactis* ssp. *lactis* « domestiquées ». 9^{ème} Congrès National de la Société Française de Microbiologie. Lille (France) 7-9 février 2013.

Genotypic and Phenotypic Analysis of Dairy *Lactococcus lactis* Biodiversity in Milk: Volatile Organic Compounds as Discriminating Markers

Amandine Dhaisne,^{a,b,c} Maeva Guellerin,^{a,d} Valérie Laroute,^{a,b,c} Sandrine Laguerre,^{a,b,c} Muriel Coccagn-Bousquet,^{a,b,c} Pascal Le Bourgeois,^{a,d} Pascal Loubiere^{a,b,c}

Université de Toulouse; INSA, UPS, INP; LISBP, Toulouse, France^a; INRA, UMR792 Ingénierie des systèmes biologiques et des procédés, Toulouse, France^b; CNRS, UMR5504, Toulouse, France^c; Laboratoire de Microbiologie et Génétique Moléculaires, UMR5100, CNRS, Université Paul Sabatier, Toulouse, France^d

The diversity of nine dairy strains of *Lactococcus lactis* subsp. *lactis* in fermented milk was investigated by both genotypic and phenotypic analyses. Pulsed-field gel electrophoresis and multilocus sequence typing were used to establish an integrated genotypic classification. This classification was coherent with discrimination of the *L. lactis* subsp. *lactis* bv. *diacetylactis* lineage and reflected clonal complex phylogeny and the uniqueness of the genomes of these strains. To assess phenotypic diversity, 82 variables were selected as important dairy features; they included physiological descriptors and the production of metabolites and volatile organic compounds (VOCs). Principal-component analysis (PCA) demonstrated the phenotypic uniqueness of each of these genetically closely related strains, allowing strain discrimination. A method of variable selection was developed to reduce the time-consuming experimentation. We therefore identified 20 variables, all associated with VOCs, as phenotypic markers allowing discrimination between strain groups. These markers are representative of the three metabolic pathways involved in flavor: lipolysis, proteolysis, and glycolysis. Despite great phenotypic diversity, the strains could be divided into four robust phenotypic clusters based on their metabolic orientations. Inclusion of genotypic diversity in addition to phenotypic characters in the classification led to five clusters rather than four being defined. However, genotypic characters make a smaller contribution than phenotypic variables (no genetic distances selected among the most contributory variables). This work proposes an original method for the phenotypic differentiation of closely related strains in milk and may be the first step toward a predictive classification for the manufacture of starters.

The mesophilic lactic acid bacterium *Lactococcus lactis* is one of the most extensively exploited microorganisms; it is used in particular in the manufacture of dairy products. Because of its industrial importance, *L. lactis* has been used as a model bacterium for academic and application-oriented studies. Taxonomically, *L. lactis* has three subspecies, *L. lactis* subsp. *horniae*, *L. lactis* subsp. *lactis* (including *L. lactis* subsp. *lactis* bv. *diacetylactis*), and *L. lactis* subsp. *cremoris* (1). Only *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* are used in starter cultures for dairy production. *L. lactis* subsp. *lactis* colonizes a wide ecological niche (dairy products, as well as animals and plant surfaces) (2), whereas *L. lactis* subsp. *cremoris* is found mostly in dairy environments (3, 4). Recent genotyping methods involving DNA fingerprinting analysis, including pulsed-field gel electrophoresis (PFGE) and comparative genomic hybridization, have been evaluated for their suitability for characterizing genomic diversity and identifying reliable genetic markers for phenotypic subspecies differentiation (5–9). Multilocus sequence typing (MLST), at the level of the gene, has proved to be a powerful method for describing *L. lactis* population structure and phylogeny with a limited number of genes sequenced (5, 6). Recent MLST analysis of *L. lactis* subsp. *lactis* (6) confirmed *L. lactis* population adaptation to the milk niche and led to the proposal of a new classification into two ecotypes, one corresponding to “domesticated” strains essentially isolated from dairy starters or fermented products and the other corresponding to “environmental” strains isolated from various sources such as plants, animals, and raw milk. The domesticated strains make only a small contribution to the genetic diversity of *L. lactis* subsp. *lactis*. Indeed, phylogenetic analysis indicates that they essentially

form two clonal complexes (CCs) that probably emerged only recently from a single founder event (6).

The characteristics of *L. lactis* subsp. *lactis* used in starter cultures for fermented milk dairy production determine their dairy phenotype, which is related to efficient growth in milk associated with fast coagulation. Coagulation is due to acidification and preserves milk from unwanted bacterial or mold growth. However, *L. lactis* subsp. *lactis* also contributes to the organoleptic quality (texture and sensory) of the fermented products. Strain selection has focused on diverse but specific technical characteristics: acidification (10), redox potential (11), texture (12), and physiological features like aroma production (13, 14), nisin synthesis (15), autolysis (16, 17), and enzymatic activities (18). In these previous studies, selection was restricted to one or a limited number of criteria and the behavior of the strains as a whole cannot be predicted. Phenotypic starter screening requires exhaustive study in a controlled dairy environment to evaluate strain performance, but this is a very complex and time-consuming approach that is incompatible with the screening of strain collections. Genotyping

Received 2 April 2013 Accepted 22 May 2013

Published ahead of print 24 May 2013

Address correspondence to Muriel Coccagn-Bousquet, coccagn@insa-toulouse.fr.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.01018-13>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/AEM.01018-13

TABLE 1 Dairy *L. lactis* subsp. *lactis* strains used in this study

Strain	Origin	CC	ST	Presence of:			Reference
				<i>citP</i>	<i>lacE</i>	<i>prtP</i>	
S86	Starter	CC1	10	—	+	+	6
UCMA5713	Grassland (France)	CC1	18	—	+	+	6
S87	Starter	CC1	15	+	+	+	6
EIP41A	Raw milk (France)	CC1	15	—	+	+	This study
EIP33A	Raw milk (France)	CC1	15	—	+	+	This study
DIA-A	Starter Choozit DIA A FRO (France-Danisco)	CC1	15	+	+	+	This study
EIP33F	Raw milk (France)	CC1	34	+	+	+	This study
EIP37F	Raw milk (France)	CC1	34	+	+	+	This study
MC70	Starter Choozit MC-70 FRO (France-Danisco)	CC2	7	—	+	+	This study

and genomic sequencing are efficient and have been largely automated. There was an expectation that this approach could be used to classify and discriminate between strains and predict their phenotypes (19). Several studies combining genomic, transcriptomic, and phenotypic data for *L. lactis* have been conducted in recent years (7, 8, 20–23), but they have failed to establish strong exploitable links among the genotype, its expression, and dairy phenotypes. Rather, they have highlighted significant transcriptomic polymorphisms among *L. lactis* subsp. *lactis* strains with the same dairy origin (21) and the difficulties associated with explaining and predicting a phenotype other than the presence or absence of relationships involving the gene (19, 20). Understanding the determinants of the phenotypic diversity of dairy strains is a prerequisite for predictive approaches and requires phenotypic screening in environmental conditions that are as close as possible to those encountered during the process of interest (18).

Therefore, this work reports an integrated approach to the assessment of the phenotypic biodiversity of nine *L. lactis* subsp. *lactis* strains exhibiting specific dairy phenotypes and selected as being representative of the “domesticated” ecotype in milk. Strain phenotypic signatures were compared with strain genotypic diversity, revealing discrepancies between the two classifications and highlighting the need for an integrated genotypic and phenotypic classification that takes into account both aspects.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Nine strains of *L. lactis* subsp. *lactis* were selected from commercial starter cultures (Danisco-France) and from various laboratory collections (Laboratoire de Microbiologie et Génétique Moléculaires, Université Paul Sabatier, Toulouse, France, for Sx and EIPx strains; LMA, Caen, France, for UCMAx strains). The origins and characteristics of the strains used in this study are listed in Table 1. Stock cultures were kept frozen (−20°C) in M17 broth (Oxoid Ltd., Basingstoke, Hampshire, England) containing 2% (wt/vol) lactose and 20% (vol/vol) glycerol. Individual strains were first grown at 37°C in sterilized (110°C, 15 min) reconstituted skim milk (10.9%, wt/vol). These overnight cultures were used to inoculate (10%, vol/vol) 500 ml of pasteurized standardized cow’s milk (fat, 35 g/kg, protein, 40 g/kg) in 1-liter flasks. The pH was unregulated (initial value of 6.60). Fermentations were continued at 28°C until the pH reached a value of 4.60, and then the milk was transferred to 4°C. The fermented milk was stored in the dark at 4°C for 1 (sample A) or 14 (sample B) days before analysis. Six biological replicates (independent batches) were performed for each strain.

Growth measurements and acidification activity. Bacterial growth in milk (μ , h^{−1}) and cell cultivability after 1 and 14 days of storage were determined in triplicate by plating appropriate dilutions onto M17 agar, incubating them at 30°C for 48 h, and counting the colonies (CFU · ml^{−1}).

Cell cultivability was calculated as the ratio of the viable count after 14 days to that after 1 day.

