


UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM  
BIOINFORMÁTICA



Dissertação de Mestrado



**Análise do potencial probiótico de *Lactococcus lactis* subsp. *lactis* NCDO 2118 por meio de genômica comparativa**

MESTRANDA: **Letícia de Castro Oliveira**


ORIENTADOR: **Prof. Dr. Vasco Ariston de Carvalho Azevedo**

CO-ORIENTADOR: **Dr. Siomar de Castro Soares**

BELO HORIZONTE

Outubro – 2014

Letícia de Castro Oliveira



**Análise do potencial probiótico de *Lactococcus lactis* subsp. *lactis* NCDO 2118 por meio de genômica comparativa**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo programa Interunidades de Pós-Graduação em Bioinformática, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais.

**Orientador: Prof. Dr. Vasco Ariston de Carvalho Azevedo**

**Co-orientador: Dr. Siomar de Castro Soares**

BELO HORIZONTE

Outubro – 2014

“Se um dia tudo lhe parecer perdido, lembre-se de que você nasceu sem nada, e que tudo que conseguiu foi através de esforços e os esforços nunca se perdem, somente dignificam as pessoas”.

Charlie Chaplin

## AGRADECIMENTOS

O motivo pelo qual cheguei até aqui primeiramente, foi graças a Deus e aos meus pais, que me proporcionaram a vida (a todo custo) e a força necessária para lutar sempre;

Agradeço ao meu orientador, Prof. Dr. Vasco Azevedo, por ter me recebido junto à equipe LGCM e me proporcionado a oportunidade de fazer parte da primeira turma de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais, me orientando e compartilhando sua experiência conosco no convívio diário;

Ao meu co-orientador, Siomar Soares, pela paciência e compreensão e por entender e me ajudar a enfrentar minhas dificuldades depois destes anos longe da vida acadêmica;

Aos membros da banca, por terem aceitado o convite em participar da avaliação deste trabalho, contribuindo para o seu enriquecimento;

Ao grupo do Laboratório de Polimorfismo de DNA, no Pará, que vem colaborando com nosso grupo desde os primeiros trabalhos feitos com *Lactococcus*;

Ao grupo do AQUACEN, também pela colaboração no desenvolvimento deste trabalho;

Às queridas, Sheila, Márcia Natália, Fernanda e Ana Paula, por serem sempre atenciosas, prestativas e carinhosas, não importando quão cinza esteja o dia;

À Bianca Mendes, por ter compartilhado comigo os primeiros artigos que precisei ao iniciar a escrita do meu manuscrito e também, por sua amizade;

Aos amigos, Fernanda Dorella, Adriana Carneiro e Rommel Ramos, pelo auxílio com os materiais que puderam completar ainda mais este trabalho e por todo carinho que sempre tiveram comigo;

À Tessália Saraiva, pelas sugestões e por trocar um pouco de sua experiência no desenvolvimento deste trabalho;

À Flávia Souza, pelas micrografias de *Lactococcus*, por sua amizade, preocupação e companheirismo;

Ao Wanderson e Ulisses, pelo auxílio na obtenção dos dados para as análises de proteômica, e pelo companheirismo e amizade de vocês;

Aos meus amigos de caminhada, Lucas e Caçula, por terem compartilhado desde os mesmos sofrimento e nervosismo, às risadas incontroláveis. Foi muito importante pra mim, poder contar firmemente com vocês durante este tempo e espero que assim continue. Pela amizade que se transformou em “quarteto fantástico”, juntamente com o quarto elemento (Carioca), que estava faltando;

À equipe LGCM, que é família, pau pra toda obra, é parceira e muito especial pra mim. Deixo meu agradecimento a cada um que partilhou comigo tanta força, amizade, companheirismo, torcida e toda energia boa que vocês têm, em especial ao nosso legendário indiano, Sandeep;

Aos amigos e colegas do Hemocentro Regional de Uberaba, que torceram e me apoiaram para que eu continuasse meus estudos, em especial ao Dr. Hélio Moraes, Marise, Gláucia Resende, Renata Volpe, Fernanda Garcia, Aline Ferreira e Darci Castro;

Aos meus pais, Sebastião e Sirlene, por todos os esforços que fizeram para que eu e meus irmãos tivéssemos uma boa educação e um bom ensino. Vocês são meus exemplos de garra, perseverança e amor;

Agradeço aos meus irmãos, Aline e Bruno, pelo apoio, torcida e por compartilharem, mesmo que de longe, todo companheirismo e carinho, e aos meus sobrinhos Noah e Mariah, por trazerem tanta luz e alegria à nossa família.

Ao meu tio Fernando (*in memoriam*);

À minha família Soares, por toda força, apoio, carinho e alegria que sempre recebo de vocês;

Aos meus amigos da faculdade, que estão longe, mas mais presentes do que nunca, torcendo sempre por mim: Vespa, Maregueza, Samucovits, Fatolina, Vanesquês e Sacramen.

Ao meu melhor amigo, parceiro e companheiro, Siomar Soares. Por ser este pesquisador criativo, ousado e ao mesmo tempo cuidadoso; não deixe que nada tire isso de você, pois foram estas características que me inspiraram a enfrentar o caminho da pesquisa, que só está começando e não é nada fácil. Agradeço todo seu apoio, sacrificando seus *deadlines* e seu tempo de descanso para me guiar neste trabalho. Sou muito feliz e sortuda por ter você ao meu lado!

## SUMÁRIO

|  |     |
|--|-----|
| <b>LISTA DE FIGURAS</b> .....  | I   |
| <b>LISTA DE TABELAS</b> .....  | II  |
| <b>LISTA DE ABREVIATURAS</b> .....   | III |
| <b>RESUMO</b> .....  | 1   |
| <b>ABSTRACT</b> .....  | 2   |
| <b>1. INTRODUÇÃO</b> .....   | 3   |
| 1.1 Estado da Arte .....   | 3   |
| 1.2 Sequenciamento Genômico .....  | 7   |
| 1.3 A genômica das Bactérias Láticas .....   | 8   |
| 1.4 Plasticidade genômica .....  | 9   |
| 1.5 As Bactérias Láticas .....   | 12  |
| 1.5.1 Definição .....  | 12  |
| 1.5.2 Importância industrial e biotecnológica .....  | 13  |
| 1.5.3 Novas utilizações das Bactérias Láticas .....  | 14  |
| 1.5.4 Efeito probiótico das Bactérias Láticas .....  | 16  |
| 1.5.4.1 Sobrevivência através do trato gastrointestinal .....                                  | 18  |
| 1.5.4.2 Atividade antimicrobiana e exclusão competitiva .....                                  | 19  |
| 1.5.4.3 Capacidade imunomodulatória .....  | 22  |
| 1.5.5 Produção do ácido gama-aminobutírico .....   | 23  |
| <b>2. JUSTIFICATIVA</b> .....  | 24  |
| <b>3. OBJETIVOS</b> .....  | 25  |
| 3.1 Objetivos específicos .....  | 25  |
| <b>4. MATERIAIS E MÉTODOS</b> .....  | 26  |
| 4.1 Obtenção de DNA da linhagem <i>Lactococcus lactis</i> subsp. <i>lactis</i> NCDO 2118 ..... | 26  |
| 4.2 Química do sequenciamento .....  | 26  |
| 4.3 Sequenciamento .....   | 28  |
| 4.4 Montagem .....   | 31  |

|   |           |
|---|-----------|
| 4.5 Anotação e curadoria de <i>frameshifts</i> . . . . .                                      | 32        |
| 4.6 Comparação genômica. . . . .  | 35        |
| 4.6.1 Filogenômica. . . . .   | 37        |
| 4.6.2 Sintenia gênica. . . . .  | 38        |
| 4.6.3 Predição de vias metabólicas. . . . .   | 39        |
| 4.6.4 Plasticidade genômica. . . . .  | 39        |
| 4.6.5 Predição de profagos. . . . .   | 41        |
| 4.6.6 Predição de bacteriocinas. . . . .  | 41        |
| 4.6.7 Representação circular da comparação das sequências genômicas. . . . .                  | 41        |
| 4.6.8 Identificação das proteínas do secretoma. . . . .                                       | 42        |
| <b>5. RESULTADOS E DISCUSSÃO. . . . .</b>   | <b>43</b> |
| 5.1 Montagens. . . . .  | 43        |
| 5.2 Anotação e curadoria de <i>frameshifts</i> . . . . .                                      | 46        |
| 5.3 Comparação genômica. . . . .  | 48        |
| 5.3.1 Análise comparativa através de similaridade nucleotídica e filogenômica. . . . .        | 48        |
| 5.3.2 Análise comparativa através de sintenia gênica. . . . .                                 | 52        |
| 5.3.3 Predição e análise de fagos em <i>L. lactis</i> NCDO 2118. . . . .                      | 56        |
| 5.3.4 Predição de vias metabólicas. . . . .   | 59        |
| 5.3.5 Predição e análises de ilhas genômicas. . . . .   | 62        |
| 5.4 Análises probióticas de <i>L. lactis</i> NCDO 2118. . . . .                               | 74        |
| 5.4.1 Identificação de genes relacionados à resistência ao estresse ácido e à adesão. . . . . | 74        |
| 5.4.2 Predição e análise de bacteriocinas. . . . .  | 77        |
| 5.4.3 Identificação <i>in silico</i> de prováveis proteínas secretadas . . . . .              | 80        |
| <b>6. CONCLUSÕES. . . . .</b>   | <b>82</b> |
| <b>7. PERSPECTIVAS. . . . .</b>   | <b>83</b> |
| <b>8. REFERÊNCIAS BIBLIOGRÁFICAS. . . . .</b>   | <b>84</b> |
| <b>9. ANEXOS. . . . .</b>   | <b>99</b> |

## LISTA DE FIGURAS

|   |    |
|---|----|
| <b>Figura 1</b> Marcos da história que contribuíram na caracterização funcional das bactérias lácticas probióticas . . . . .  | 4  |
| <b>Figura 2</b> Integração das áreas ômicas. . . . .  | 10 |
| <b>Figura 3</b> Filogenética das bactérias lácticas . . . . .   | 13 |
| <b>Figura 4</b> Figura representativa das vias metabólicas de consumo da glicose pelas bactérias lácticas. . . . .  | 14 |
| <b>Figura 5</b> Mecanismos pelos quais as bacteriocinas podem contribuir para a funcionalidade probiótica. . . . .  | 20 |
| <b>Figura 6</b> Possíveis dinucleotídeos codificados por cada cor. . . . .  | 27 |
| <b>Figura 7</b> Pipeline dos processos de sequenciamento, montagem e anotação de <i>L. lactis</i> subsp. <i>lactis</i> NCDO 2118 . . . . .                                | 28 |
| <b>Figura 8</b> Pipeline das etapas de comparação genômica do gênero <i>Lactococcus</i> . . . . .   | 35 |
| <b>Figura 9</b> Sintenia entre <i>L. lactis</i> subsp. <i>lactis</i> NCDO 2118 ( <i>draft</i> ) e <i>L. lactis</i> subsp. <i>lactis</i> KF147. . . . .                    | 43 |
| <b>Figura 10</b> Sintenia entre <i>L. lactis</i> subsp. <i>lactis</i> KF147 e <i>L. lactis</i> subsp. <i>lactis</i> NCDO 2118 (genoma completo) . . . . .                 | 44 |
| <b>Figura 11</b> Curadoria manual de <i>frameshift</i> . . . . .  | 47 |
| <b>Figura 12</b> Análise filogenômica do gênero <i>Lactococcus</i> . . . . .  | 51 |
| <b>Figura 13</b> Análise de sintenia gênica entre <i>Lactococcus lactis</i> subsp. <i>cremoris</i> . . . . .  | 54 |
| <b>Figura 14</b> Análise de sintenia gênica entre <i>Lactococcus garvieae</i> . . . . .   | 54 |
| <b>Figura 15</b> Análise de sintenia gênica entre <i>Lactococcus lactis</i> subsp. <i>lactis</i> . . . . .  | 55 |
| <b>Figura 16</b> Análise de sintenia gênica entre os plasmídeos de <i>L. lactis</i> subsp. <i>lactis</i> NCDO 2118 e <i>L. lactis</i> subsp. <i>lactis</i> KF147. . . . . | 55 |
| <b>Figura 17</b> Regiões de profagos preditas em <i>L. lactis</i> subsp. <i>lactis</i> NCDO 2118. . . . .   | 58 |
| <b>Figura 18</b> Comparação entre as sequências do genoma de <i>L. garvieae</i> Lg2 e das outras espécies do gênero <i>Lactococcus</i> . . . . .                          | 59 |
| <b>Figura 19</b> Comparação entre as sequências do genoma de <i>L. lactis</i> subsp <i>cremoris</i> MG1363 e as outras espécies do gênero <i>Lactococcus</i> . . . . .    | 62 |
| <b>Figura 20</b> Comparação entre as sequências do genoma de <i>L. lactis</i> subsp. <i>lactis</i> NCDO 2118 e as outras espécies do gênero <i>Lactococcus</i> . . . . .  | 64 |
| <b>Figura 21</b> Predição de bacteriocinas para <i>L. lactis</i> subsp. <i>lactis</i> NCDO 2118. . . . .  | 66 |
| <b>Figura 22</b> Comparação entre as sequências do genoma de <i>L. lactis</i> NCDO 2118 e as outras espécies do gênero <i>Lactococcus</i> . . . . .                       | 67 |
| <b>Figura 23</b> Predição de bacteriocinas para <i>L. lactis</i> NCDO 2118. . . . .   | 79 |