Acidification activity of lactococcal cells was monitored by using the CINAC system (Alliance Instrument). For each culture, four descriptors were defined to characterize the acidification activity during fermentation: the maximal acidification rate (V_{max}), expressed in 10^{−3} pH units per minute and calculated from the slope of the pH curve as a function of time; the time (in hours) necessary to reach V_{max} ; the pH at which V_{max} is attained; and the time (in hours) necessary for the pH to decrease to 4.60.

Fermentation analysis. Substrate (lactose and citrate) and fermentation end product (lactate, formate, acetate, acetoin, diacetyl, and ethanol) concentrations in fermented milk after 14 days of storage were determined by high-pressure liquid chromatography (HPLC) as previously described (24). Briefly, a solution of 0.3 M barium hydroxide and zinc sulfate was used to precipitate proteins from the samples. The samples were then analyzed in a Bio-Rad HPX87H column maintained at a temperature of 48°C in a 1200 series preparative HPLC apparatus (Agilent Technologies, Waldbronn, Germany), and 5 mM H₂SO₄ was used as the eluent at a flow rate of 0.5 ml · min^{−1}. Free amino acid (FAA) concentrations in milk culture supernatants at 14 days were measured by HPLC as follows. Proteins in the samples were precipitated by adding 4 volumes of methanol to 1 volume of the sample and incubating the mixture overnight at 4°C. The mixture was then centrifuged, and the supernatant was kept for amino acid analysis as previously described (25). The amino acids were automatically derived with *ortho*-phthalic aldehyde (OPA) and 9-fluorenylmethyl chloroformate (FMOC), and the derivatives were separated on a Hypersil AA octadecyl silane column (Agilent Technologies, Waldbronn, Germany) at 40°C by using a linear gradient of acetate buffer (pH 7.2) with triethylamine (0.018%), tetrahydrofuran (0.03%), and acetonitrile (60%). A diode array detector was used at 338 nm for OPA derivatives and at 262 nm for FMOC derivatives. These analyses were performed on the six biologically independent batches for each strain.

Identification and semiquantification of volatile organic compounds (VOCs) in fermented milk. VOCs in fermented milk were identified and semiquantified by the Analytical Chemical Service of ISVV (Institut des Sciences du Vin et de la Vigne), Bordeaux, France. Separation and semiquantification were carried out by solid-phase microextraction (SPME)-gas chromatography (GC)-mass spectrometry (MS) analysis.

The internal standards used for GC analysis were [2H₅]ethyl acetate (ethyl acetate-d₅) and [2H₁₅]octanoic acid (octanoic acid-d₁₅) from Sigma-Aldrich (Saint-Quentin-Fallavier, France) and [2,2-2H]3-methylbutanal (3-methylbutanal-d₂) from Euriso-top (Saint-Aubin, France). Solutions were prepared at 100 mg · liter^{−1} with MilliQ water and kept at 4°C.

A 5-g sample of fermented milk was mixed with 250 μ l of internal standard mix solution (at 10, 20, and 20 mg · liter^{−1} for ethyl acetate-d₅, octanoic acid-d₁₅, and 3-methylbutanal-d₂, respectively) in a 20-ml glass sample vial. The vial was then tightly sealed, and the contents were homogenized with a vortex shaker at 55°C for 30 min. The SPME fiber

(Supelco, Bellefonte, PA) used was coated with a 50/30- μm layer of divinylbenzene-carboxen-polydimethylsiloxane (57299-U; StableFlex) and maintained in the headspace at 55°C for 90 min. The volatile analytes adsorbed to the SPME fiber were analyzed by GC-MS with an HP 5890 GC system coupled to an HP 5972 quadrupole mass spectrometer (Agilent Technologies, Wilmington, DE) equipped with a Gerstel MPS2 autosampler. The compounds were separated on a BP 21 capillary column (60 m by 0.32 mm, 0.25- μm film thickness; SGE, Courtaboeuf, France). The injection port was programmed to heat at 250°C. The oven temperature was programmed at 40°C for 5 min, raised to 240°C at 3°C $\cdot$ min⁻¹, and then held at that temperature for 10 min. Volatile compounds were identified in SCAN mode (mass range of 29 to 200 m/z at a scan rate of 1.49 scans $\cdot$ s⁻¹), and the ionization energy was set at 70 eV. Only compounds for which the area was at least 2-fold the background noise were selected. The eluted compounds were identified by their retention times and by comparison of their mass spectra with those in the National Institute of Standards and Technology database (<http://www.nist.gov/srd/> [July 2002 version]).

Selected volatile compounds were detected and semiquantified with the mass spectrometer operating in selected-ion-monitoring mode with electron ionization at 70 eV. VOCs were semiquantified by determining the ratio of the total ion count (TIC) of each compound to the TIC of the corresponding internal standard for 1 g of fermented milk. The internal standard was chosen according to the chemical function and the retention time. For each strain, analytical triplicates were performed. Detection and quantification limits and coefficients of variation were determined for each compound.

DNA manipulation and phylogenetic analysis. Bacteria were grown at 30°C on M17 broth. Genomic DNA was extracted from overnight cultures with the DNeasy tissue kit according to the manufacturer's instructions (Qiagen, Hilden, Germany). Genetic markers of important industrial traits (*lacE*, encoding lactose-specific enzyme II of the phosphotransferase [PTS] system; *prtP*, encoding the cell envelope-associated serine proteinase; and *citP*, encoding the membrane-bound citrate permease involved in citrate uptake) were detected by PCR amplification and standard agarose gel electrophoresis. PCR primers for the *lacE* (*lacE*-F2, 5'-AGCGTCTATGGTAGGGTCC-3'; *lacE*-R2, 5'-GATGGCACGGTTACGATCTG-3'; PCR product size of 606 bp), *prtP* (*prtP*-F2, 5'-GAGGCAGTGAAGCTGTTAGTC-3'; *prtP*-R2, 5'-TCATTCGCAGCAGTACATC-3'; PCR product size of 713 bp), and *citP* (*CitP*1, 5'-ATGATGAATCA CCGG-3'; *CitP*2, 5'-ACTTCATGAATATGAC-3'; PCR product size of 1,327 bp) genes were designed by standard procedures with Clone Manager version 9.0 software (Sci-Ed Software). The cycling conditions used were 95°C for 5 min followed by 30 cycles of 95°C for 30 s, 55°C for 1 min, and 72°C for 1 min for the *lacE* and *prtP* genes or 98°C for 5 min followed by 30 cycles of 98°C for 10 s, 42°C for 30 s, and 72°C for 1 min for the *citP* gene, using an MJ Mini thermocycler (Bio-Rad, Hercules, CA). Each PCR involved a 25- μl mixture containing 10 ng of genomic DNA, 0.5 mM each primer, and 12.5 μl of iQ SYBR green Supermix (Bio-Rad, Hercules, CA).

MLST analysis was performed by the six-locus MLST scheme exactly as described in reference 6. The concatenated sequences generated were used for phylogenetic analysis with MEGA5 software (26).

For PFGE, preparation of genomic DNA embedded in an agarose matrix, digestion of DNA by restriction endonucleases, and electrophoresis were performed as previously described (27). The genomic relatedness of bacterial strains was estimated from pairwise comparisons of PFGE *Sma*I macrorestriction patterns, and a matrix of binary data was constructed from the presence or absence of each band. Dice coefficients (S_D) and corresponding genomic distances ($1 - S_D$) for each pairwise comparison were calculated from the matrix of binary data with the WINDIST program (28). Dendrograms based on the unweighted-pair group method using average linkages (UPGMA) were constructed with the NEIGHBOR program of the PHYLIP package v3.69 (29).

Data analysis. All variables (physiological descriptors, metabolite concentrations, and genetic distances) were first normalized (centering

[subtracting the population mean] and then scaling [dividing this difference by the standard deviation of the population]) to allow an unbiased comparison of these heterogeneous data. Missing data in COV semiquantification for a third of the biological replicates were filled in by using the geometric mean of the available replicates of that strain. To investigate the relationships between strains and the phenotypic variables of fermented milk, an analysis of variance (ANOVA) of the 82 phenotypic variables was performed. The level of significance for all statistical analyses was set to a P value cutoff of 0.05. Spearman's rank-order correlations with P values adjusted by the Benjamini-Hochberg method to control the false-discovery rate (30) were performed to investigate pairwise associations between the variables involved in strain phenotypes in milk. Multidimensional methods such as principal-component analysis (PCA), hierarchical ascendant classification (HAC), partial least-squares discriminant analysis (PLS-DA), and sparse PCA were carried out with R free statistical software (2.15.0 version) and the mixOmics package (31). The Euclidian distance metric and Ward's criterion were used for HAC in phenotypic and genotypic classifications to constitute hierarchical groups of mutually exclusive subsets in which members are maximally similar with respect to their specified characteristics (32). Bootstrap analysis was used with 1,000 simulations by cluster analyses. Consensus trees were built by bootstrapping (bootstrap value >0.50). To identify discriminatory variables, a variable selection method was developed on the basis of the principal components (PCs). The variable selection combined a test of strain dependency (removal of one strain from the sample in order to differentiate markers highly dependent on a strain from common markers) and a selection of the most contributory variables for this set of individuals. The contribution of a variable to the selected PCs is obtained by the ratio of the sum of the squared factor score divided by the sum of the eigenvalues. The contributory threshold was set to 75% for this procedure; therefore, only variables with a contribution above the third quartile of the contributory distribution were selected as most contributory variables. This procedure was done iteratively for each strain, and strain clustering based on the variables selected was then performed. The cluster robustness and consequently the reliability of the selected variable set were assessed by bootstrapping. In this work, the variables selected represented 25% of the total variables (i.e., 20 variables for phenotypic analyses and 22 for phenotype-genotype investigation, respectively) with the highest head count among each set of the most contributory variables. In order to provide a deeper analysis of the robustness of the variable selection method developed, results were compared to those of variable selection through sparse PCA (33).