## LISTA DE TABELAS

|   |    |
|---|----|
| <b>Tabela 1</b> Atual aplicação das bactérias lácticas. ....  | 15 |
| <b>Tabela 2</b> Lista dos genomas completos utilizados para comparação genômica. ....   | 36 |
| <b>Tabela 3</b> Resultados das diferentes abordagens de sequenciamento, montagem e anotação do genoma de <i>L. lactis</i> NCDO 2118. .... | 43 |
| <b>Tabela 4</b> Genes e proteínas presentes nas regiões de fagos preditas pelo PHAST. . . . .   | 57 |
| <b>Tabela 5</b> Ilhas simbióticas e metabólicas preditas para <i>L. lactis</i> NCDO 2118. ....  | 68 |
| <b>Tabela 6</b> Ilhas metabólicas preditas para <i>L. lactis</i> NCDO 2118. ....  | 71 |
| <b>Tabela 7</b> Genes codificantes de proteínas envolvidas na resistência ao estresse ácido. . .  | 74 |
| <b>Tabela 8</b> Proteínas potencialmente envolvidas nos mecanismos de adesão de <i>L. lactis</i> ao hospedeiro. ....                      | 76 |
| <b>Tabela 9</b> Codificantes de bacteriocinas, muramidases e macrolídeos. ....  | 78 |
| <b>Tabela 10</b> Predição de proteínas secretadas exclusivas de <i>L. lactis</i> NCDO 2118. ....  | 80 |

## LISTA DE ABREVIATURAS

|              |  |
|--------------|--|
| <b>BL</b>    | bactérias lácticas                       |
| <b>BLAST</b> | <i>basic local alignment search tool</i> |
| <b>BRIG</b>  | <i>blast ring image generator</i>        |
| <b>CDS</b>   | <i>coding DNA sequence</i>               |
| <b>2CRS</b>  | <i>two-component regulatory systems</i>  |
| <b>CTAB</b>  | brometo de cetiltrimetil amônio          |
| <b>dNTP</b>  | desoxiribonucleotídeo                    |
| <b>EC</b>    | <i>enzyme commission</i>                 |
| <b>ECM</b>   | componentes da matriz extracelular       |
| <b>EDTA</b>  | ácido etilenodiaminotetracético          |
| <b>Fbp</b>   | <i>fibronectin binding protein</i>       |
| <b>Fn</b>    | fibronectina                             |
| <b>FTP</b>   | <i>file transfer protocol</i>            |
| <b>GABA</b>  | ácido gama-aminobutírico                 |
| <b>GAD</b>   | ácido glutâmico descarboxilase           |
| <b>Gb</b>    | gigabase                                 |
| <b>GEI</b>   | ilha genômica                            |
| <b>GIT</b>   | trato gastrointestinal                   |
| <b>GPCR</b>  | <i>G-protein coupled receptors</i>       |
| <b>GRAS</b>  | <i>generally recognized as safe</i>      |
| <b>G+C</b>   | conteúdo G+C do genoma                   |
| <b>HPK</b>   | proteína histidina quinase               |
| <b>IBD</b>   | doença inflamatória intestinal           |
| <b>IEC</b>   | células epiteliais intestinais           |
| <b>IL</b>    | interleucina                             |

|                 |   |
|-----------------|---|
| <b>Kb</b>       | kilobase  |
| <b>kDa</b>      | kilodalton  |
| <b>Lan</b>      | lantionina  |
| <b>MAMPS</b>    | <i>microbial-associated molecular patterns</i>                            |
| <b>Mb</b>       | megabase  |
| <b>MeLan</b>    | metilantionina  |
| <b>MEM</b>      | proteína de membrana  |
| <b>MHC</b>      | complexo principal de histocompatibilidade                                |
| <b>MI</b>       | ilha metabólica   |
| <b>Mub</b>      | mucina  |
| <b>MSCRAMM</b>  | <i>microbial surface components recognizing adhesive matrix molecules</i> |
| <b>NCBI</b>     | <i>national center for biotechnology information</i>                      |
| <b>NGS</b>      | <i>next generation sequencing</i>   |
| <b>Nm</b>       | nanômetros  |
| <b>NICE</b>     | <i>nisin controlled gene expression system</i>                            |
| <b>NJ</b>       | <i>neighbor-joining</i>   |
| <b>NOD-like</b> | <i>nucleotide oligomerization domain-like receptor</i>                    |
| <b>NR</b>       | não redundante  |
| <b>ORF</b>      | <i>open reading frame</i>   |
| <b>PAI</b>      | ilha de patogenicidade  |
| <b>pb</b>       | par de base   |
| <b>PCR</b>      | <i>polymerase chain reaction</i>  |
| <b>pH</b>       | potencial de hidrogênio   |
| <b>PRR</b>      | receptor de reconhecimento de padrão                                      |
| <b>PSE</b>      | proteínas expostas na superfície  |
| <b>Pxylt</b>    | gene da xilose permease   |
| <b>RBS</b>      | sítio de ligação ao ribossomo   |

|                                |  |
|--------------------------------|--|
| <b>RI</b>                      | ilha de resistência                        |
| <b>RR</b>                      | regulador de resposta                      |
| <b>SDS</b>                     | dodecil fosfato de sódio                   |
| <b>SI</b>                      | ilha simbiótica                            |
| <b>SlpA</b>                    | proteína <i>S-layer</i>                    |
| <b>SP</b>                      | peptídeo sinal                             |
| <b>ScFvs</b>                   | <i>Single-chain variable fragment</i>      |
| <b>TE</b>                      | tris EDTA                                  |
| <b>TLR</b>                     | receptores semelhantes ao <i>Toll</i>      |
| <b>TNF-<math>\alpha</math></b> | fator de necrose tumoral $\alpha$          |
| <b>Usp45</b>                   | <i>unknown secreted protein of 45 kDa</i>  |
| <b>XIES</b>                    | sistema de expressão por indução de xilose |

## RESUMO

As bactérias lácticas (BL) compõem um dos grupos bacterianos mais importantes da área biotecnológica. Nele, *Lactococcus lactis* se destaca pelo seu uso na produção de produtos fermentados, macromoléculas e na sua aplicação de benefícios à saúde, uma vez que são espécies geralmente reconhecidas como seguras (GRAS – do inglês: *generally recognized as safe*). Neste cenário, *Lactococcus lactis* subsp. *lactis* NCDO 2118 (NCDO 2118) se sobressai como uma linhagem fermentadora de xilose e produtora de ácido gama-aminobutírico (GABA) isolada de ervilha congelada, além de sua capacidade anti-inflamatória, comprovada recentemente. O GABA pode contribuir no relaxamento muscular, além de apresentar atividade hipotensora. Entretanto, apesar destas características importantes, pouco ainda se conhece sobre os mecanismos envolvidos nos efeitos probióticos da linhagem. Neste contexto, a área da genômica pode auxiliar esclarecendo estas questões. A genômica comparativa tem revolucionado os novos estudos, por meio da determinação da plasticidade genômica entre diferentes espécies. Neste trabalho, o genoma da linhagem NCDO 2118 foi sequenciado e curado manualmente. Além disso, comparamos o seu genoma com outras linhagens, onde 5 são de *L. lactis* subsp. *lactis*, 6 *Lactococcus lactis* subsp. *cremoris* e 2 *Lactococcus garvieae*. Foram utilizados os softwares Gegenees, Mauve e BRIG para: análises filogenômicas, sintenia gênica e comparações de genomas entre as espécies anteriormente mencionadas. Foram trabalhadas em conjunto, as ferramentas GIPSY e Phast, para predição de ilhas genômicas (GEI) e fagos, respectivamente. Para predição de proteínas secretadas, o genoma foi analisado por meio da ferramenta SurfG+. Nas análises filogenômicas, uma alta similaridade foi observada entre NCDO 2118 e *Lactococcus lactis* subsp. *lactis* KF147 (KF147), ambas isoladas de plantas. Além das análises de sintenia terem mostrado claramente o alto grau de conservação da ordem gênica entre ambas linhagens. Foram preditas também, 15 ilhas genômicas, onde 4 delas são ilhas metabólicas (MI) e 7 são ilhas simbióticas (SI). As quatro MI são compartilhadas em comum com todas as linhagens analisadas aqui, enquanto duas das SI são compartilhadas pelas NCDO 2118 e KF147. Dentre os 98 genes relacionados a bacteriocinas (que podem auxiliar a regulação e secreção destas), identificamos 3, uma de cada classe (classes I, II e III); além disso, identificamos 5 regiões de fago, sendo 3 delas preditas como intactas. As proteínas preditas como secretadas em NCDO 2118 foram comparadas com as proteínas de *Lactococcus lactis* subsp. *lactis* IL1403, considerada, em estudos recentes, como uma linhagem que não possui atividade probiótica. Assim, foram preditas 34 proteínas exclusivas de NCDO 2118, e neste conjunto, alguma(s) dela(s) pode(m) estar relacionadas com a sua capacidade probiótica. No geral, o alto grau de similaridade entre todas as linhagens apontam que as SI compartilhadas em comum entre NCDO 2118 e KF147 foram responsáveis pela relação próxima nas análises filogenômicas e também pela adaptação destas linhagens a plantas. As MI, por outro lado, são altamente conservadas entre as linhagens, como esperado, dado o uso destas em vários processos metabólicos industriais. Finalmente, as três classes de bacteriocinas podem ter um papel de grande importância contra invasão de linhagens competitivas ou influência no sistema imune do hospedeiro, além de poder estar envolvidas na característica probiótica desta linhagem.

**Palavras-chave:** genômica comparativa, *Lactococcus lactis* subsp. *lactis* NCDO 2118, GABA, probiótico.

## ABSTRACT

Lactic Acid Bacteria (LAB) compose one of the most biotechnologically important group of bacteria. In this group, *Lactococcus lactis* is highly important for its use in the production of many fermented products and macromolecules and its application in health improvement it is a Generally Regarded As Safe (GRAS) species. Since this scenario, *Lactococcus lactis* subsp. *lactis* NCDO 2118 (herein, NCDO2118) stands out as a xylose fermenter and GABA (gamma-aminobutyric acid) producer strain isolated from frozen peas. GABA can contribute to smooth muscle relaxation and present a hypotensor activity. However, despite of those important characteristics, few is known about the mechanisms involved in probiotic effects of this strain. In this context, the genomic area can help elucidating unresolved issues. The comparative genomics has revolutionized the recent studies, through determining the genome plasticity between different species and/or strains, helping in driving experimental studies. In this work, we have sequenced and manually curated the genome of NCDO2118. Besides, we have compared the genome of NCDO2118 with those of 5 additional *L. lactis* subsp. *lactis*, 6 *Lactococcus lactis* subsp. *cremoris* and 2 *Lactococcus garvieae* strains. We used the software Gegenees, Mauve and BRIG to perform phylogenomics, gene synteny and circular genome comparisons between the above mentioned species. Additionally, we have used GIPSy and Phast to predict genomic islands (GEIs) and phages, respectively. We have predicted secreted proteins using SurfG+ tool. In phylogenomics analyses, a high similarity was observed between NCDO 2118 and *L. lactis* subsp. *lactis* KF147 (herein, KF147), both isolated from plants. Besides, the gene synteny analyses have clearly shown a highly conserved gene order between both strains. We have also predicted 15 putative GEIs, where four are metabolic islands (MIs) and seven are symbiotic islands (SIs). The four MIs are commonly shared by all strains analyzed here, whereas two of the SIs are only shared by NCDO2218 and KF147. Finally, we have identified five phage regions and 3 bacteriocins, one of each class(classes I, II and III), along with 98 bacteriocin-related genes, which may support the regulation and secretion of the bacteriocins. The predicted secreted proteins of NCDO 2118 were compared with *Lactococcus lactis* subsp. *lactis* IL1403 proteins, considered recently as a non-probiotic strain. Therefore, 34 proteins were present exclusively in NCDO 2118 and some of them can have relationship with its probiotic ability. Altogether, the high degree of similarity between all strains point that the SIs commonly shared by both NCDO2118 and KF147 were responsible for the close relationship in phylogenomic analyses and also for the adaptation of those strains to plants. The MIs, on the other hand, are highly conserved between all strains, which is a expected feature given the use of those strains in metabolic processes in industry. Finally, the three classes of bacteriocins may have an important role against invasion of competing strains or influencing the host immune system, which may be involved in the probiotic characteristic of this strain.

**Keywords:** comparative genomics, NCDO 2118, GABA, probiotic.

---

## 1. INTRODUÇÃO

---

### 1.1 ESTADO DA ARTE

Durante muito tempo, considerou-se que o consumo de bactérias alóctones (transientes, benéficas à saúde), provenientes de suplementos ou produtos, tem influenciado de forma positiva na saúde em geral e no bem estar, por meio de interações com o sistema imune do trato gastrointestinal (GIT) e a microbiota residente (Johnson & Klaenhammer, 2014). Estes organismos, conhecidos como probióticos, foram definidos como “microrganismos vivos” pela Organização Mundial de Saúde, os quais sendo administrados adequadamente, proporcionam benefícios à saúde do hospedeiro (FAO & WHO, 2002).

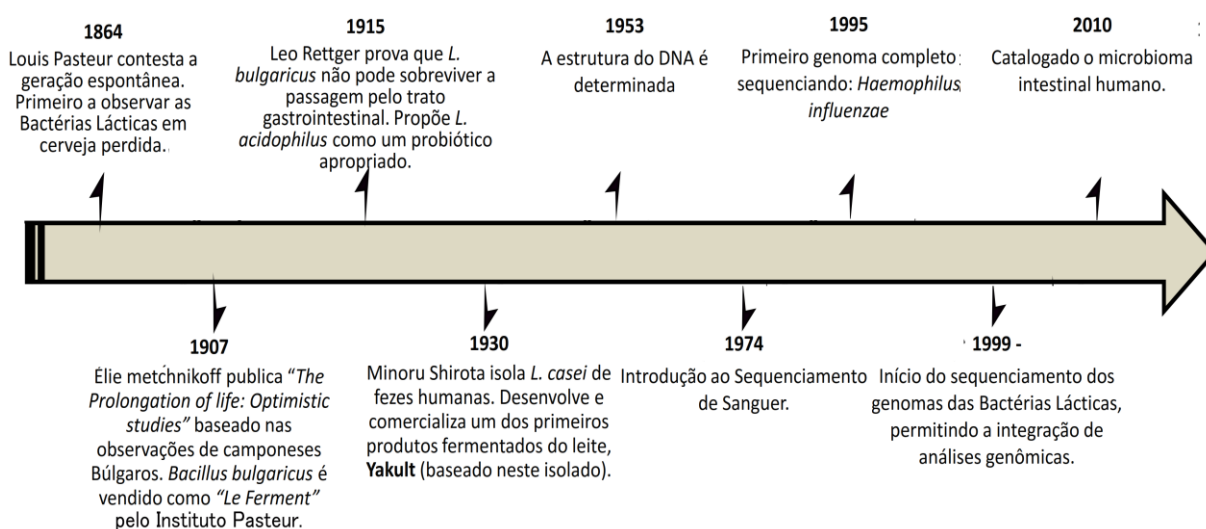
O consumo de microrganismos (com capacidade de evitar o processo de putrefação) vem sendo feito pelos humanos por meio de alimentos fermentados, durante muitos anos. É muito provável que os fermentados surgiram devido à contaminação, condição climática e um ambiente favorável, resultando em produtos como *kefir*, *leben*, *koumiss*, iogurte e creme azedo (Hosono, 1992). Além de serem produtos seguros para o consumo, os alimentos fermentados tiveram grande importância cultural. Estes alimentos datam do início da civilização e sua menção é encontrada na Bíblia e nos livros sagrados do Hinduísmo (Hosono, 1992; Bibel, 1988; Shortt, 1999).

No século XIX, o bioquímico francês Louis Pasteur foi premiado por suas descobertas, que o levaram ao melhor entendimento científico sobre fermentação. Estudando sobre os processos fermentativos do vinho e da cerveja, Pasteur demonstrou que as reações presentes eram realizadas por microrganismos. Ademais, através de seus estudos, foi transformada a ideia de que os micróbios cresciam por geração espontânea, como prevalecia até então, pois Pasteur demonstrou que este processo era devido à biogênese, a qual postula que tudo que é vivo vem apenas de outras coisas vivas (Johnson & Klaenhammer, 2014).

Na mesma fundação de pesquisas de Pasteur, um russo foi premiado com Nobel. Élie Metchnikoff se destacou com a descoberta da fagocitose e, além disso, foi o primeiro a popularizar o conceito de probiótico por volta da virada do século XX. Em seu livro *The Prolongation of Life: Optimistic Studies* (**Fig. 1**), Metchnikoff propôs que as bactérias envolvidas na fermentação do iogurte, *Lactobacillus bulgaricus* e

*Streptococcus thermophilus* suprimiam a fermentação do tipo putrefativa da flora intestinal e participavam da manutenção da saúde (Johnson & Klaenhammer, 2014). Esta teoria foi reforçada baseada em observações de alta prevalência de centenários na Bulgária, região conhecida pelo seu consumo de leite acidificado. Um colega de Pasteur, Michael Cohendy forneceu dados que enfatizariam a hipótese de Metchnikoff. Em seus experimentos com seres humanos, Cohendy descobriu que *Bacillus bulgaricus* (agora conhecido como *Lactobacillus delbrueckii* subsp. *bulgaricus*) além de serem recuperados das fezes, poderiam também, reduzir a prevalência de toxinas de putrefação e auxiliarem no tratamento de colite após serem transplantados para o intestino grosso (Cohendy, 1906a; Cohendy, 1906b).

Apesar das observações feitas por Metchnikoff e seus colaboradores a respeito do conceito probiótico, ainda existia poucos indícios científicos sugerindo uma linhagem específica probiótica ou o conhecimento do seu suposto mecanismo efetor. Em 1915, Leo Rettger e colaboradores observaram a incapacidade de *L. bulgaricus* de sobreviver à passagem pelo trato gástrico, colonizando o intestino delgado (Rettger, 1915). Este estudo colocou em questão quais linhagens estariam presentes na administração terapêutica original, dos estudos realizados por Cohendy e que foi, por conseguinte, vendido como “*Le Ferment*” (Rettger & Cheplin, 1921).



**Fig. 1** Marcos da história que contribuíram na caracterização funcional das bactérias lácticas probióticas. (Adaptado de Johnson & Klaenhammer, 2014).

Frente às crescentes evidências experimentais a respeito das bactérias probióticas, o físico japonês Minoru Shirota, isolou linhagens humanas derivadas de *Lactobacillus* para aplicações terapêuticas. Em 1930, Shirota selecionou uma



espécie de *Lactobacillus*, conhecida hoje como *Lactobacillus casei* Shirota, de fezes humanas, que sobreviveram ao GIT com o objetivo de utilizá-las no desenvolvimento do leite fermentado para distribuição em sua clínica. Seu primeiro produto contendo *L. acidophilus* Shirota, chamado posteriormente de *L. casei* Shirota, foi base do primeiro produto de leite fermentado, *Yakult* (Shortt, 1999). Desde então, o mercado de alimentos se expandiu, principalmente no que diz respeito aos produtos lácteos contendo bactérias probióticas (Sanders & in't Huis-Veld, 1999). Embora os produtos lácteos fermentados sejam os mais conhecidos pelos consumidores, as bactérias lácticas (BL) também são utilizadas na preservação de produtos vegetais, grãos e carnes (Chaillou *et al.*, 2005).

Existem duas subespécies de *Lactococcus lactis* (inicialmente definidas como *Streptococcus lactis* e *Streptococcus cremoris*), reclassificadas como *Lactococcus lactis* subsp. *lactis* e *Lactococcus lactis* subsp. *cremoris*, respectivamente (Schleifer *et al.*, 1985). A primeira subespécie é mais utilizada na fabricação de queijos macios e a última para queijos duros. Ambas têm sido muito estudadas, principalmente devido ao interesse industrial, e têm se tornado excelentes modelos para pesquisas em fisiologia, genética, metabolismo e biologia molecular das BL (Bolotin *et al.*, 2001).

A eficácia do uso de *Lactococcus* na indústria de produtos lácteos requer um melhor entendimento dos aspectos fisiológicos da bactéria, como: o mecanismo de conversão de açúcares à lactato, o uso dos açúcares e proteínas do leite para o crescimento, a ação de síntese de substâncias envolvidas nos diversos sabores de queijo e, assim, a relação entre os diferentes tipos de fermentação (Bolotin *et al.*, 2001).

Além disso, os probióticos estão se expandindo do mercado de alimentos funcionais para o mercado farmacêutico e terapêutico. Esta expansão cresce de acordo com os avanços dos aspectos científicos e regulatórios dos mecanismos probióticos e de entrega (Foligné *et al.*, 2013).

Um exemplo das novas utilizações dos probióticos pode ser observado no tratamento de distúrbios funcionais do trato gastrointestinal (FGID). Para muitas destas, como a síndrome do intestino irritável (IBS), existem poucas opções de tratamentos farmacológicos, devido à baixa eficácia e aos graves efeitos colaterais (Shen & Nahas, 2009). Além disso, a IBS é bastante comum e acredita-se ser causada por mudanças do microbiota gastrointestinal (Porter *et al.*, 2011).

## **A linhagem *Lactococcus lactis* subsp. *lactis* NCDO 2118**

Os projetos desenvolvidos no Laboratório de Genética Celular e Molecular (LGCM) com a linhagem *Lactococcus lactis* subsp. *lactis* NCDO 2118 (aqui, *L. lactis* NCDO 2118), foram possíveis em virtude da colaboração entre Brasil e França. Esta linhagem veio de uma das coleções da Unidade de Genética Microbiana do *Institut National de La Recherche Agronomique* (INRA), em Jouy-en-Josas na França, e hoje, faz parte da bacterioteca do LGCM.

*L. lactis* NCDO 2118 é um isolado de ervilha congelada e vem sendo utilizada na rotina laboratorial do LGCM para produção de proteínas heterólogas. Esta linhagem foi descrita por Mazzoli e colaboradores, em 2010, como produtora do ácido gama-aminobutírico (GABA) (Mazzoli *et al.*, 2010), conhecido por agir positivamente na saúde humana. O GABA é um neurotransmissor distribuído de forma abundante no sistema nervoso central dos vertebrados e é capaz de diminuir a pressão sanguínea de pacientes levemente hipertensos (Inoue *et al.*, 2003), induzir um efeito tranquilizante e diurético (Jakobs *et al.*, 1993; Wong *et al.*, 2003), prevenir diabetes (Adeghate & Ponery, 2002), além de reduzir a resposta inflamatória de artrite reumatóide em modelo murino (Tian *et al.*, 2011).

Dos trabalhos realizados com esta linhagem, merecem destaque os projetos desenvolvidos pelo grupo LGCM. Em um deles, Miyoshi e colaboradores desenvolveram um sistema de expressão e endereçamento proteico utilizando *L. lactis* NCDO 2118. Este, se mostrou eficaz e duradouro quanto à produção e secreção dos níveis da proteína nuclease B de *Staphylococcus aureus*, por meio de um promotor *Pxylt* (gene da xilose permease) e da proteína Usp45 (Miyoshi *et al.*, 2004). Em outro trabalho, Luerce e colaboradores demonstraram a atividade anti-inflamatória e imunomodulatória de *L. lactis* NCDO 2118 por meio da avaliação de três linhagens de *L. lactis in vitro*, com o uso de células epiteliais intestinais (Luerce *et al.*, 2014).

Compreendendo a importância prática das BL no processo fermentativo e em outras áreas de aplicação: bioprocessamento, agricultura, alimentos e, atualmente, na medicina, este grupo de bactérias tem sido o objeto de várias pesquisas considerando seu uso no comércio durante décadas. Auxiliando estes estudos, grandes esforços foram e estão sendo feitos para determinar as sequências genômicas das espécies e linhagens representativas das BL (Klaenhammer *et al.*, 2002).

## 1.2 SEQUENCIAMENTO GENÔMICO

As realizações de Sanger e Coulson (1975) e Maxam e Gilbert (1977) foram um marco no sequenciamento genômico. A tecnologia de sequenciamento de terminação de cadeia dideoxi de Sanger forneceu, por décadas, a base para as pesquisas fundamentadas em sequências, conhecida como tecnologia de sequenciamento de DNA e referida comumente como sequenciamento de Sanger (Kaur & Malik, 2013).

Na área de bacteriologia, a era genômica teve início em 1995 com o trabalho de Fleischmann e colaboradores através do primeiro sequenciamento completo do genoma da bactéria *Haemophilus influenzae* por meio da técnica de Sanger (Fleischmann *et al.*, 1995). Desde então, grandes orçamentos e anos de esforços têm sido necessários para o desenvolvimento dos projetos de sequenciamento. Dez anos depois, em 2005, com o advento da primeira tecnologia de sequenciamento *high throughput* ou NGS (sequenciamento de próxima geração), houve um significativo avanço no custo, e a facilidade do sequenciamento favoreceu a obtenção das *reads* (leituras) dos genomas (Loman *et al.*, 2012).

Dentre as tecnologias de NGS, podemos destacar as plataformas de segunda geração: 454 GS FLX *system* (Roche), Illumina GA IIx (Illumina), HeliScope (Helicos) e SOLiD 5500 XL *system* (ABI). Já entre as plataformas consideradas de terceira geração figuram o PacBio RS *system* (Pacific Biosciences) e a Personal Genome Machine - Ion PGM™ (Life Technologies) (Kaur & Malik, 2013).

As plataformas *high throughput* podem ser divididas em dois grandes grupos (tendo como base o tipo de *template* necessário para as reações do processo de sequenciamento): as que dependem de bibliotecas de *templates* amplificados clonalmente (amplificação de uma biblioteca imobilizada feita de uma única molécula de DNA) e as plataformas mais novas, capazes de determinar a sequência de uma única molécula sem a necessidade de amplificação (Loman *et al.*, 2012), onde a primeira plataforma para sequenciamento de uma única molécula foi o HeliScope (Helicos BioSciences) (Bowers *et al.*, 2009).

Devido à complexidade do organismo, duas plataformas *high throughput* de *templates* amplificados clonalmente foram utilizadas para o sequenciamento da linhagem de *L. lactis* NCDO 2118: SOLiD 5500 XL e Ion PGM™.

### 1.3 A GENÔMICA DAS BACTÉRIAS LÁTICAS

O primeiro genoma completo do grupo das BL a ser sequenciado foi o da linhagem *Lactococcus lactis* subsp. *lactis* IL1403 (aqui, *L. lactis* IL1403), publicado em 2001. Nele, foram reveladas vias biossintéticas, profagos e parte de componentes que participam no metabolismo aeróbico (Bolotin *et al.*, 2001).

Em 2002, um programa destinado ao sequenciamento em massa dos genomas das BL foi anunciado pelo “*Lactic Acid Bacteria Genome Sequencing Consortium*” (Consórcio de Sequenciamento Genômico das Bactérias do Ácido Láctico) (Klaenhammer *et al.*, 2002).

Atualmente, mais de 30 genomas de *Lactococcus* estão disponíveis no banco de dados do NCBI (National Center for Biotechnology Information - <http://www.ncbi.nlm.nih.gov/genome/genomes/156>), onde 15 deles são completos.

As BL apresentam genomas relativamente pequenos, com tamanho e quantidade de genes de aproximadamente 2 megabases (mb) e 2000 genes, respectivamente, podendo abranger uma faixa de 1600 a 3000 genes em diferentes espécies. Esta variação resulta da evolução das BL via processos de perda, duplicação e aquisição gênica (Khalid, 2011).

Durante a evolução das BL, pôde-se observar tanto uma redução do seu genoma e eliminação de sistemas desnecessários durante sua adaptação a ambientes específicos, quanto um ganho de genes durante a sua adaptação a locais como o leite, por exemplo. Com relação à perda destes genes, o exemplo mais notável ocorreu em *Streptococcus thermophilus*, onde esta divergiu de outras espécies de *Streptococcus* através da perda de genes associados à virulência, tais como aqueles envolvidos na adesão e resistência a antibióticos (Bolotin *et al.*, 2004).