RESULTS

Genotypic characterization of *L. lactis* subsp. *lactis* strains. To investigate the phenotypic diversity of *L. lactis* subsp. *lactis* strains under growth conditions close to those of relevant industrial processes, nine strains isolated from raw milk and starter cultures (Table 1) were selected according to their dairy phenotype as pure cultures (ability to grow efficiently in milk inducing fast coagulation). All of these strains were considered to belong to the "domesticated" ecotype (6). The genetic markers of industrial traits checked by PCR amplification were the *lacE* gene, encoding lactose-specific enzyme II of the PTS system, and the *prtP* gene, encoding the cell envelope-associated serine protease. Both were present in all nine strains. *L. lactis* subsp. *lactis* bv. diacetylactis isolates are widely used in the dairy industry because of their ability to take up and consume citrate to produce diacetyl and acetoin, two components essential for a creamy and buttery aroma (8, 18, 22). Four strains in our panel were assigned to *L. lactis* subsp. *lactis* bv. diacetylactis on the basis of results of PCR amplification of *citP*, a plasmid-borne gene encoding the permease involved in citrate uptake (34).

"Domesticated" *L. lactis* subsp. *lactis* currently contains 13 se-

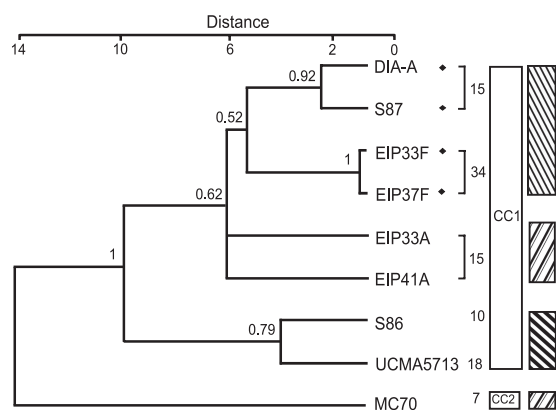


FIG 1 Integrative genotypic classification of nine domesticated *L. lactis* subsp. *lactis* strains from genetic (MLST) and genomic (PFGE) data sets. The Euclidean distance metric and Ward's criterion were used for hierarchical ascendant clustering on scaled MLST and PFGE distances. From left to right, a strain dendrogram, strain names, the presence of the *citP* gene (black diamonds), STs, CCs, and the three main clusters based on PFGE classification (shading) are shown. Bootstrap values above 0.50 are indicated at the corresponding nodes ($n = 1,000$).

sequence types (STs) organized into two main CCs (<https://www-mlst.biotoul.fr/Lactococcuslactissubsp. lactis/>). The main CC (CC1) contains 11 STs (ST1, ST6, ST9, ST10, ST15, ST16, ST18, ST22, ST23, ST34, and ST36), with ST15 predicted to be the ancestor genotype, whereas CC2 comprises ST7 and ST4. MLST analysis revealed that eight of the nine strains used in this study belonged to CC1 (four strains of ST15, two strains of ST34, one strain of ST10, and one strain of ST18) and that one in CC2 was ST7 (Table 1). We therefore considered our strain sample to be genetically representative of the “domesticated” ecotype. As strain redundancy in bacterial collections cannot be excluded, especially for strains belonging to the same ST, SmaI macrorestriction analysis and PFGE were used to confirm that each of these strains was unique and different from the other eight. The genomic relatedness of the selected strains was assessed by computing S_D s from pairwise comparisons of the SmaI macrorestriction patterns (see Fig. S1A in the supplemental material): 52% of the calculated S_D values were lower than 0.35, the value observed when comparing the two different subspecies *L. lactis* subsp. *cremoris* and *lactis* (6). These findings confirmed previous descriptions of the substantial genomic variability within *L. lactis* subsp. *lactis* (6). The consensus tree built from UPGMA-based clustering of PFGE distances yielded three major clusters (bootstrap value >0.50) and clustered the four strains of *L. lactis* subsp. *lactis* bv. *diacetylactis* together (bootstrap value, 0.61) (see Fig. S1B). However, strains of the same ST or the same CC were not necessarily grouped together.

We thus considered unsupervised hierarchical clustering integrating both genetic (MLST) and genomic (PFGE) data sets as an alternative approach to improve strain clustering by taking into account genotypic aspects. For this purpose, HAC with the Ward criterion was performed to constitute hierarchical groups of mutually exclusive subsets (32). The accuracy of the genotypic dendrogram was assessed by bootstrapping, and a consensus tree was built (bootstrap value >0.50) (Fig. 1). This integrated genotype classification is consistent with the mean features of gene phylogeny, *L. lactis* subsp. *lactis* bv. *diacetylactis* discrimination, and strain genomic uniqueness (Fig. 1). However, it failed to group

strains of ST15 in one class. The latter result can be explained by evolutionary concepts. Knowing that clonal diversification of lactococcal strains is mostly dependent on genome rearrangements that have large effects on PFGE fingerprints (35), ST15 may be more prone to macrorestriction polymorphism than other STs of its CC (CC1) because it is the ancestor genotype.

Phenotypic biodiversity and subpopulation structure. (i) **Phenotypic data set.** Eighty-two variables selected as being descriptive of dairy performance, including physiological indicators (growth, acidification) and extracellular metabolic products (sugars, FAAs, organic acids, and VOCs) were assayed for the nine *L. lactis* subsp. *lactis* strains grown in pasteurized standardized cow's milk. Data for some physiological indicators (acidification and growth kinetics) were collected at various times during the milk fermentations, but most were collected after 14 days of storage (corresponding to the estimated half-life of fermented milk). Under all conditions, growth started immediately after inoculation, with maximal growth rates ranging from 0.60 to 0.80 h⁻¹. No significant difference was found for cell populations: there were 2.1×10^9 to 2.6×10^9 CFU · ml⁻¹ at the end of the culture (when the pH reached 4.60). Cell cultivability after 14 days was between 50 and 70%, confirming the growth similarities of all nine strains.

The acidification properties of the nine strains were compared. As expected, the mean acidification rate of the four *L. lactis* subsp. *lactis* bv. *diacetylactis* strains was 40% lower than that of the other strains; this was due to the 4-fold higher acetoin concentration and lower lactic acid production of the *L. lactis* subsp. *lactis* bv. *diacetylactis* strains. The time taken for the fermentation to reach a pH of 4.60 was 2 h longer for the *L. lactis* subsp. *lactis* bv. *diacetylactis* strains than for the other strains, a delay considered to be significant for the dairy industry. Supernatant from 14-day fermented milk was assayed for 18 FAAs; all were detected, and their concentrations could be quantified. However, the biogenic amine γ -aminobutyric acid (GABA) was detected in only three strains, EIP33A (0.1 mM), DIA-A, and MC70 (1 mM). A total of 47 different VOCs, belonging to the hydrocarbon, alcohol, aldehyde, ketone, ester, sulfide compound, and free fatty acid (FFA) families, were identified in the samples and quantified. To investigate the relationships between strains and the phenotypic variables of fermented milk, an ANOVA of the 82 variables was performed. This analysis revealed that only VOCs were strongly dependent on the strains (P value of <0.05 for the ANOVA test).