Vários estudos destacam a importância do sequenciamento genômico na descoberta de novas características relacionadas às BL. Por meio desta abordagem, Cogan e colaboradores conseguiram identificar vários genes codificantes de enzimas proteolíticas (participantes na maturação do queijo) de *Lactobacillus helveticus* (Cogan *et al.*, 2007). Assim como o uso da abordagem de bioinformática auxiliou na identificação da via catabólica do citrato em *Lactobacillus casei* (Díaz-Muñoz *et al.*, 2006), vários outros estudos têm identificado genes responsáveis pela descarboxilação da cadeia ramificada dos alfa-cetoácidos em *Lactococcus lactis* (Smit *et al.*, 2005; de la Plaza *et al.*, 2004). Outro exemplo interessante diz respeito

ao sequenciamento dos genomas de *Lactobacillus* com nicho no intestino que ilustram os mecanismos pelos quais estas espécies têm se adaptado ao GIT, como tolerância ao estresse, absorção de carboidratos, além da adesão a células epiteliais e muco (Siezen *et al.*, 2011).

Além das análises de sequências genômicas, informações adicionais sobre a genômica das BL têm sido encontradas por meio da identificação de plasmídeos. Estes foram identificados em vários gêneros de BL e muitas linhagens albergam múltiplos plasmídeos. Apesar de serem geralmente considerados dispensáveis para a célula hospedeira, alguns plasmídeos carregam genes que podem ser essenciais para a sobrevivência de bactérias sob determinadas condições. No gênero *Lactococcus*, os plasmídeos têm sido associados com a resistência a fagos, utilização de citrato, produção de bacteriocinas, metabolismo de açúcares, proteólise e transferência gênica (Wood & Warner, 2003).

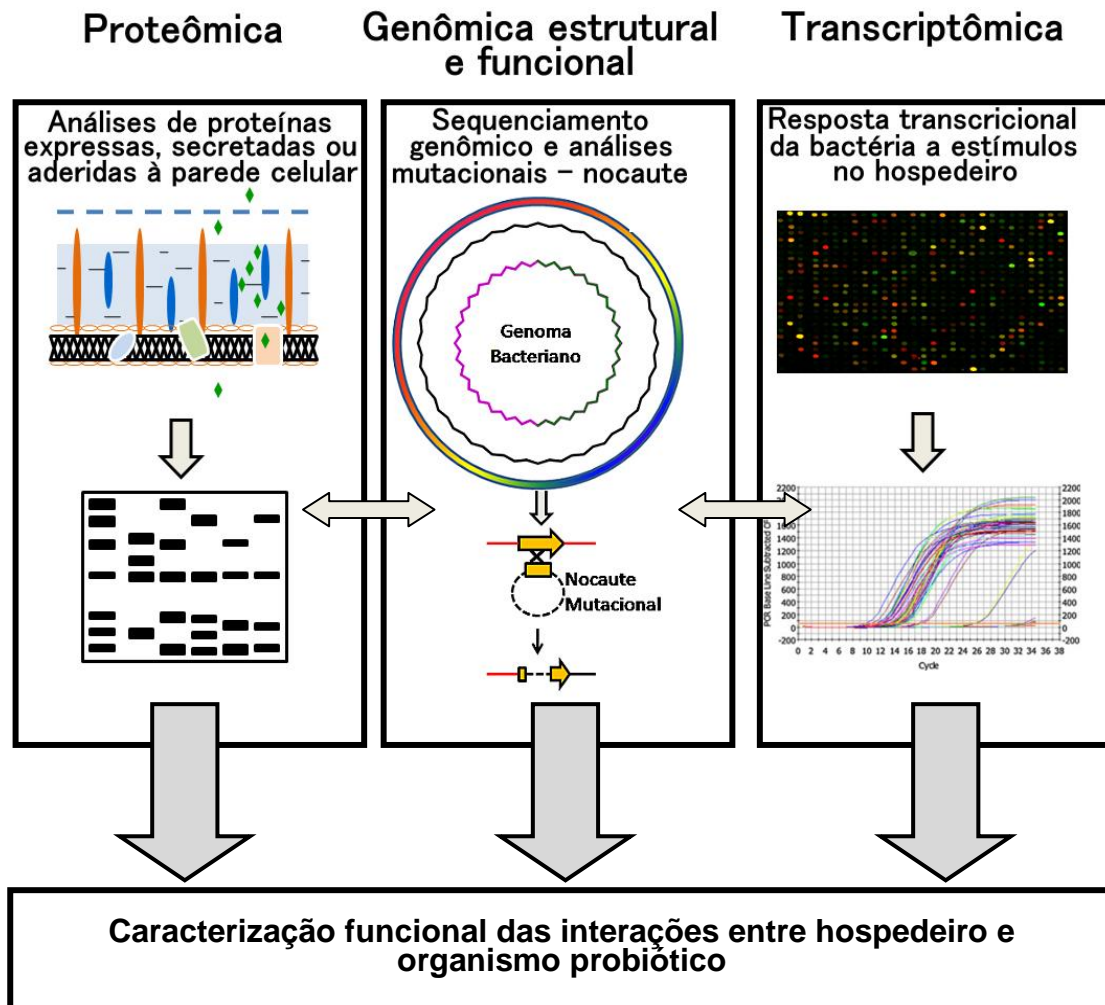
Finalmente, além da abordagem genômica, outras abordagens ômicas (**Fig. 2**) vêm proporcionando a análise da divergência e evolução das mais variadas espécies ao longo do tempo (Pfeiler & Klaenhammer, 2007). Por meio das abordagens ômicas é possível, por exemplo, correlacionar dados de proteínas importantes para a sobrevivência dentro do hospedeiro durante situações de estresse ou de proteínas secretadas que podem exercer um papel nos efeitos probióticos de dadas linhagens, através de análises de interação bactéria-hospedeiro.

#### 1.4 PLASTICIDADE GENÔMICA

No decorrer do tempo, é possível observar mudanças no conteúdo e na ordem da informação genética dos organismos. A pressão evolutiva atua no hospedeiro conduzindo as bactérias a passarem por modificações genômicas. Este processo de mudanças no DNA é chamado de plasticidade genômica (Soares *et al.*, 2011).

A plasticidade genômica é caracterizada por quatro formas de mudanças do DNA: conversão gênica e mutações pontuais, rearranjos (por meio de tranlocação e inversão, por exemplo), deleção e inserção de material gênico de outros organismos por meio de: plasmídeos, transposons, bacteriófagos dentre outros. Estes mecanismos são capazes de alterar o estilo de vida das bactérias, contribuindo para

sua adaptação em diversos ambientes específicos, além de influenciar na sua evolução (Schmidt & Hensel, 2004).



**Fig. 2** Integração das áreas ômicas. Através do sequenciamento genômico, aliado às técnicas de proteômica, genômica funcional e transcriptômica, mecanismos da interação entre organismos probióticos e hospedeiro vêm sendo caracterizados (Adaptado de Johnson & Klaenhammer, 2014).

Com relação à inserção de material genômico, o estudo de regiões de fago ganha maior atenção, principalmente em se tratando de algumas bactérias como as pertencentes ao grupo das BL, pelo fato de serem utilizadas em produtos fermentados para consumo humano. Fagos são parasitas obrigatórios e a maioria deles possui ciclo de multiplicação por meio de lise celular, onde centenas de partículas virais são liberadas, prontas para infectar células próximas (Kutter & Sulakvelidze, 2005).

Para entender a presença dos bacteriófagos, ou apenas fagos, no planeta, em apenas uma gota de água do oceano, é possível encontrar até  $10^8$  fagos (Wommack & Colwell, 2000).

Todo processo industrial ou biotecnológico que requer o uso de bactérias na produção de produtos alimentícios, ou de moléculas, poderá ser interrompido, em algum momento, devido à presença de fagos virulentos. Eles são a causa primária de falhas nos processos fermentativos durante a transformação industrial do leite (Garneau & Moineau, 2011).

A primeira descrição de fagos afetando culturas iniciadoras de laticínios data de 1935 e, desde então, melhoramentos importantes têm sido feitos, particularmente nas áreas de ecologia, genômica de fagos e resistência a fatores ambientais (Whitehead & Cox, 1935). A literatura também relata regiões de profagos em espécies do gênero *Lactococcus*, como: *Lactococcus lactis* (Brussow, 2001; Cavanagh *et al.*, 2014; Geller *et al.*, 2005).

Contudo, mesmo com esses avanços, ainda é possível observar contaminação por fagos, prejudicando a variabilidade dos produtos e reduzindo a sua produtividade (Moineau & Lévesque, 2005). Os fagos podem apresentar vários locais de origem, por isso, é de grande importância conhecer desde as fontes de fagos em potencial até a sua ação na produção de produtos lácteos, prejudicando os processos de fermentação (Garneau & Moineau, 2011).

Outro fator que caracteriza a adição de material genômico é a presença de ilhas genômicas (GEI). Elas podem ser classificadas como ilhas de patogenicidade (PAI), ilhas metabólicas (MI), ilhas simbióticas (SI) e ilhas de resistência (RI) e são identificadas por serem regiões de transferência e incorporação de um conjunto significativo de genes em blocos (genes que codificam funções similares e operons), com potencial de trazer grandes mudanças evolutivas (Soares *et al.*, 2011).

Estas regiões de ilha são caracterizadas primeiramente, por serem regiões adquiridas de outros organismos, além de ocuparem regiões genômicas grandes de 10 a 200 kilobases (kb). Outra característica diz respeito a sua localização, frequentemente são encontradas em regiões adjacentes a genes de tRNA (Schmidt & Hensel, 2004).

## 1.5 AS BACTÉRIAS LÁTICAS

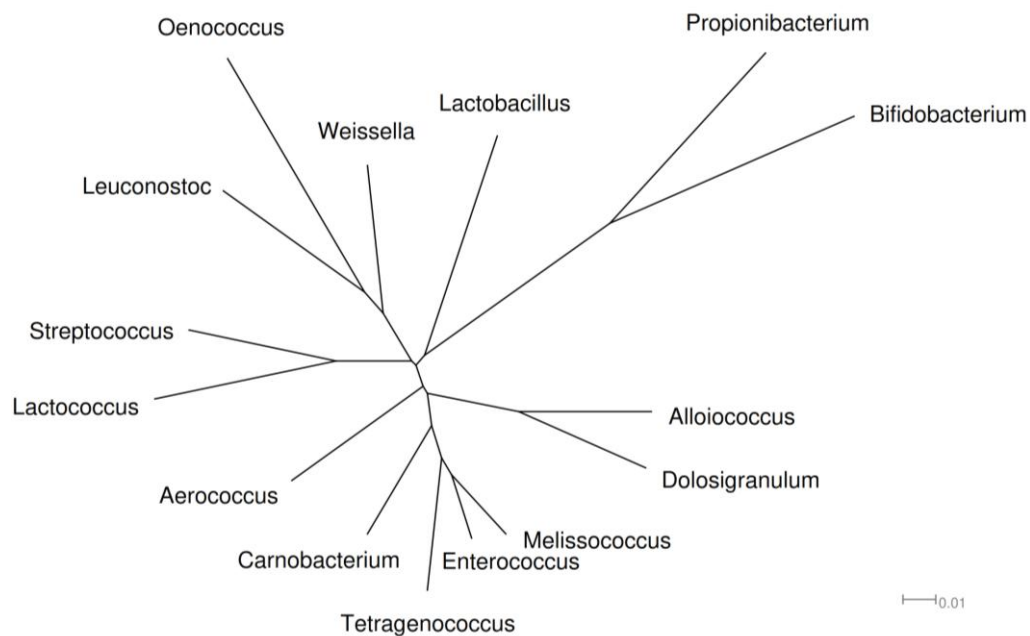
### 1.5.1 DEFINIÇÃO

As BL fazem parte de um grupo bastante heterogêneo de bactérias gram-positivas, não esporuladas, que alberga espécies dos gêneros: *Streptococcus*, *Lactococcus*, *Leuconostoc*, *Pediococcus* e *Lactobacillus*. Revisões quanto à taxonomia destas bactérias foram feitas propondo vários gêneros novos ao grupo das BL como: *Carnobacterium*, *Aerococcus*, *Dolosigranulum*, *Alloiococcus*, *Enterococcus*, *Globicatella*, *Oenococcus*, *Tetragenococcus*, *Weissella* e *Vagococcus* (Khalid, 2011).

O termo BL as define principalmente com relação à sua capacidade metabólica comum; salvo poucas exceções, elas são responsáveis pela conversão de açúcares (principalmente glicose) em ácido láctico através de vias metabólicas específicas. É possível observar outras características compartilhadas entre estas bactérias como: (i) serem anaeróbicas facultativas; (ii) não produtoras de catalase; e, (iii) serem imóveis. Além disso, existe uma proximidade filogenética entre as bactérias deste grupo que pode ser observada na **figura 3** (Pfeiler & Klaenhammer, 2007).

Devido à sua grande diversidade, essas bactérias podem ocupar vários nichos diferentes, como superfícies de plantas e animais e no GIT destes e, em algum período da vida, todo ser vivo entrará em contato com estas bactérias, seja através da alimentação ou do meio ambiente. Acredita-se, ainda, que as BL se encontram em estado latente na superfície de plantas e que podem se multiplicar no trato gastrointestinal de animais, após a ingestão dessas (Bolotin *et al.*, 2001).





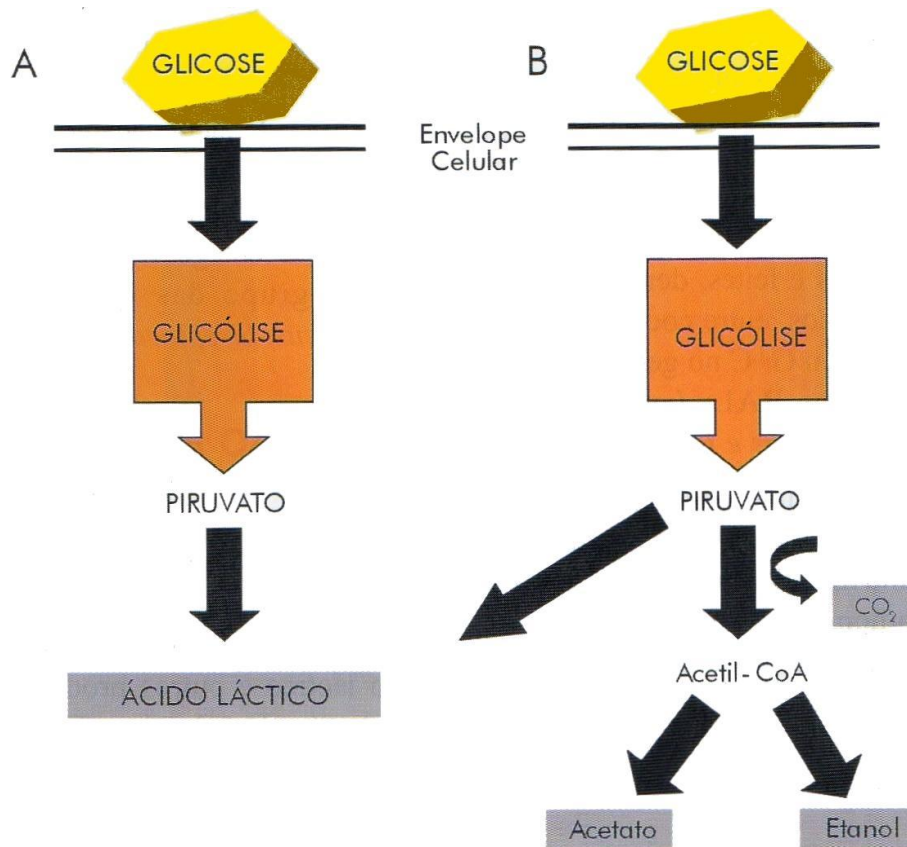
**Fig. 3** Filogenética das bactérias lácticas. Análise filogenética das principais bactérias lácticas, baseada na comparação de sequências de 16S. O alinhamento global das sequências nucleotídicas foi feito com ClustalW e a geração da árvore foi realizada por meio do *software* SplitsTree. (Este trabalho).

### 1.5.2 IMPORTÂNCIA INDUSTRIAL E BIOTECNOLÓGICA

As BL "domesticadas", ou seja, manipuladas nas indústrias de laticínios, vivem em um nicho diferente, definidos por razões tecnológicas, tais como o crescimento e produção rápida de ácido láctico no leite (Bolotin *et al.*, 2001).

A maioria das BL adquire energia através da transformação de açúcares, principalmente da glicose em ácido láctico (**Fig. 4**) (via homofermentativa, ou homolática) e/ou ácido láctico e outros produtos (via heterofermentativa ou mista) (Carr *et al.*, 2002). De um modo geral, estes organismos estão associados ao preparo de alimentos, na preservação de carnes, grãos e vegetais (Bolotin *et al.*, 2001).

Dessa forma, estudos relacionados às análises de sequência do genoma das BL despertam grande interesse da indústria alimentícia. Além disso, devido à sua utilização ao longo dos anos na manutenção e preservação de alimentos, as espécies do gênero *Lactococcus* são consideradas bactérias seguras ou "GRAS" (*Generally Recognized As Safe*), mesmo seu grupo apresentando algumas poucas espécies patogênicas. No grupo das bactérias lácticas, a *Lactococcus lactis* é uma das espécies melhor caracterizada e que figura como um microrganismo modelo para estudos deste grupo (Bolotin *et al.*, 2001).



**Fig. 4** Figura representativa das vias metabólicas de consumo da glicose pelas bactérias lácticas. (A) Via hemoláctica - A glicose é transformada em piruvato (Glicólise) e este, é convertido em ácido láctico. (B) Via mista – O processo de conversão da Glicose é semelhante ao da via hemoláctica até a formação do piruvato. A partir daí o piruvato é transformado em Acetil-CoA, e esta, convertida em acetato, etanol e CO<sub>2</sub> (Azevedo & Miyoshi, 2004).

### 1.5.3 NOVAS UTILIZAÇÕES DAS BACTÉRIAS LÁCTICAS

BL também possuem um grande potencial de utilização na produção de biomoléculas. Proteínas de interesse biotecnológico e médico, como antígenos, citocinas (Marinho *et al.*, 2010), enzimas (Du *et al.*, 2010) e alérgenos (**Tabela 1**), já foram produzidos por meio de engenharia genética pelas BL (Bermudez *et al.*, 2004; Nouaille *et al.*, 2003).

No grupo das BL, a espécie *Lactococcus lactis* tem sido intensivamente utilizada na produção de proteínas heterólogas (Mierau & Kleerebezem, 2005). Isto se deve ao fato de que poucas proteínas são conhecidas como secretadas por esta espécie e apenas uma, Usp45 (*Unknown Secreted Protein of 45 kDa*) é secretada em quantidades detectáveis por gel corado com *coomassie*; outro motivo refere-se ao fato de que as proteínas secretadas são propensas, na maior parte do tempo, à degradação extracelular (Lee *et al.*, 2000; Wu *et al.*, 1991; Wu *et al.*, 1998) e as

linhagens laboratoriais de *L. lactis* não produzem nenhuma protease extracelular (Nouaille *et al.*, 2003).

Se tratando da linhagem em estudo, *L. lactis* NCDO 2118 é uma robusta linhagem isolada de vegetal que possui a capacidade de crescer em meios menos complexos que a maioria das linhagens lácteas, permitindo seu uso em sistemas de produção de menor porte (Miyoshi *et al.*, 2004).

| PRODUTO   | MODELO                 | APLICAÇÃO   |
|-----------|------------------------|---|
| DNA       | camundongos            | Vacinas   |
| Peptídeo  | camundongos            | HIV-1, fusão de inibidores, colite                            |
| ScFvs     | camundongos            | Anti-infectante   |
| Antígenos | camundongos e galinhas | Vacinas   |
| Citocinas | camundongos e humanos  | Doença inflamatória intestinal e colite, adjuvantes de vacina |
| Enzimas   | camundongos            | Terapia para colite e deficiência de enzimas                  |
| Alérgenos | camundongos            | Terapia e prevenção de alergia                                |

**Tabela 1.** Atual aplicação das bactérias lácticas. Várias moléculas têm sido produzidas nas BL e muitos modelos já foram utilizados (Wells & Mercenier, 2008). ScFvs = *Single-chain variable fragment*.

Em 1999, Langella & Loir desenvolveram ferramentas para a secreção de proteínas heterólogas onde relataram o aumento de expressão do gene *nuc* (nuclease), uma proteína repórter, por meio da clonagem de várias cópias do vetor e de um promotor deste gene de *Lactococcus*. Com este trabalho, foi relatado a possibilidade de *L. lactis* ser uma forte candidata para desenvolvimento de vacinas vivas como veículos de entrega (Langella & Loir, 1999).

Já Miyoshi e colaboradores desenvolveram um sistema de expressão e endereçamento protéico utilizando *L. lactis*. Este sistema, nomeado XIES (*Xylose-Inducible Expression System*), utiliza-se do promotor *Pxylt* (gene da xilose permease) e dos elementos genéticos: sítio de ligação ao ribossomo (RBS); e sequência codificadora do peptídeo sinal (SP) da proteína Usp45 de *L. lactis*. Este sistema de expressão foi utilizado com sucesso na linhagem *L. lactis* NCDO 2118, onde demonstrou ser capaz de: (i) produzir e secretar, por um período mais longo, elevados níveis da proteína modelo Nuc (nuclease B) de *Staphylococcus aureus*, na presença de xilose; tais níveis foram comparados apenas com os obtidos utilizando-

se o sistema NICE (*Nisin-Controlled gene Expression*), produzidos na fase estacionária de crescimento; (ii) endereçar o produto final para o citoplasma ou meio extracelular e (iii) permitir ativar e desativar a expressão gênica por meio da adição de xilose ou glicose, respectivamente. Ademais, este sistema é mais vantajoso que os atuais sistemas de expressão por ser de mais fácil manipulação, menos dispendioso e principalmente, por ser mais seguro para o uso humano e animal.

A linhagem *L. lactis* NCDO 2118 foi descrita como produtora do ácido gama-aminobutírico - GABA (Mazzoli *et al.*, 2010), um neurotransmissor distribuído no sistema nervoso central dos vertebrados e capaz de diminuir a pressão sanguínea de pacientes levemente hipertensos (Inoue *et al.*, 2003), além de induzir um efeito tranquilizante e diurético (Jakobs *et al.*, 1993; Wong *et al.*, 2003), prevenir diabetes (Adeghate & Ponery, 2002), e reduzir a resposta inflamatória de artrite reumatóide em modelo murino (Tian *et al.*, 2011).

Devido a estas características e alta capacidade industrial de *Lactococcus lactis* como principais componentes na produção de produtos lácteos, é grande o interesse relacionado a esta espécie (Siezen *et al.*, 2011). Apesar de sua ampla utilização na indústria de alimentos e nas novas áreas que foram citadas anteriormente, as BL só ganharam evidências experimentais há pouco tempo (Johnson & Klaenhammer, 2014). Dentre os vários papéis benéficos das bactérias probióticas no GIT, podem-se destacar: a inibição no crescimento de *H. pylori* (Ushiyama *et al.*, 2003), redução dos biomarcadores de câncer do colo retal (Rafter *et al.*, 2007), diminuição dos níveis de colesterol no sangue (Ataie-Jafari *et al.*, 2009), diminuição dos riscos de diarreia associada a antibióticos (Gao *et al.*, 2010), dentre outros.

Se tratando da linhagem *L. lactis* NCDO 2118, a atenção estará voltada para o seu potencial probiótico relatado recentemente na literatura por Luerce e colaboradores (Luerce *et al.*, 2014), e o que já conhecido na literatura sobre outros organismos com tais características.

#### 1.5.4 EFEITO PROBIÓTICO DAS BACTÉRIAS LÁTICAS

A definição de bactéria probiótica está relacionada a culturas vivas que mantêm a saúde e o equilíbrio da microbiota intestinal (Cronin *et al.*, 2011). Os alimentos fermentados, produzidos a partir do gênero *Lactobacillus*, têm sido amplamente empregados devido às suas propriedades benéficas, como: obtenção

de alívio diante os sintomas de intolerância à lactose e diminuição da diarreia por rotavírus (Ouwehand *et al.*, 2002).

Diferentes gêneros de bactérias são utilizados como probióticas como, por exemplo: *Lactobacillus casei* Shirota, *Lactobacillus delbrueckii*, *Lactobacillus acidophilus* La5, *Lactobacillus plantarum* 299v, *Lactobacillus fermentum* KLD, *Lactobacillus reuteri* SD2112, *Bifidobacterium breve*, *Bifidobacterium longum* BB536, *Bifidobacterium lactis* Bb12, *Propionibacterium freudenreichii* JS, *Bacillus subtilis*, *Bacillus cereus* toyoi, *Escherichia coli* Nissle 1917 e *Enterococcus faecium* SF68. Todas estas espécies apresentam aplicações terapêuticas na prevenção e tratamento de distúrbios como diarreia em recém-nascidos (Ouwehand *et al.*, 2002).

O primeiro relatório científico amplamente referenciado pela comunidade científica a respeito das bactérias probióticas, data do ano de 1907. Nele, Elie Metchnikoff demonstrou a correlação existente entre a ingestão de bactérias produtoras do ácido láctico e o aumento da longevidade dos búlgaros e outras populações (Howell, 1988).

Recentemente, a maioria dos estudos e utilização de probióticos está relacionada às BL, evidenciando em especial os lactobacilos, isolados do GIT. Em estudos de Luerce e colaboradores, 2014, três linhagens de *L. lactis* subsp. *lactis* (MG1363, IL1403 e NCDO 2118) foram analisadas *in vitro*, quanto a capacidade de modulação do sistema imunológico. A cultura com *L. lactis* NCDO 2118 foi capaz de apresentar uma resposta anti-inflamatória significativa, o que não foi observado nas outras linhagens analisadas (Luerce *et al.*, 2014).

Uma curiosidade sobre a utilização destas bactérias para consumo humano é em relação a quantidade necessária para que o hospedeiro tenha uma resposta desejada em relação ao uso de um produto probiótico, que pode variar de acordo com a linhagem, tipo de emprego e a formulação, sendo geralmente recomendado pelo menos  $10^7$  microorganismos por mililitro (mL) diários (Corcoran *et al.*, 2008).

Os mecanismos e interações dos probióticos com o GIT do hospedeiro vêm sendo desvendados por meio da fusão de técnicas de genômica estrutural e funcional. Dentro deste contexto, três pontos principais merecem atenção quanto aos mecanismos da ação probiótica: (i) a sobrevivência através do trânsito no GIT; (ii) a atividade antimicrobiana e a exclusão por competitividade e (iii) a modulação do sistema imune do GIT do hospedeiro (Johnson & Klaenhammer, 2014).

#### 1.5.4.1 SOBREVIVÊNCIA ATRAVÉS DO TRATO GASTROINTESTINAL

Uma das principais qualidades de um microrganismo probiótico é sua habilidade de sobreviver ao variado ambiente que compõe o GIT. Ele deve ser capaz de se adaptar ao ácido gástrico e ao suco biliar no intestino delgado. Como em muitas bactérias que crescem em meio ácido, as membranas lipídicas de *Lactobacillus*, por exemplo, expostas ao ácido e bile, são alteradas com o objetivo de aumentar a sobrevivência. A membrana lipídica de *Lactobacillus casei* demonstrou um aumento acentuado de ácidos graxos monoinsaturados em resposta a acidificação (Fozo *et al.*, 2004).