(ii) **Metabolic exploration of fermented milk by correlation investigations.** The analysis of these 82 scaled variables as a single integrated phenotypic data set allowed exploration of the metabolic network in fermented milk. Spearman's rank-order correlations after P value multiple-testing corrections by the Benjamini-Hochberg method (P value cutoff, 0.05) were then performed to investigate pairwise associations between the variables involved in strain signatures in milk (see Fig. S2 in the supplemental material). The principal finding was a strong correlation ($R \geq 0.70$) among VOCs of the FFAs (C_4 to $C_{9,2}$), methyl-ketone (2-undecanone, 2-tridecanone, 2-pentadecanone), aldehyde (hexadecanal), alcohol (heptanol, octanol, nonanol), and lactone (δ -octalactone, δ -decalactone, δ -dodecalactone) families (see Fig. S2A). All of these metabolites are involved in the four main steps of lipolysis, i.e., release of FFAs from milk triglycerides, followed by β -oxidation of released FFAs to α -keto acids, then decarboxylation to alkan-2-ones, and finally reversible reduction to the correspond-

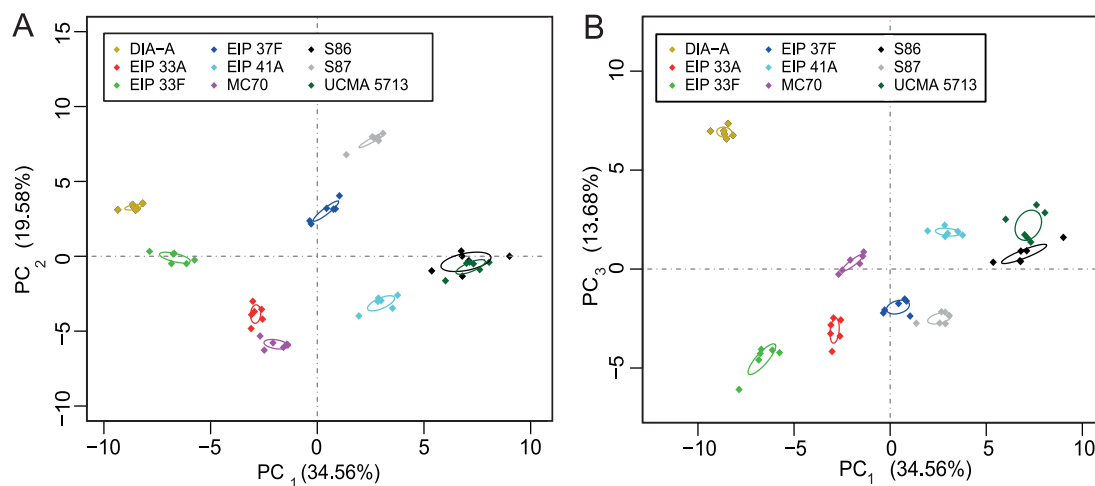


FIG 2 PCA of domesticated *L. lactis* subsp. *lactis* strains in milk from an integrative phenotypic data set. Both panels A and B are score plots of PC₁, PC₂, and PC₃ scores. Diamonds of the same color correspond to biological replicates of the same strain. Ellipses show the strain category models at a 95% confidence level.

ing alkan-2-ols (13, 36–38). The strong relationships among these metabolites confirmed that they are involved in the same pathway.

The metabolites derived from glycolysis, more specifically, those involved in pyruvate metabolism (acetaldehyde, lactic acid, acetoin, diacetyl, and butanediol), displayed strong positive correlations with each other ($R \geq 0.67$). The contribution of citrate consumption to the production of diacetyl, acetoin, acetic acid, lactic acid, and butanediol was confirmed by strong negative correlations ($R \leq -0.75$) (see Fig. S2B).

Several correlations related to proteolysis were also identified. First, the catabolic link between benzaldehyde and its amino acid precursor, phenylalanine, was confirmed by a strong negative correlation ($R < -0.70$). Metabolites of methionine catabolism (sulfide and thiol VOCs) were positively correlated (see Fig. S2B). Lastly, 3-methylbutanal and 2-methylpropanol, the main products of branched-chain amino acid catabolism, were positively correlated ($R = 0.61$).

(iii) High strain phenotypic diversity. A PCA of the 54 fermented milk samples (six cultures of each of the nine strains) was used to study phenotypic strain diversity. This analysis was limited to the first three PCs because the cumulative variance analysis of PC₁ to PC₃ accounted for 67.8%, a percentage sufficiently high to ensure that the PCA plots were representative of the main features of the data set. On the basis of visual explorations of the score plots (Fig. 2A and B), nine independent signatures, i.e., one per strain, were observed. This result illustrates the high phenotypic diversity of genetically closely related strains, all exhibiting a dairy phenotype (efficient growth in milk associated with fast coagulation). This result also confirms the suitability of the 82 monitored variables for establishing a discriminatory phenotypic data set for investigation of the diversity of the “domesticated” *L. lactis* subpopulation.

Thus, each of these nine strains has a unique phenotypic signature. We classified them according to their metabolic proximity. Although the PCA score plots allowed intuitive strain grouping, a clustering analysis was preferred to allow the robustness of the strain classification to be assessed. The 82 variables were used to build a dendrogram by hierarchical clustering (Fig. 3). The accuracy of the classification was checked by bootstrap analysis,

which revealed that all of the biological replicates of a strain (six independent experiments) were repeatedly grouped together (bootstrap value >0.94), confirming the quality of the data. Thus, our integrated phenotypic approach allows robust and accurate strain identification and discrimination by analysis of fermented milk. It would be interesting to analyze how these nine strains could be grouped together on the basis of the proximity of their phenotypes. To choose the relevant number of strain clusters in this classification, two criteria were examined: the decrease in interclass variance (data not shown) and bootstrap analysis (Fig. 3). However, with these criteria, four, five, and nine clusters were possible options, without it being possible to conclude which was the most appropriate. Consequently, further investigations were required to determine the number of clusters and, by extension, the organization of the “domesticated” *L. lactis* subsp. *lactis* subpopulation.

(iv) Subpopulation organization and variable selection. To identify the relevant number of strain clusters in the “domesticated” *L. lactis* subsp. *lactis* subpopulation, a robustness analysis was carried out. This strain cluster analysis also aimed to reduce the number of variables used from 82, since strain discrimination is no longer required. This approach allowed the identification of the most significant phenotypic markers. We focused our study on the variables making the largest contribution, i.e., those most responsible for strain positioning. From the variable selection (for details, see Materials and Methods), 20 variables, corresponding to 25% of the 82 initial variables, was found to be the minimal number of most contributory variables allowing robust strain clustering in each case (data not shown). Four strain clusters were thereby defined (bootstrap value >0.82) and named I to IV. Their compositions are shown in a dendrogram in Fig. 3. A supplementary PLS-DA was then used to estimate the error rate of strain affiliation with the four clusters identified. With only the 25% of the variables that contributed the most to this clustering, perfect accuracy (100%) in strain cluster affiliation was predicted.

Thus, 20 of the 82 variables studied were identified as the most contributory. Variable selection through sparse PCA confirmed the suitability of the set of discriminatory variables. All of these selective variables were related to VOCs and were classified into

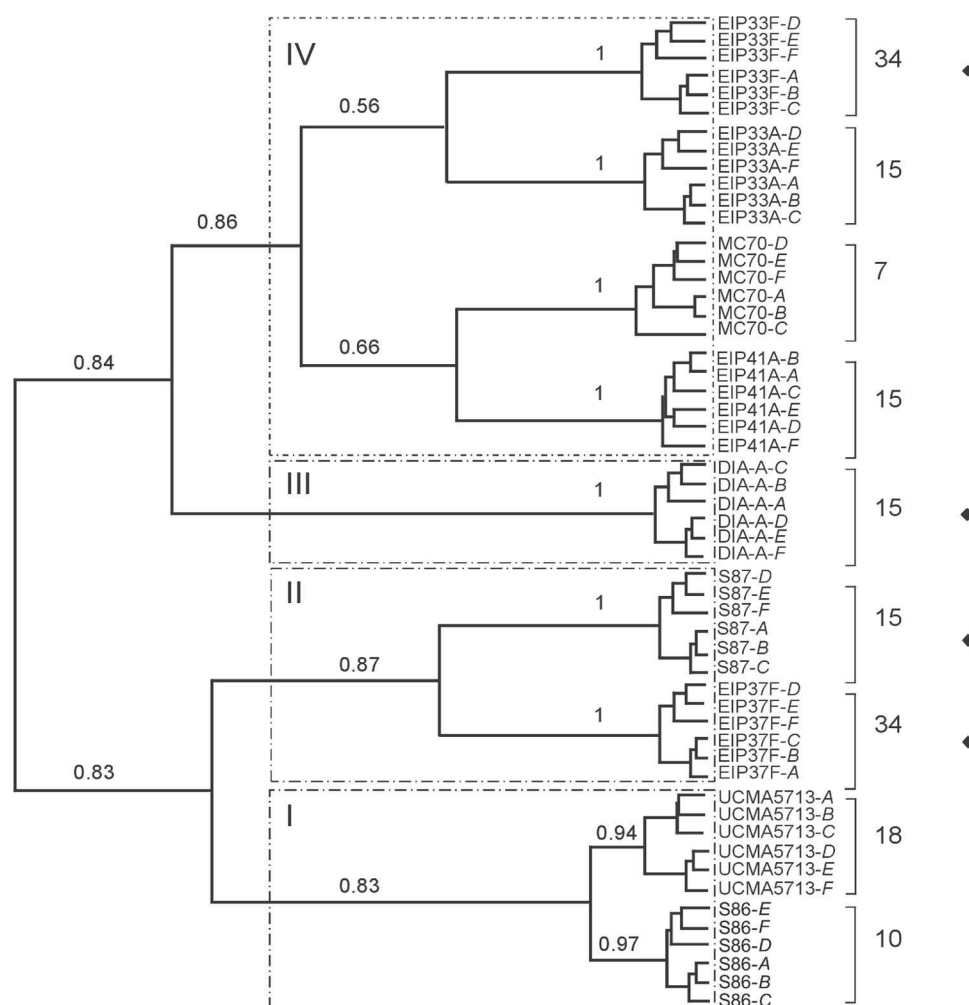


FIG 3 Integrative phenotypic classification of nine domesticated *L. lactis* subsp. *lactis* strains grown in milk. The Euclidian distance metric and Ward's criterion were used for hierarchical ascendant clustering based on 82 variables. From left to right, a strain dendrogram, the names of the 54 experiments (nine strains with their six replicates, from A to F), STs from MLST, and the presence of the *citP* gene (black diamond) are shown. Bootstrap values above 0.50 are indicated at the corresponding nodes ($n = 500$). Dotted squares with roman numerals I to IV in the left panel correspond to the four main conserved clusters identified by robustness analysis.

three main groups (A, B, and C) (Table 2). These groups are representative of the three major catabolic pathways involved in flavor formation, consistent with the findings of the correlation analysis. Group A is related to glycolysis (and citrate products except for nonan-2-one), group B is involved in proteolysis, and group C is involved in lipolysis. Analysis of the correlations between these three groups of selected variables and the first three PCs of the PCA (Table 2) demonstrated the relevance of the PCs for the analysis of metabolism.