Espécies de *Lactobacillus* possuem respostas transcricionais aos fatores de estresse, geralmente por meio de sistemas regulatórios de dois componentes (2CRS – *two-component regulatory systems*) (Lebeer *et al.*, 2008). A bactéria rapidamente percebe a mudança das condições do ambiente via 2CRS, através da detecção via domínios de transmembrana das proteínas histidina quinase (HPK). Após receber o sinal do ambiente, a HPK é ativada a autofosforilar um resíduo de histidina específica, o qual é transferido para um domínio regulador de resposta (RR), um regulador transcricional de ligação ao DNA. Portanto, é possível realizar a predição dos 2CRS da sequência genômica bacteriana, baseado na presença de HPK e RR putativos, onde ambos estão próximos (Altermann *et al.*, 2005 ; Morita *et al.*, 2011).

Além da sobrevivência pelo trânsito gastrointestinal, um fator que pode contribuir para a atividade probiótica, é a adesão ao epitélio intestinal. Estudos preliminares *in vitro* utilizando células epiteliais intestinais Caco-2 revelaram múltiplos *Lactobacillus* probióticos com capacidade de produzir adesinas (Chauvière *et al.*, 1992; Tuomola & Salminen, 1998). A adesividade de *Bifidobacterium spp.* ao muco intestinal humano também foi demonstrada no trabalho de He e colaboradores, em 2001 (He *et al.*, 2001). O acesso aos dados de sequência genômica, combinado às técnicas genômicas, ajudaram a elucidar mediadores de adesão. A maioria destes são secretados ou ligados à parede celular em uma forma de sortase-dependente, no intuito de interagir com o epitélio intestinal (Lebeer *et al.*, 2008; Vélez *et al.*, 2007).

A adesão microbiana ao tecido do hospedeiro é um evento crítico na patogênese de muitas infecções e, como tal, é um alvo no desenvolvimento de novas terapias antimicrobianas. Adesinas de superfície celular que interagem com componentes da matriz extracelular e que são capazes de reconhecer fibronectina,

fibrinogênio, colágeno e polissacarídeos relacionados à heparina, são designadas MSCRAMMs (componentes microbianos de superfície que reconhecem moléculas adesivas da matriz - do inglês: *microbial surface components recognizing adhesive matrix molecules*) (Patti *et al.*, 1994).

As proteínas de ligação à fibronectina A e B (FnbpA e FnbpB) são componentes importantes na etapa inicial da adesão de *S. aureus* à superfícies do hospedeiro (Bingham *et al.*, 2008; O'Neill *et al.*, 2008). Estas moléculas fazem parte da família dos MSCRAMMs e formam ligações com os ligantes dos hospedeiros como a fibronectina (Fn), responsáveis pelo revestimento de dispositivos prostéticos *in vivo* (Patti *et al.*, 1994).

Em *L. acidophilus* NCFM, uma busca *in silico* no genoma permitiu a identificação de cinco proteínas preditas como sendo de adesão de superfície celular, incluindo uma proteína de ligação à fibronectina (FbpA), uma proteína S-layer (SlpA), uma proteína de ligação de mucina (Mub) e duas proteínas R28 homólogas envolvidas na adesão de *Streptococcus* (Buck *et al.*, 2005). Por meio de análises mutacionais, FbpA, SlpA e Mub foram identificadas por contribuir na adesão às células epiteliais Caco-2. Similarmente, uma proteína de resposta ao estresse e um fator promotor de agregação (ambos proteínas de superfície), foram encontrados mais tarde em estudos contribuindo na aderência à células Caco-2 (Goh & Klaenhammer, 2010; O'Flaherty & Klaenhammer, 2010).

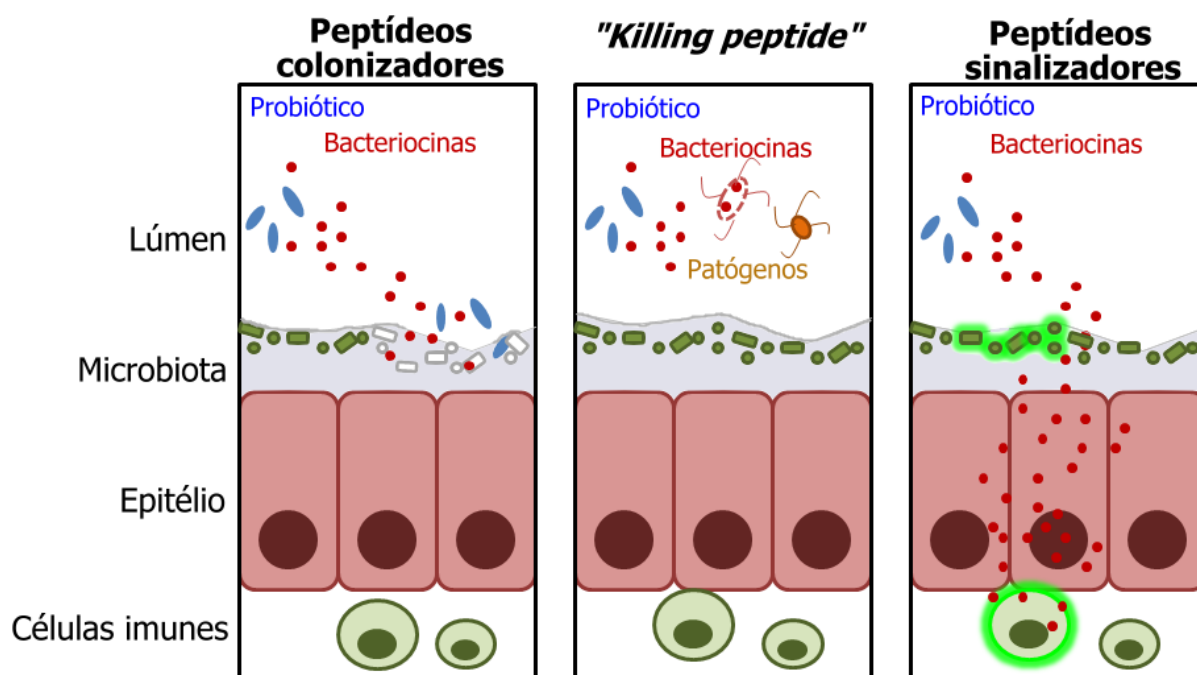
Uma mesma abordagem utilizada para duas linhagens de *L. rhamnosus*, revelaram a presença de uma ilha genômica em *L. rhamnosus* GG que possuía três proteínas secretadas, pili sortase-dependentes, codificadas pelo operon *spaCBA*, que foram confirmadas posteriormente como genes codificantes de pili através de análises experimentais (Kankainen *et al.*, 2009).

#### 1.5.4.2 ATIVIDADE ANTIMICROBIANA E EXCLUSÃO COMPETITIVA

A ação das BL na conservação dos alimentos deve-se tanto à acidificação do meio (pH 4.5 a 3.5), quanto à produção de inúmeros agentes bacterianos, como compostos orgânicos e bacteriocinas (van de Guchte *et al.*, 2001).

As bacteriocinas são peptídeos produzidos por bactérias que agem contra outros microrganismos, cujo produtor apresenta mecanismos de imunidade específica (Cotter *et al.*, 2005). Elas são capazes de inibir o crescimento de organismos indesejáveis (**Fig. 5**) e/ou influenciarem no desenvolvimento de

propriedades organolépticas, como aroma, textura e sabor do produto final (Pfeiler & Klaenhammer, 2007).



**Fig. 5** Mecanismos pelos quais as bacteriocinas podem contribuir para a funcionalidade probiótica. As bacteriocinas podem colonizar peptídeos, facilitando a competição de um organismo probiótico presente na microbiota (Gillor *et al.*, 2009); elas podem funcionar como peptídeos "assassinos", eliminando diretamente um patógeno (Corr *et al.*, 2007), ou também, elas podem sinalizar peptídeos, sinalizando outras bactérias ou o sistema imune (Kleerebezem *et al.*, 1997). Adaptado de (Dobson *et al.*, 2012).

Os primeiros trabalhos relatando o mecanismo de ação por meio da inibição mediada por bacteriocinas foi descrito há 89 anos com a descoberta de antagonistas entre linhagens de *Escherichia coli* (Gratia, 1925). Apesar do uso de bacteriocinas ter sido formalmente proposto em 1951 (Hirsch *et al.*, 1951), é provável que a humanidade já se beneficiava casualmente da produção de bacteriocinas nos 8.000 anos desde a produção dos primeiros produtos como o queijo e alimentos fermentados (Cotter *et al.*, 2005).

Um exemplo da ação das bacteriocinas pode ser observado no estudo de Corr e colaboradores, em 2007. Nele, identificaram que a bacteriocina Abp118, produzida por *Lactobacillus salivarius* UCC118, é bastante eficiente na redução de infecções por *Listeria monocytogenes* em camundongos, sugerindo que estas bacteriocinas possuem um papel como agentes anti-infecciosos (Corr *et al.*, 2007).

Em 1995, Kuipers e colaboradores descreveram a autoindução da expressão de bacteriocinas nisinas em *Lactococcus*, além de experimentos realizados com este mesmo mecanismo, para direcionar a indução da expressão de proteínas



heterólogas: a  $\beta$ -glucuronidase de *Escherichia coli*, através de nisina. Este estudo marcou o surgimento de um dos sistemas mais amplamente utilizados para genes de expressão em bactérias gram-positivas, nomeado mais tarde como NICE (Kuipers *et al.*, 1995).

As bacteriocinas foram classificadas pela primeira vez por Klaenhammer em 1993 (Klaenhammer, 1993). Desde então, algumas sugestões quanto à classificação destas vêm sendo estudadas (Cotter *et al.*, 2005). Segundo Karpinski & Szkaradkiweicz, as bacteriocinas se dividem nas classes I, II, III e IV.

Na classe I de bacteriocinas produzidas por bactérias gram-positivas, estão incluídas as lantibióticas ou peptídeos termostáveis de tamanho molecular abaixo de 5 kDa. Elas apresentam aminoácidos atípicos como lantionina (Lan), metilantionina (MeLan), dentre outras (Karpiński & Szkaradkiewicz, 2013). A bacteriocina melhor reconhecida desta classe é a nisina, produzida por determinadas linhagens de *Lactococcus lactis*. Ela possui uma ampla ação antibacteriana contra outras bactérias gram-positivas como *Leuconostoc*, *Lactococcus*, *Enterococcus*, *Lactobacillus*, dentre outras, além disso, ela previne a formação de esporos (Delves-Broughton *et al.*, 1996).

Já as bacteriocinas da classe II são representadas pelas bacteriocinas não lantibióticas, ou seja, que não possuem lantionina. Estas também são termoestáveis, porém com peso molecular de 10 kDa, pouco maior que as da classe I. Devido às diferenças de estrutura, existe uma divisão em subclasses entre as bacteriocinas deste grupo (Karpiński & Szkaradkiewicz, 2013), como por exemplo, a pediocina (IIa), lactacina F (IIb), enterocina (IIc) e lactococcina A (IId) (Cotter *et al.*, 2005).

As bacteriocinas com peso acima de 30 kDa estão classificadas na classe III. Estas são termolábeis e são produzidas, sobretudo, por bactérias gram positivas (van Belkum & Stiles, 2000). Nesta classe também existe uma subdivisão, onde um grupo é representado pelas enzimas bacteriolíticas (bacteriolisinas), as quais facilitam a morte de linhagens sensíveis por lise celular e o grupo não-lítico de proteínas antimicrobianas, representado pelas lisostafina e enterolisina A, respectivamente (Cotter *et al.*, 2005; Karpiński & Szkaradkiewicz, 2013).

Por fim, a classe IV inclui as bacteriocinas que, para possuírem uma atividade completa, requerem a presença de uma porção de carboidrato ou lipídeo na sua molécula (Jack *et al.*, 1995).

#### 1.5.4.3 CAPACIDADE IMUNOMODULATÓRIA

Algumas BL também são importantes por apresentarem atividades imunoestimulatórias e ação anti-inflamatória, atuando na prevenção e tratamento das IBD (Doença Inflamatória Intestinal) e doenças alérgicas. Por conta destas características, vários produtos lácteos estão disponíveis no mercado para consumo, sendo que os mais difundidos são os leites fermentados (de Moreno de Leblanc *et al.*, 2011).

Lebeer e colaboradores propuseram a presença de três mecanismos de ação de *Lactobacillus* probióticos, que seriam capazes de contribuir para a saúde humana. Um dos mecanismos ocorreria através da inibição de patógenos e restauração da homeostase microbiana através das interações existentes entre microrganismos. Outra maneira de ação consistiria no aumento da função da barreira epitelial e o terceiro mecanismo seria a modulação da resposta imune (Lebeer *et al.*, 2008).

Em estudos realizados recentemente, Luerce *et al.* (2014) demonstraram os efeitos anti-inflamatórios e imunomodulatórios de *L. lactis* NCDO 2118, onde a linhagem pode agir por meio de dois importantes mecanismos. Um deles seria na modulação da produção de citocinas do tecido intestinal, onde foi encontrado o aumento precoce de IL-6 (interleucina 6), responsável pela ativação de linfócitos e produção aumentada de anticorpos. Outras análises demonstraram a manutenção dos níveis de IL-10 que atua, por exemplo, na inibição de citocinas pró-inflamatórias e também pode interferir no aumento do complexo principal de histocompatibilidade (MHC - *major histocompatibility complex*) classe II. Ademais, foi observado o aumento expressivo de TNF- $\alpha$  (fator de necrose tumoral  $\alpha$ ), que induz uma resposta inflamatória local ajudando na contenção de infecções (Luerce *et al.*, 2014).

O segundo mecanismo está relacionado à ação de *L. lactis* NCDO 2118 na diminuição da produção de IL-8, que atua na mobilização, ativação e degranulação de neutrófilos (Janeway *et al.*, 2007).

Diante das diversas funções fisiológicas do epitélio intestinal, além da absorção de nutrientes, ele também é capaz de criar uma barreira física entre os ambientes interno e externo, sendo responsável pela fiscalização imunológica de comensais e possíveis bactérias patogênicas (Rocha, 2011).