To interpret the metabolic significance of each PC, the major coefficients of correlation between the three variable groups and each PC were investigated (Table 2). The first PC, with which all of the group C variables (plus nonan-2-one) were positively correlated, could be interpreted as the lipolytic activity of the strains in milk. Group A variables were positively correlated with the second PC, suggesting that PC2 is related to glycolysis. The third PC, correlated only with group B variables, may be interpreted as an indicator of proteolytic activity. From these results, the metabolic orientations of the signatures of single strains could be easily iden-

tified. Thus, strains UCMA5713 and S86 display the highest lipolytic activity; S87 produces the largest amounts of the end products of the citrate pathway; and strain DIA-A displays the highest concentrations of VOCs from amino acid catabolism (Fig. 2A and Table 2).

The biological significance of these four clusters was investigated according to the metabolic interpretation of each of the PCs (as described by using the 20 most contributory variables, groups A, B, and C) and by positioning the four strain clusters in PCA score plots (Fig. 2A and B). In cluster I, strains UCMA5713 and S86 produced predominantly VOCs by intense lipolysis (group C) whereas the levels of these compounds were relatively low in fermentations with cluster III strains (DIA-A). In cluster II, strains EIP37F and S87 were characterized mainly by the production of acetoin, diacetyl, and butanediol (group A). Note that the DIA-A strain was a singleton (cluster III), emphasizing its extreme phenotypic specificity. DIA-A was also associated with strong amino acid catabolism (group B). Finally, the four remaining strains (EIP41A, MC70, EIP33A, and EIP33F) belonged to cluster IV;

TABLE 2 Coefficients of correlation of the 20 most contributory variables in the PCA to the three PCs^a

Group and variable	PC ₁		PC ₂		PC ₃	
	R	P value	R	P value	R	P value
A						
Acetoin	−0.55	1.45×10^{-5}	0.77	1.54×10^{-11}	−0.37	5.37×10^{-3}
Acetaldehyde			0.86	5.51×10^{-16}		
2-Nonanone	0.31	2.19×10^{-2}	0.89	9.87×10^{-20}		
Butan-2,3-diol D ⁺	−0.46	5.24×10^{-4}	0.81	8.77×10^{-14}		
B						
2-Methylbutanal	−0.57	5.50×10^{-6}	0.29	3.05×10^{-2}	0.74	1.95×10^{-10}
Butyl acetate	−0.54	2.24×10^{-5}	−0.29	3.05×10^{-2}	0.76	3.83×10^{-11}
2-Methylpropan-1-ol	−0.33	1.54×10^{-2}	0.51	7.56×10^{-5}	0.73	2.64×10^{-10}
2-Hydroxypentan-3-one	−0.61	8.44×10^{-7}	0.43	1.10×10^{-5}	0.60	1.39×10^{-6}
Ethanol	−0.28	4.22×10^{-2}	−0.43	1.19×10^{-3}	0.90	8.02×10^{-21}
C						
Octanol	0.79	1.28×10^{-12}	0.53	3.21×10^{-5}		
Undecane-2-one	0.90	1.94×10^{-20}	0.36	7.82×10^{-20}		
Butanoic acid	0.96	8.93×10^{-30}				
2-Tridecanone	0.88	8.68×10^{-19}	0.44	7.48×10^{-4}		
Hexanoic acid	0.93	5.21×10^{-24}	0.28	3.73×10^{-2}		
Heptanoic acid	0.92	3.38×10^{-23}	0.35	8.78×10^{-3}		
2-Pentadecanone	0.89	8.43×10^{-20}	0.42	1.43×10^{-3}		
Octanoic acid	0.92	3.71×10^{-22}	0.34	1.23×10^{-2}		
Nonanoic acid	0.89	8.73×10^{-20}	0.37	5.51×10^{-3}		
Decanoic acid	0.93	1.00×10^{-23}	0.28	3.94×10^{-2}		
δ-Dodecalactone	0.90	2.95×10^{-29}	0.32	1.73×10^{-2}		

^a Significance was defined as a *P* value below 0.05. *R*, coefficient of correlation of the variable to the PC. Hierarchical clustering of ascendant classification using the Euclidian distance metric and Ward's criterion is shown for groups A to C.

their phenotypic signatures are not characteristic of a particular metabolic pathway, with near mean concentrations of the most contributory VOCs. Therefore, we considered cluster IV to be characterized by no dedicated metabolic pathway.

These findings proved that within a single cluster, other phenotypic variables were responsible for strain-specific signatures. To identify the variables that are potentially specific markers of each strain, a PCA of the complete phenotypic data set for each strain cluster was performed. For metabolically undefined cluster IV, this analysis revealed that EIP41A was further specifically identified by high sulfide VOC concentrations (responsible for onion and sulfur aromas) whereas EIP33F was distinguished by high residual amino acid (His, Ala, Pro, Gln, Val, Leu, Ile, Phe) concentrations in milk. Both a high growth rate and a high ethanol concentration were characteristics of EIP33A, whereas MC70 appeared to be characterized by high furfural and 2-furanmethanol concentrations and the presence of GABA. The metabolites furfural and 2-furanmethanol can be produced in the presence of L-alanine and glucose under acidic conditions (especially at pHs ≤5) and are characterized by a “bread and burnt” aroma (39).

Genotype-phenotype integration. Phenotype-based strain clustering (Fig. 2) appeared to correspond to neither MLST phylogeny (strains of the same ST did not cluster together) nor genomic PFGE profiles (four clusters formed rather than the three groups of related strains suggested by PFGE) (Fig. 1). Moreover, the four strains belonging to *L. lactis* subsp. *lactis* bv. diacetylactis were distributed among different phenotypic clusters (II, III, and IV, Fig. 3). Comparisons of the integrated phenotypic classification (Fig. 3) with genotypic strain diversity (Fig. 1) confirmed the

discrepancies among the classifications. In particular, the specificity of strain MC70 is underlined in the genotypic classification because of its specific CC2 affiliation of the “domesticated” strains, but its phenotypic signature was identified as not dedicated to a particular metabolic pathway. These results highlight the need for a system of classification integrating both genotypes and phenotypes to describe *L. lactis* subsp. *lactis* biodiversity.

Using the integrated genotype-phenotype data set, strains were subjected to both PCA and clustering analysis (Fig. 4). This led to a five-cluster organization rather than the four strain clusters (I to IV) generated with discriminatory phenotypic markers. Two groups in these two classifications were similar (clusters II and III). The previous cluster IV was the most affected by the new classification. In particular, MC70, which was previously classified in cluster IV, was now singled out. The variable analysis underlined the contribution of MLST phylogenetic distances to PC2 (data not shown) to this new classification and thus the position of MC70. The use of independent genetic (MLST) and genomic (PFGE) data sets in addition to phenotypic variables revealed the relevance of MLST in the classification. Indeed, unlike MLST data, the PFGE genomic diversity data set did not change the phenotypic PCA or clustering results (data not shown). The other strains previously classified in cluster IV (EIP41A and EIP33F) were classified in existing clusters (I and IV) by the integrated genotype-phenotype classification.

Variable selection (as previously described) demonstrated that no genotypic variable (either MLST or PFGE distances) was selected as being one of the most contributory variables. The con-

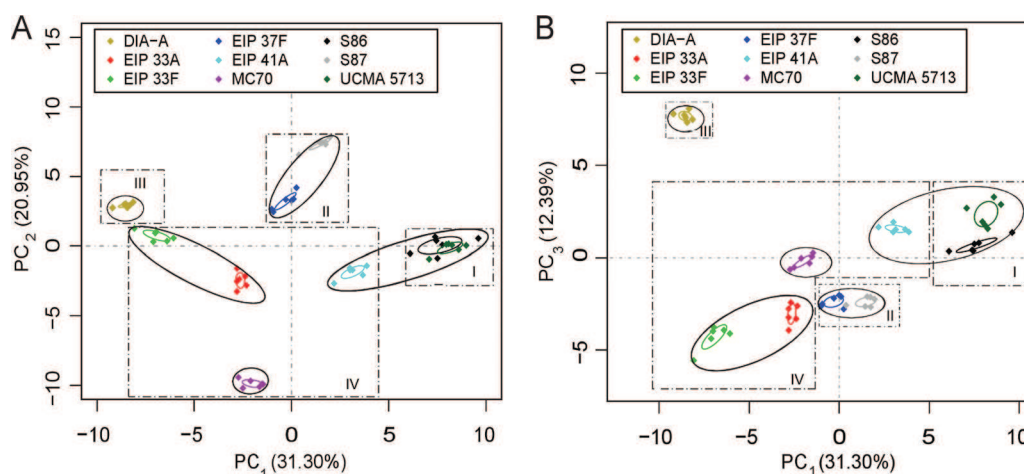


FIG 4 Integrative phenotypic and genotypic PCA of nine domesticated *L. lactis* subsp. *lactis* strains grown in milk. Score plots of PC₁-PC₂ (A) and PC₁-PC₃ (B) are shown. Diamonds with the same color are biological replicates of the same strain, and colored ellipses show the strain category models (at a 95% confidence level). Dotted squares with roman numerals I to IV correspond to the four phenotypic clusters (Fig. 3), and black ellipses represent the five integrative genotype-phenotype clusters.

tributary variables were, again, mostly VOCs (21/23), acetic acid, and fermentation time.

DISCUSSION

We report an analysis of the phenotypic diversity of nine strains of *L. lactis* subsp. *lactis* representative of the “domesticated” ecotype and exhibiting a dairy phenotype (efficient growth in milk and fast coagulation). Eighty-two variables related to important dairy features, including physiological descriptors (growth, acidification) (10), carbon and nitrogen metabolites, and flavor determinants (14, 40), were monitored in 54 fermentations (6 per strain). Our integrated phenotypic approach confirms the suitability of these 82 variables for the establishment of well-characterized dairy strain signatures in fermented milk. Furthermore, this study reveals large strain-to-strain phenotypic divergences that are sufficient for the accurate identification of the nine strains from their milk fermentation signatures. Consequently, these 82 variables provide a useful data set for the classification of dairy strains and provide new possibilities for applications, particularly for strain identification and selection.

Unexpectedly, these results illustrate the substantial phenotypic diversity in a limited genetic subpopulation of strains. Indeed, phylogenetic studies suggest that dairy strains are poor contributors to the overall genetic diversity of *L. lactis* subsp. *lactis* (8, 21, 41). “Domesticated” strains arose from a genetic bottleneck or were the result of the successful adaptation to milk of a common “environmental” ancestor due to genomic flexibility (6, 42, 43). Consequently, it has been suggested that the phenotypic diversity of dairy strains is expected to be low because of the small pool of strains (44, 45). However, despite the extensive genotypic similarities, we report substantial phenotypic divergence between “domesticated” strains: consequently, these strains have different potentials for different industrial applications.

The phenotypic biodiversity of these dairy strains could be subdivided into four robust clusters based on metabolic orientations (lipolysis, proteolysis, and glycolysis). These clusters may be helpful in assessing strain relatedness. The analysis of more *L. lactis* subsp. *lactis* collections, including strains representative of the

“domesticated” ecotype from several dairy environments, would probably lead to the compositions of these clusters being modified and to the identification of additional clusters.

However, these findings imply that it would be useful to have appropriate markers to define the specific abilities of a strain. Indeed, a clear description of the differences between strain groups is necessary for better control of starter choice. Therefore, we used the integrated phenotypic data set to identify the 20 variables that contribute the most to discrimination between the strains. All of these variables are related to VOCs. Our study confirms that selected VOCs can be used as discriminatory phenotypic markers to predict strain signatures in milk processing (46). Combined with the recent development of ultrafast GC-time of flight MS technology for fermented milk (46, 47), monitoring of these markers should help in the high-throughput screening of strain collections. Further analyses of these 20 VOCs revealed that the three major pathways involved in flavor, i.e., lipolysis, proteolysis, and glycolysis, were represented. Proteolysis and glycolysis in milk have been studied and described (13, 48, 49), whereas lipolysis in fermented milk has been regularly underestimated; this is largely because many experiments are carried out with skim milk media. This work underlines the importance of lipolysis, because more than half of the most discriminatory VOC markers belonged to this class.

L. lactis subsp. *lactis* bv. diacetylactis strains are used in the dairy industry to produce acetoin and notably diacetyl, which imparts a high level of buttery flavor notes. Surprisingly, the four strains in our sample belonging to *L. lactis* subsp. *lactis* bv. diacetylactis (DIA-A, EIP33F, S87, and EIP33F) were not grouped in a single cluster at a phenotypic level but were classified together by integrated genotypic classification. Interestingly, this feature demonstrates that acetoin and diacetyl production is not relevant for discriminating *L. lactis* subsp. *lactis* bv. diacetylactis strain signatures in milk at a phenotypic level. A recent study of diacetyl- and acetoin-producing *L. lactis* strains isolated from diverse origins indicated that the production of this aroma is not restricted to strains able to grow in milk (50). These observations illustrate the intragroup metabolic diversity of these strains and show the need

for further metabolic investigations to characterize these strains thoroughly (notably, amino acid and lipid catabolism).

The phenotypic classification was not entirely consistent with the genetic (MLST) and genomic (PFGE) dendrograms or with the new integrated genotypic classification. The proposed integrated genotypic classification of these dairy strains was coherent with the CC phylogenetic organization of “domesticated” strains (MLST) (6), discrimination of the *L. lactis* subsp. *lactis* bv. diacetyl-lactis lineage (based on results of PCR amplification of *citP*), and strain genomic uniqueness (nonredundancy of *Sma*I macro-restriction by PFGE).

Genotype and phenotype matching is not straightforward, as previously demonstrated, because of the regulation of genome and proteome expression (18, 21, 43). However, we identified contributory phenotypic determinants that are promising for the discovery of reliable genetic markers. This approach has been used previously but only at the *L. lactis* subsp. *lactis* and *cremoris* level (9, 22). On the basis of the three groups of VOCs identified as the most contributory variables, the corresponding genes (i.e., those involved in lipolysis, proteolysis, and glycolysis and their regulation) could be sequenced and their variability analyzed.

We also considered a genotype-phenotype classification to integrate all aspects of dairy *L. lactis* subsp. *lactis* diversity. We found that the genotype made only a small contribution to diversity, which depends on large phenotypic differences. Nevertheless, integration of genotypic diversity modified the phenotypic classification and led to a five-cluster organization rather than the four strain clusters generated with only discriminatory phenotypic markers. In particular, this increased the diversity of the strains “not dedicated to a metabolic pathway” of cluster IV. Moreover, “domesticated” strains of *L. lactis* subsp. *lactis* are characterized by noticeable mobility and substantial genomic variability of their genomes (6). This genotypic variability may contribute significantly to variations in the properties of the starter culture. Adding genotypic characteristics would ensure that all of the strain potentialities encoded in the genotype would be covered, including those not expressed in our model dairy process but which may be of relevance to other technological applications (such as osmotic resistance in cheese ripening or phage resistance). The genotypic-phenotypic classification tends toward a predictive classification for starter manufacturing.

Finally, this study illustrates the feasibility of establishing well-characterized strain signatures. This approach could be extended to mixed-culture profiling. Indeed, it may be possible to assess the effect of strain adjunction in cocultures of lactococci by comparing single-strain signatures to the resulting mix signatures. This method may complement traditional trial-and-error methods used to design targeted starters for particular applications and facilitate the rational development of defined mixed cultures.

ACKNOWLEDGMENTS

We acknowledge Helene Tormo from INP-EI Purpan (Toulouse, France) for providing EIP bacterial strains. We thank Gilles De Revel and Laurent Riquier from the Analytical Chemical Service of ISVV for SPME-GC-MS analyses.

This research was supported by Yoplait S.A.