Uma estratégia importante presente nos mamíferos para manutenção da homeostase no ambiente intestinal é minimizar o contato entre microrganismos do

lúmen e a superfície das células epiteliais intestinais (IEC). Diferentes tipos de receptores de reconhecimento de padrão (PRR) são expressos pelas IEC como: receptores *Toll-like* (TLR), receptores *NOD-like* (*Nucleotide oligomerization domain*) e receptores ligados a proteína G (GPCR), que reconhecem componentes microbianos (MAMPS, padrões moleculares associados a micróbios) e respostas de células moduladoras (Hill & Artis, 2010).

#### 1.5.5 PRODUÇÃO DE ÁCIDO GAMA-AMINOBUTÍRICO

Muito além da capacidade probiótica e imunomodulatória, algumas bactérias que potencialmente possuem tais características podem, também, produzir uma molécula biogênica de grande interesse nutricional e farmacêutico, o ácido  $\gamma$ -aminobutírico (GABA). A sua produção pelas bactérias probióticas é possível em razão da descarboxilação do glutamato. O GABA é um neurotransmissor amplamente distribuído no sistema nervoso central em vertebrados e vários estudos têm relatado o seu efeito positivo na saúde humana (Mazzoli *et al.*, 2010). Ele modula atividades de neurotransmissores como serotonina, noradrenalina e dopamina, responsáveis pela modulação do ritmo circadiano e humor (Machado-Vieira *et al.*, 2005).

Esta molécula tem um papel trófico e age como regulador no pâncreas, podendo prevenir condições relacionadas ao diabetes (Adeghate & Ponery, 2002), além de baixar a pressão sanguínea de pacientes com hipertensão leve (Inoue *et al.*, 2003).

No estudo de Inoue *et al.*, realizado em 2003, pacientes com hipertensão leve foram submetidos a um período de testes com leite fermentado contendo GABA, por meio do uso de *Lactobacillus casei* linhagem Shirota e *Lactococcus lactis* YIT 2027. Foi observado, no grupo experimental que consumiu o leite, a diminuição dos valores de pressão sanguínea, mostrando ainda a ausência de efeitos colaterais mediante ingestão do leite fermentado (Inoue *et al.*, 2003). Além disso, outros estudos da literatura relatam casos onde foram encontrados baixos níveis de GABA no líquido, cérebro e plasma de pacientes bipolares (Machado-Vieira *et al.*, 2005).

Diante das várias possibilidades de aplicações e estudos das BL, a linhagem *L. lactis* subsp. *lactis* NCDO 2118, produtora de GABA (Mazzoli *et al.*, 2010) e detentora de efeito anti-inflamatório (Luerce *et al.*, 2014) tem despertado o interesse deste estudo devido ao seu potencial probiótico promissor.

---

## 2. JUSTIFICATIVA

---

Por ser uma bactéria de grande interesse biotecnológico, tanto na área da saúde, quanto na produção de alimentos de qualidade (incluindo os processos de fermentação), *Lactococcus lactis* vem sendo estudada com o intuito de esclarecer características ainda desconhecidas, e assim, ampliar sua aplicabilidade nas mais diversas áreas.

Estudos da literatura vêm destacando o uso de *L. lactis* NCDO 2118 na produção de proteínas heterólogas; na capacidade anti-inflamatória e probiótica através de análises *in vitro* do seu sobrenadante. Além disso, atualmente está disponível apenas um genoma de *Lactococcus lactis* isolada de planta, sendo *L. lactis* NCDO 2118, o segundo genoma publicado na literatura.

Buscando desvendar os fatores que fazem de *L. lactis* NCDO 2118 tão promissora, estudos ômicos como: Genômica, Transcriptômica e Proteômica, se unem, auxiliando fortemente nas pesquisas e estudos, como na genômica comparativa.

Através das análises de bioinformática, busca-se estudar os genes envolvidos com a capacidade probiótica e imunomoduladora de *L. lactis* NCDO 2118, identificando os possíveis mecanismos envolvidos em sua sobrevivência através do trato gastrointestinal, a sua forma de adesão às células epiteliais intestinais e também, prever possíveis fatores relacionados à atividade microbiana e sua capacidade de exclusão por competitividade, além dos possíveis mecanismos específicos envolvidos na modulação do sistema imune do trato gastrointestinal do hospedeiro.

---

### 3. OBJETIVOS

---

Caracterizar o genoma de *L. lactis* NCDO 2118 e procurar pistas que expliquem a capacidade probiótica desta linhagem através de genômica comparativa.

---

#### 3.1 OBJETIVOS ESPECÍFICOS

---

- Montar, anotar e curar manualmente o genoma de *L. lactis* NCDO 2118;
- Caracterizar e anotar o genoma quanto: ao conteúdo G+C, número de regiões codificantes de proteína, à quantidade de RNAs ribossomais e transportadores;
- Analisar a similaridade e sintenia entre os diferentes genomas completos já depositados no banco de dados do NCBI por meio de técnicas filogenéticas;
- Identificar possíveis ilhas genômicas das espécies de *Lactococcus*;
- Identificar possíveis regiões adquiridas de fagos na linhagem de interesse;
- Analisar a capacidade imunomodulatória de *L. lactis* NCDO 2118 através de análises comparativas;
- Identificar possíveis bacteriocinas presentes no genoma de interesse.

---

## 4. MATERIAIS E MÉTODOS

---

### 4.1 OBTENÇÃO DE DNA DA LINHAGEM *Lactococcus lactis* subsp. *lactis* NCDO 2118

Esta linhagem foi crescida em meio M17 (Difco) a 30°C, contendo 0,5% de glicose (GM17), sem agitação durante 18h (Luerce *et al.*, 2014). No tratamento da linhagem, para romper a parede das bactérias, foram centrifugados 1,5 mL de cultura e adicionados 500µL de Tris EDTA (TE) 1x ao pellet. As bactérias foram inativadas a 80°C em banho-maria durante 10 minutos. Em seguida, foram acrescentados 50µL de dodecil sulfato de sódio (SDS) a 10% e 3µL de Lisozima. Ao final do tratamento da linhagem, as amostras foram incubadas a 65°C durante 30 minutos; nos 10 minutos finais foram acrescentados 10µL de RNase.

Para a purificação do DNA, foram adicionados 100µL de NaCl 5M e 80µL de Brometo de Cetiltrimetil Amônio (CTAB)/NaCl. Em seguida, as amostras foram incubadas a 65°C durante 10 minutos. Nestas, foram adicionadas 700µL de clorofórmio seguido de homogeneização por inversão (de 15 a 20 vezes) e centrifugação durante 5 minutos. A fase superior, um volume de 550µL, foi transferida para outro tubo. Seguindo o protocolo, foram acrescentados 0,6 v de isopropanol à amostra e esta foi incubada a -20°C durante 30 minutos. A amostra foi centrifugada durante 10 minutos e o sobrenadante, desprezado. O precipitado foi lavado em 500µL de etanol a 70% e então, invertido por 3 vezes; em seguida, foi realizada a centrifugação durante 5 minutos. O sobrenadante foi desprezado e o precipitado, secado à temperatura ambiente. O DNA foi ressuspenso em 50µL de H<sub>2</sub>O e adicionado 10 mg/mL de RNase. Para que o precipitado tivesse uma ressuspensão ótima, este foi colocado *overnight* em baixa temperatura.

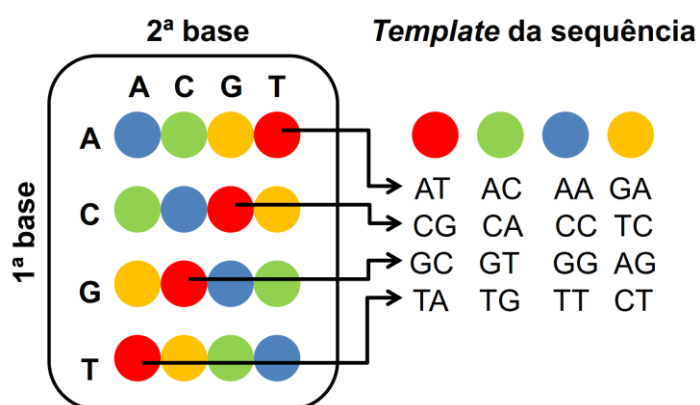
### 4.2 QUÍMICA DO SEQUENCIAMENTO

Como mencionado anteriormente, as plataformas SOLiD 5500 XL e Ion PGM™ foram utilizadas para o sequenciamento da linhagem de *L. lactis* NCDO 2118.

#### **SOLiD 5500 XL**

A plataforma SOLiD 5500 XL emprega o sequenciamento através da detecção de ligações de oligonucleotídeos de tamanhos fixos. Neste método, é realizada uma amplificação do DNA por PCR (*polymerase chain reaction*) em

emulsão, onde pequenas *beads* são utilizadas para amplificar os fragmentos, e posteriormente, a reação de sequenciamento ocorre por meio da utilização de DNA ligase. Cada ligação é seguida por uma detecção de uma fluorescência específica onde cada sinal representa dois nucleotídeos (**Fig. 6**). Este processo de codificação por duas bases permite a distinção entre erros de sequenciamento e polimorfismos nas sequências pela mudança da fluorescência, respectivamente: do ponto onde o erro ocorreu até o final da leitura; ou, em duas ligações consecutivas, mantendo o padrão das demais ligações até o fim da leitura (Henson *et al.*, 2012; Kaur & Malik, 2013; Mardis, 2008).



**Fig. 6** Possíveis dinucleotídeos codificados por cada cor (Adaptado de Mardis, 2008).

### Ion Torrent Personal Genome Machine (PGM)

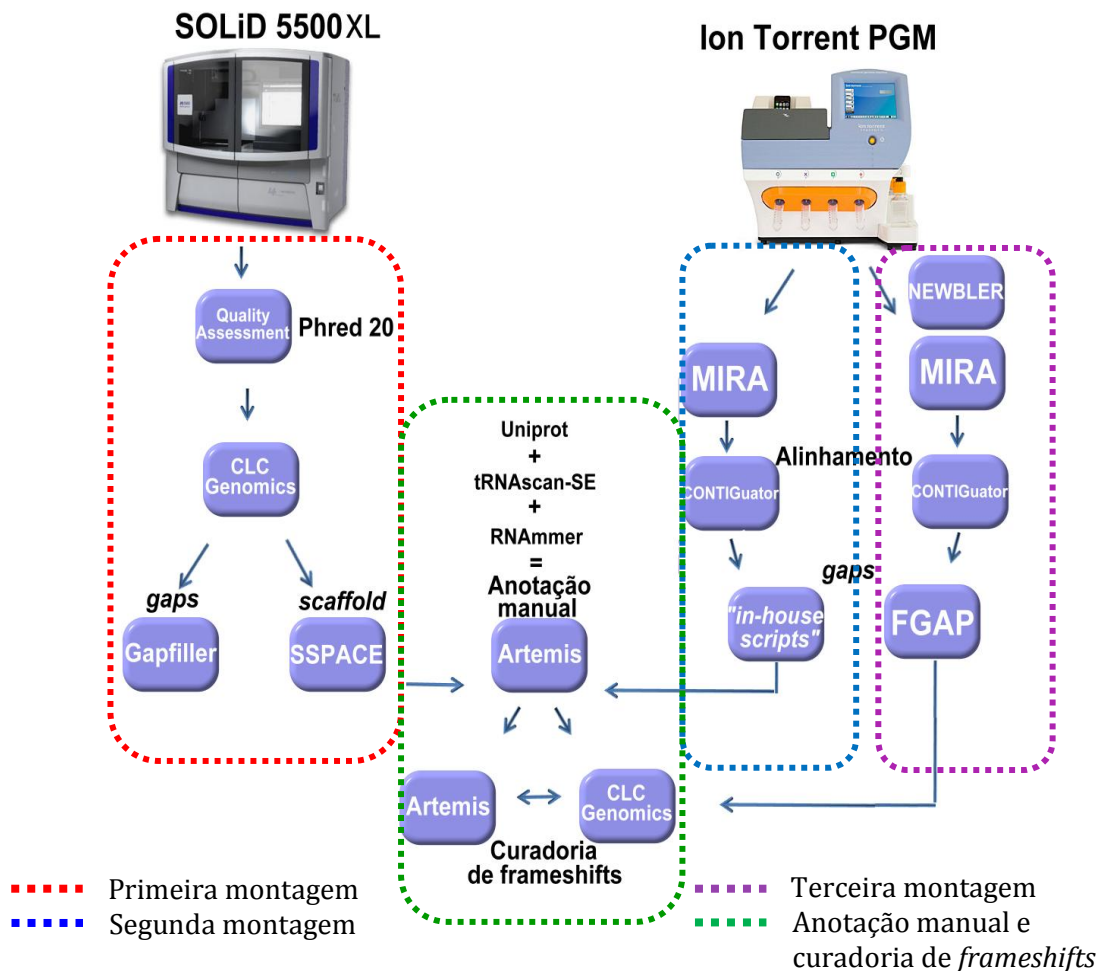
Representando as plataformas de terceira geração, o Ion PGM™ utiliza um padrão de sequenciamento por síntese e detecção por intermédio de um sistema de semicondução (Rothberg *et al.*, 2011). Através deste, o sequenciamento é realizado por meio de detecção de íons de hidrogênio, que são liberados durante a polimerização do DNA. Por meio da detecção direta dos dNTP (desoxiribonucleotídeos), o pH do meio é alterado e este é detectado por um transistor, convertendo a mudança de pH em um sinal elétrico (Varuzza, 2013). Esta tecnologia é capaz de gerar *reads* de tamanho aproximado de 400 pb (pares de bases) com maior acurácia, sendo que seu último chip: PGM 318, pode produzir 2.0 Gb (gigabase) de dados em apenas 2h de corrida ([www.iontorrent.com](http://www.iontorrent.com)).