REFERENCES

- Schleifer KH, Kraus J, Dvorak C, Kilpper-Bälz R, Collins MD, Fischer W. 1985. Transfer of *Streptococcus lactis* and related streptococci to the genus *Lactococcus* gen. nov. Syst. Appl. Microbiol. 6:183–195.
- Klijn N, Weerkamp AH, de Vos WM. 1995. Detection and characterization of lactose-utilizing *Lactococcus* spp. in natural ecosystems. Appl. Environ. Microbiol. 61:788–792.
- Kelly W, Ward L. 2002. Genotypic vs. phenotypic biodiversity in *Lactococcus lactis*. Microbiology 148:3332–3333.
- Salama MS, Sandine WE, Giovannoni SJ. 1993. Isolation of *Lactococcus lactis* subsp. *cremoris* from nature by colony hybridization with rRNA probes. Appl. Environ. Microbiol. 59:3941–3945.
- Rademaker JL, Herbert H, Starrenburg MJ, Naser SM, Gevers D, Kelly WJ, Hugenholtz J, Swings J, van Hylckama Vlieg JE. 2007. Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel multilocus sequence analysis scheme and (GTG)₅-PCR fingerprinting. Appl. Environ. Microbiol. 73:7128–7137.
- Passerini D, Beltramo C, Coddeville M, Quentin Y, Ritzenthaler P, Davaan-Mingot ML, Le Bourgeois P. 2010. Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. PLoS One 5:e15306. doi:10.1371/journal.pone.0015306.
- Fernández E, Alegría A, Delgado S, Martín MC, Mayo B. 2011. Comparative phenotypic and molecular genetic profiling of wild *Lactococcus lactis* subsp. *lactis* strains of the *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* genotypes, isolated from starter-free cheeses made of raw milk. Appl. Environ. Microbiol. 77:5324–5335.
- Siezen RJ, Bayjanov JR, Felis GE, van der Sijde MR, Starrenburg M, Molenaar D, Wels M, van Hijum SA, van Hylckama Vlieg JE. 2011. Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays. Microb. Biotechnol. 4:383–402.
- Taibi A, Dabour N, Lamoureux M, Roy D, LaPointe G. 2010. Evaluation of the genetic polymorphism among *Lactococcus lactis* subsp. *cremoris* strains using comparative genomic hybridization and multilocus sequence analysis. Int. J. Food Microbiol. 144:20–28.
- Cachon R, Jeanson S, Aldarf M, Divies C. 2002. Characterisation of lactic starters based on acidification and reduction activities. Lait 82:281–288.
- Brasca M, Morandi S, Lodi R, Tamburini A. 2007. Redox potential to discriminate among species of lactic acid bacteria. J. Appl. Microbiol. 103:1516–1524.
- Kristo E, Biliaderis CG, Tzanetakis N. 2003. Modelling of the acidification process and rheological properties of milk fermented with a yogurt starter culture using response surface methodology. Food Chem. 83:437–446.
- Cheng H. 2010. Volatile flavor compounds in yogurt: a review. Crit. Rev. Food Sci. Nutr. 50:938–950.
- Smit G, Smit BA, Engels WJ. 2005. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. FEMS Microbiol. Rev. 29:591–610.
- Aly S, Flourey J, Piot M, Lortal S, Jeanson S. 2012. The efficacy of nisin can drastically vary when produced in situ in model cheeses. Food Microbiol. 32:185–190.
- Hickey RM, Ross RP, Hill C. 2004. Controlled autolysis and enzyme release in a recombinant lactococcal strain expressing the metalloendopeptidase enterolysin A. Appl. Environ. Microbiol. 70:1744–1748.
- Valence F, Deutsch SM, Richoux R, Gagnaire V, Lortal S. 2000. Autolysis and related proteolysis in Swiss cheese for two *Lactobacillus helveticus* strains. J. Dairy Res. 67:261–271.
- Bachmann H, Starrenburg MJ, Dijkstra A, Molenaar D, Kleerebezem M, Rademaker JL, van Hylckama Vlieg JE. 2009. Regulatory phenotyping reveals important diversity within the species *Lactococcus lactis*. Appl. Environ. Microbiol. 75:5687–5694.
- Bayjanov JR, Molenaar D, Tzeneva V, Siezen RJ, van Hijum SA. 2012. PhenoLink—a web-tool for linking phenotype to ~omics data for bacteria: application to gene-trait matching for *Lactobacillus plantarum* strains. BMC Genomics 13:170. doi:10.1186/1471-2164-13-170.
- Bayjanov JR, Starrenburg MJ, van der Sijde MR, Siezen RJ, van Hijum SA. 2013. Genotype-phenotype matching analysis of 38 *Lactococcus lactis* strains using random forest methods. BMC Microbiol. 13:68. doi:10.1186/1471-2180-13-68.
- Tan-a-ram P, Cardoso T, Davaan-Mingot ML, Kanchanatawee S, Loubière P, Girbal L, Coccagn-Bousquet M. 2011. Assessment of the diversity of dairy *Lactococcus lactis* subsp. *lactis* isolates by an integrated approach combining phenotypic, genomic, and transcriptomic analyses. Appl. Environ. Microbiol. 77:739–748.
- Kütahya OE, Starrenburg MJ, Rademaker JL, Klaassen CH, van Hylckama Vlieg JE, Smid EJ, Kleerebezem M. 2011. High-resolution ampli-

- fied fragment length polymorphism typing of *Lactococcus lactis* strains enables identification of genetic markers for subspecies-related phenotypes. *Appl. Environ. Microbiol.* 77:5192–5198.
23. Taïbi A, Dabour N, Lamoureux M, Roy D, LaPointe G. 2011. Comparative transcriptome analysis of *Lactococcus lactis* subsp. *cremoris* strains under conditions simulating cheddar cheese manufacture. *Int. J. Food Microbiol.* 146:263–275.
 24. Coccagn-Bousquet M, Lindley ND. 1995. Pyruvate overflow and carbon flux within the central metabolic pathways of *Corynebacterium glutamicum* during growth on lactate. *Enzyme Microb. Technol.* 17:260–267.
 25. Nouaille S, Even S, Charlier C, Le Loir Y, Coccagn-Bousquet M, Loubière P. 2009. Transcriptomic response of *Lactococcus lactis* in mixed culture with *Staphylococcus aureus*. *Appl. Environ. Microbiol.* 75:4473–4482.
 26. Tanigawa K, Kawabata H, Watanabe K. 2010. Identification and typing of *Lactococcus lactis* by matrix-assisted laser desorption/ionization–time of flight mass spectrometry. *Appl. Environ. Microbiol.* 76:4055–4062.
 27. Le Bourgeois P, Lautier M, van den Berghe L, Gasson MJ, Ritzenthaler P. 1995. Physical and genetic map of the *Lactococcus lactis* subsp. *cremoris* MG1363 chromosome: comparison with that of *Lactococcus lactis* subsp. *lactis* IL 1403 reveals a large genome inversion. *J. Bacteriol.* 177:2840–2850.
 28. Nelson RJ. 1994. Relationship between phylogeny and pathotype for the bacterial blight pathogen of rice. *Appl. Environ. Microbiol.* 60:3275–3283.
 29. Felsenstein J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
 30. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57:289–300.
 31. Lê Cao K-A, González I, Déjean S. 2009. integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* 25:2855–2856.
 32. Ward JH. 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58:236–244.
 33. Shen H, Huang JZ. 2008. Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivariate Anal.* 99:1015–1034.
 34. David S, van der Rest ME, Driessen AJ, Simons G, de Vos WM. 1990. Nucleotide sequence and expression in *Escherichia coli* of the *Lactococcus lactis* citrate permease gene. *J. Bacteriol.* 172:5789–5794.
 35. Le Bourgeois P, Daveran-Mingot ML, Ritzenthaler P. 2000. Genome plasticity among related *Lactococcus* strains: identification of genetic events associated with macrorestriction polymorphisms. *J. Bacteriol.* 182:2481–2491.
 36. Collins YF, McSweeney PL, Wilkinson MG. 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *Int. Dairy J.* 13:841–866.
 37. Fox PF, Wallace JM. 1997. Formation of flavor compounds in cheese. *Adv. Appl. Microbiol.* 45:17–85.
 38. McSweeney PL, Sousa MJ. 2000. Biochemical pathways for the production of flavour compounds in cheeses during ripening: a review. *Lait* 80:32.
 39. Bemiller J, Huber K. 2007. Carbohydrates, p 83–155. In Damodaran S, Parkin KL, Fennema OR (ed), *Fennema's food chemistry*, 4th ed. CRC Press, New York, NY.
 40. Gutiérrez-Méndez N, Vallejo-Cordoba B, González-Córdova AF, Nevárez-Moorillón GV, Rivera-Chavira B. 2008. Evaluation of aroma generation of *Lactococcus lactis* with an electronic nose and sensory analysis. *J. Dairy Sci.* 91:49–57.
 41. Bayjanov JR, Wels M, Starrenburg M, van Hylckama Vlieg JE, Siezen RJ, Molenaar D. 2009. PanCGH: a genotype-calling algorithm for pan-genome CGH data. *Bioinformatics* 25:309–314.
 42. van Hylckama Vlieg JE, Rademaker JL, Bachmann H, Molenaar D, Kelly WJ, Siezen RJ. 2006. Natural diversity and adaptive responses of *Lactococcus lactis*. *Curr. Opin. Biotechnol.* 17:183–190.
 43. Kelly WJ, Ward LJ, Leahy SC. 2010. Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. *Genome Biol. Evol.* 2:729–744.
 44. Salama MS, Musafija-Jeknic T, Sandine WE, Giovannoni SJ. 1995. An ecological study of lactic acid bacteria: isolation of new strains of *Lactococcus* including *Lactococcus lactis* subspecies *cremoris*. *J. Dairy Sci.* 78:1004–1017.
 45. Hansen EB. 2002. Commercial bacterial starter cultures for fermented foods of the future. *Int. J. Food Microbiol.* 78:119–131.
 46. de Bok FA, Janssen PW, Bayjanov JR, Sieuwerts S, Lommen A, van Hylckama Vlieg JE, Molenaar D. 2011. Volatile compound fingerprinting of mixed-culture fermentations. *Appl. Environ. Microbiol.* 77:6233–6239.
 47. Smit BA, Engels WJ, Bruinsma J, van Hylckama Vlieg JE, Wouters JT, Smit G. 2004. Development of a high throughput screening method to test flavour-forming capabilities of anaerobic micro-organisms. *J. Appl. Microbiol.* 97:306–313.
 48. Chipman D, Barak Z, Schloss JV. 1998. Biosynthesis of 2-aceto-2-hydroxy acids: acetolactate synthases and acetohydroxyacid synthases. *Biochim. Biophys. Acta* 1385:401–419.
 49. Smit BA, Engels WJ, Smit G. 2009. Branched chain aldehydes: production and breakdown pathways and relevance for flavour in foods. *Appl. Microbiol. Biotechnol.* 81:987–999.
 50. Passerini D, Laroute V, Coddeville M, Le Bourgeois P, Loubière P, Ritzenthaler P, Coccagn-Bousquet M, Daveran-Mingot ML. 2013. New insights into *Lactococcus lactis* diacetyl- and acetoin-producing strains isolated from diverse origins. *Int. J. Food Microbiol.* 160:329–336.

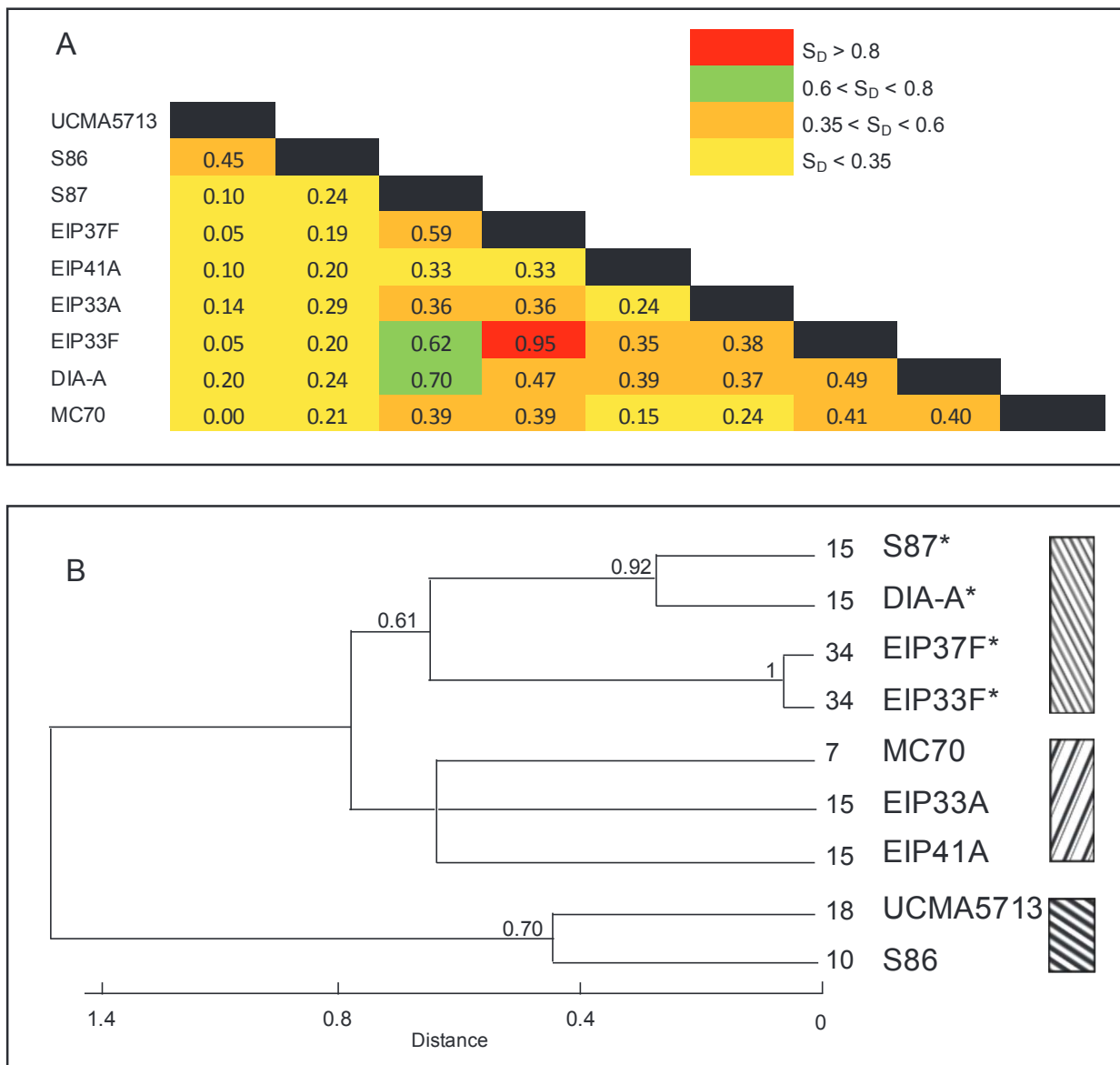


Figure S1: Pulsed field gel electrophoresis (PFGE) of nine strains of *Lactococcus lactis* subsp. *lactis* exhibiting a dairy phenotype. A) Matrix of S_D values for all pairwise comparisons of PFGE fingerprints. B) Consensus tree built from UPGMA classification of S_D values. Bootstrap values above 0.5 ($n=1000$) are indicated on each corresponding node. First column corresponds to ST number, second column to strain name and the third to the three main clusters (shading). Biovar *diacetylactis* (assigned from PCR targeting the plasmid-borne *citP* gene) is indicated with asterisk.

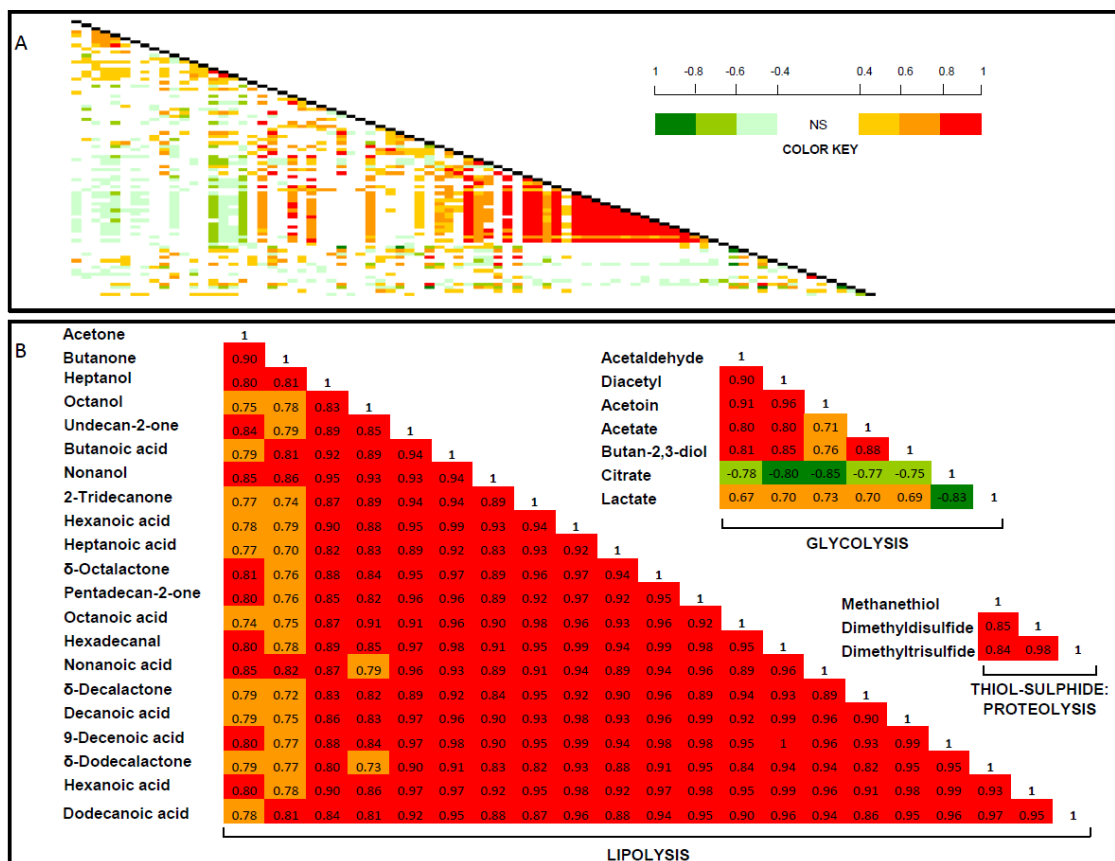


Figure S2: Correlation matrix for the integrative phenotypic dataset (Spearman coefficients with Benjamini & Hochberg multiple testing corrections, p-value cut-off of $5 \cdot 10^{-2}$). A) Integral view of correlation matrix. B) Selected variable clusters for the three main pathways: lipolysis, glycolysis and proteolysis.