Um desafio com relação a esta plataforma está relacionado às taxas de erros de inserção/deleção que são gerados devido aos picos heterogêneos provenientes das regiões de homopolímeros. Na publicação de Jünemann e colaboradores (2013), o desempenho de alguns sequenciadores foi comparado e observaram que

96% dos erros por *indel* (inserção ou deleção) no Ion PGM™ foram resultantes de regiões homopoliméricas (Jünemann *et al.*, 2013 - Supplementary Information).

### 4.3 SEQUENCIAMENTO

O sequenciamento de *L. lactis* NCDO 2118 aconteceu em três diferentes momentos, devido a sua complexidade. No sequenciamento em SOLiD 5500XL, não foi possível fechar o genoma, sendo este depositado como genoma *draft*, ou seja, incompleto. Um barcode foi realizado no primeiro sequenciamento em Ion PGM™ e assim, foi possível fechar o genoma, porém, devido a grande quantidade de pseudogenes, um novo sequenciamento na mesma plataforma foi realizado. Neste segundo sequenciamento, um chip completo (318) foi utilizado e com os dados do mapeamento, foi possível resolver grande parte dos pseudogenes restantes. As etapas dos sequenciamentos à obtenção do genoma curado, podem ser observadas na **Figura 7** abaixo:



**Fig. 7** Pipeline dos processos de sequenciamento, montagem e anotação de *L. lactis* NCDO 2118 (Este trabalho).



## Sequenciamento em SOLiD 5500XL

De acordo com protocolo disponibilizado para o sequenciador SOLiD 5500XL, foi construída a biblioteca *mate-paired*. Ela é constituída de *reads* de 60 pb cada; estas se formaram das extremidades de um mesmo fragmento, que estava separado por um adaptador interno.

Para a construção desta biblioteca, foram necessários 5 µg de DNA genômico, com intuito de assegurar a amplificação final do material e minimizar possíveis falhas de cobertura. Além disso, foi realizada uma avaliação da integridade da amostra e quantificação da mesma através de visualização em gel e dosagem, respectivamente.

A distância dos fragmentos foi determinada com base na complexidade e tamanho dos fragmentos. Primeiramente, foram testados todos os parâmetros de fragmentação deste DNA para, então, ser iniciada a preparação da biblioteca. O DNA foi então fragmentado através do sistema *HydroShear (GeneMachines)*, controlado pelo *software HydroShear v 2.0.9*.

Em seguida, a amostra foi concentrada em *speed vac* que reduziu seu volume com o objetivo de tornar mais precisa a seleção do tamanho dos fragmentos em gel de agarose 1%. Após esta etapa, uma corrida em eletroforese foi feita, paralelamente a um marcador específico de tamanho; as bandas foram separadas e o gel foi, então, colocado sob luz UV. Com auxílio de um bisturi, o gel foi recortado nas regiões dos insertos.

O DNA recuperado na etapa anterior passou por um processo de correção de extremidades através das enzimas *end polyshing E1* e *end polysing E2*. Logo depois, foi feita a ligação dos adaptadores MPR e MPL que formaram um adaptador interno após circularização, este processo aconteceu sob ação de alta temperatura seguido de resfriamento. O DNA circularizado foi, então, isolado através da enzima *plasmid-safe*, responsável pela degradação do DNA que não se circularizou. O DNA circular foi clivado pela enzima DNA polimerase I, onde foram originados os 60 pb em cada extremidade da biblioteca. A amostra restante foi eliminada pelas enzimas T7 exonuclease e S1 nuclease.

Com o intuito de aumentar a eficiência da ligação dos adaptadores P1-T e P2-T, uma cauda poli-A foi adicionada às extremidades de cada fragmento, que foi catalizada pela enzima *A Tailing Enzyme II*. A biblioteca foi, então, capturada pela *Dynabeads® Myone™ Streptavidin C1*, que se liga de forma específica à biotina do

adaptador interno da biblioteca. Uma reação de ligação, catalizada pela T4 DNA ligase, aderiu os adaptadores P1-T e P2-T às extremidades dos fragmentos, fazendo com que a biblioteca ficasse pronta para amplificação. Esta última, foi feita com os *primers* PCR *primer 1* e PCR *primer 2*, complementares aos adaptadores P1 e P2, respectivamente.

Neste momento, ficaram prontas as amostras para PCR (*polymerase chain reaction*) em emulsão, onde os fragmentos de DNA formados foram inseridos às esferas magnéticas. A técnica de PCR acontece em um sistema SOLiD *EZbead Emulsifier* (Life Technologies), onde microgotículas de fase aquosa foram geradas no interior da emulsão (microrreatores). A reação foi depositada em recipiente apropriado para ciclagem em SOLiD *EZbead Amplifier* (Life Technologies). Ao fim da emulsão, foi feito um processo de enriquecimento pelo sistema SOLiD *EZbead Enricher* (Life Technologies), onde somente as microesferas que apresentavam os fragmentos inseridos foram capturadas. Em seguida, estas microesferas passaram por uma etapa de modificação da extremidade 3' através da terminal transferase, para que então fosse possível a adesão à lâmina de sequenciamento. Terminado este processo, foi iniciado o sequenciamento conforme normas do fabricante, onde foram produzidas leituras de 60 nucleotídeos em cada *tag*.

### **Sequenciamento em Ion Torrent PGM**

A fragmentação enzimática do DNA genômico foi realizada com o kit Ion Shear™ Plus Reagents. A interação entre o DNA e a enzima varia de acordo com o tempo, podendo gerar fragmentos de diferentes tamanhos e com pontas cegas (blunt-ended). Para fragmentos com tamanho de 200 pb, foi estipulado um tempo de 5 minutos a 37°C.

O DNA, após ter sido fragmentado, foi então ligado a adaptadores Ion-compatíveis, Ion Xpress™ Barcode Adapters, seguido de *nick-repair* para fechar a ligação entre o inserto de DNA e os adaptadores. Para a corrida de um único DNA por chip, usou-se um adaptador padrão. Nesta etapa de *Size-selection*, foi realizada a escolha da biblioteca, ligada ao adaptador, de acordo com o tamanho do fragmento desejado.

O Ion Library Equalizer™ kit disponibiliza um método de normalização das bibliotecas em ~100 pM sem que seja necessário quantificá-las. Para isso, foi preciso realizar a amplificação da biblioteca após a etapa do *Size-selection* usando

Platinum® PCR SuperMix High Fidelity e a captura foi feita pelas Equalizer™ Beads. Após a eluição, a biblioteca equalizada pôde ser sequenciada. Para o caso de mais de uma amostra, diferentes bibliotecas podem ser combinadas após esta etapa.

Após a fase supracitada de equalização, as amostras seguiram para PCR em emulsão. Nesta etapa, os fragmentos de DNA foram ligados a partículas esféricas (Ion Sphere™ particles), através do sistema Ion One Touch™ e assim, foram clonalmente amplificados.

No enriquecimento, as partículas que possuíam DNA ligados a elas (partículas positivas), foram separadas por meio de *beads* magnéticas no Ion OneTouch™ ES. A partir disso, estas partículas puderam ser carregadas no chip semicondutor do Ion.

Nesta fase de sequenciamento, foram adicionados às partículas: Control Ion Sphere™ Particles, Sequencing Primer e Ion PGM™ Sequencing Polymerase. Esta mistura foi então carregada em uma parte do chip 318 e acomodada no Ion PGM™ pré-programado.

#### 4.4 MONTAGEM

Após as três etapas do sequenciamento, *L. lactis* NCDO 2118 passou por três diferentes etapas de montagem, realizadas em duas plataformas diferentes. Com os dados da terceira montagem (Ion PGM™) uma nova curadoria manual foi realizada para solução dos pseudogenes que ainda existiam.

##### **Montagem de dados oriundos de SOLiD 5500 XL**

Primeiramente, *L. lactis* NCDO 2118 foi sequenciada com SOLiD 5500 XL (Applied Biosystems) com biblioteca *mate-paired*. A montagem do genoma foi feita no CLC Genomics Workbench. As *reads* foram submetidas ao filtro de qualidade *Phred* 20 usando Quality Assessment (Ramos *et al.*, 2011). As sequências foram orientadas e ordenadas com base no genoma de referência *L. lactis* KF147 (uma linhagem associada à planta; número de acesso CP001834) (Siezen *et al.*, 2010). Os *softwares* SSPACE (Boetzer *et al.*, 2011) e Gapfiller (Nadalin *et al.*, 2012) foram usados para gerar o *scaffold* e resolver os *gaps*, respectivamente.

## Montagem de dados oriundos de Ion Torrent PGM

**1ª Montagem** - No Ion PGM™, o genoma foi sequenciado utilizando biblioteca de fragmentos e, para a sua montagem, foi utilizado o Mira (Chevreux *et al.*, 1999)(versão 3.9). A referência foi alinhada usando o CONTIGuator (Galardini *et al.*, 2011) e as sequências sobrepostas redundantes foram removidas com "*in-house scripts*", fechando os *gaps* remanescentes.

**2ª Montagem** - Na segunda montagem de *L. lactis* NCDO 2118, foram utilizadas bibliotecas de fragmentos. Nesta etapa, os *softwares* utilizados foram Mira (versão 4.0.1) e Newbler (Ohnishi *et al.*, 2014) (versão 2.9). Para realização do alinhamento e fechamento dos *gaps*, foram utilizados CONTIGuator e FGAP 1.7 (Piro *et al.*, 2014), respectivamente.

### 4.5 ANOTAÇÃO E CURADORIA DE *FRAMESHIFTS*

Do primeiro sequenciamento (SOLiD 5500XL), foi realizada a anotação manual do genoma *draft* de *L. lactis* NCDO 2118 e este, depositado como genoma incompleto no banco de dados do NCBI sob o número de acesso ASAE00000000.

Já do sequenciamento em Ion PGM™, obteve-se um genoma completo, que foi anotado manualmente e onde foram identificados genes e proteínas por meio de alinhamento de aminoácidos realizados no Uniprot (<http://www.uniprot.org/>). No processo de anotação manual, o arquivo .embl, representado abaixo, foi trabalhado no *software* Artemis®. Nesta etapa, cada CDS (regiões codificantes, do inglês *coding DNA sequence*) passou por uma verificação quanto à posição correta dos codons iniciais e finais utilizando a tabela de uso codon 11, de procariotos.

A partir daí, foi feita a curadoria manual de *frameshifts*, onde grande parte das regiões foram resolvidas através do primeiro sequenciamento. Pelo fato do genoma ainda apresentar muitos pseudogenes, um segundo sequenciamento em Ion PGM™ foi realizado. Deste segundo processo, foram utilizados apenas os dados de mapeamento para fechamento dos pseudogenes restantes.

- Arquivo .embl: formato de arquivo para anotação do genoma:

```
ID  NCDO_2118; circular; genomic DNA; STD; UNC; 2561391 BP.
FT  CDS                    582..1949
FT                          /locus_tag="L_lactis_NCDO_00002"
FT                          /colour=11
FT                          /product="Chromosomal replication initiator protein DnaA"
```

```

FT          /gene="dnaA"
FT          /note=""
FT   CDS    2106..3248
FT          /locus_tag="L_lactis_NCDO_00006"
FT          /colour=3
FT          /product="DNA polymerase III subunit beta"
FT          /gene="dnaN"
FT          /note=""
SQ   Sequence 2554693 BP; 833673 A; 443529 C; 448387 G; 829104 T; 0 other;
gattttcatt acttttttatt ttttcttatt ttttagtaag ataaggtgct tatttttagtc      60
ttctttacct tctgagaata aaattccttg atttttcacg tagaaatgat gtttttaagc      120
ttaactaaat ttttattata aataatagaa aattgaattc tcttttattt tgtagcaaat      180
agtttgctc  tatattccgt ctttatatag tagaaaactt aaaatttaac atttttatttt      240
caaattcctg gattttatta cattagagat aaaatatccc caaagataga actgactgtc      300
taaaagctgt ggtaaatfff tcttaatfff ttctttttcc acagtaaata atctataaat      360
ttgactffff ttgatttggtg gataagtctg tatgcatttg atattactca tttttttctg      420
tggataagaa aactttttaca cattatccac agttttgtta ataactffff ctatgggtgg      480
aaaactttct tttgttaatt tcaacagact gtggaaaaat ataatgtgaa atgatataat      540
gagttatcaa caaactaaaa agtaaaggag taatatggca tcccttaatg aaaatcaaaa      600
attttgggcc cgtgtaaccg aactagctca gcagagtatc ggtaacaag cttatgattt      660
ttttattgag cctgctcaat taatgtctgt cgagcaagat actgccaata tcttgcttga      720
tagtggaatg aaaaaagatt attggaaaaa acagtccgat ttgatcacia cggcgggatt      780
tgaggtcttt ggtcggatga ttgattatga actctatgca aatgatgagt taactgaact      840
ggaacttcat cgtttaaata atcaatcttc tattgaagag cagccgcgtt ctactgctaa      900
accgccttct ctttggtta gtggtctcaa cgagaaatat aattttgaaa attttggtgca      960

```

Além disso, as sequências de aminoácidos foram comparadas às sequências existentes no banco de dados do Uniprot, onde foi possível encontrar informações de proteínas curadas e informações quanto à função e classificação. O Uniprot busca a sequência de interesse em duas seções (i) Swiss-Prot, onde encontramos as anotações feitas manualmente e que foram revisadas e o (ii) TrEMBL, que conta apenas com as informações anotadas manualmente. As *features* de cada CDS foram preenchidas conforme os resultados do BLAST (*Basic Local Alignment Search Tool*) de proteínas feito contra o banco do Uniprot, levando em consideração a informação de produto revisado, sequências de outras *Lactococcus* com melhor identidade, valor do *e-value*, conservação no nome do gene, entre outros aspectos. Algumas CDS foram criadas com o intuito de preencher o máximo possível com informações sobre o genoma da *L. lactis* NCDO 2118, e na comparação feita no banco de dados, em sua maioria foram encontradas proteínas hipotéticas.

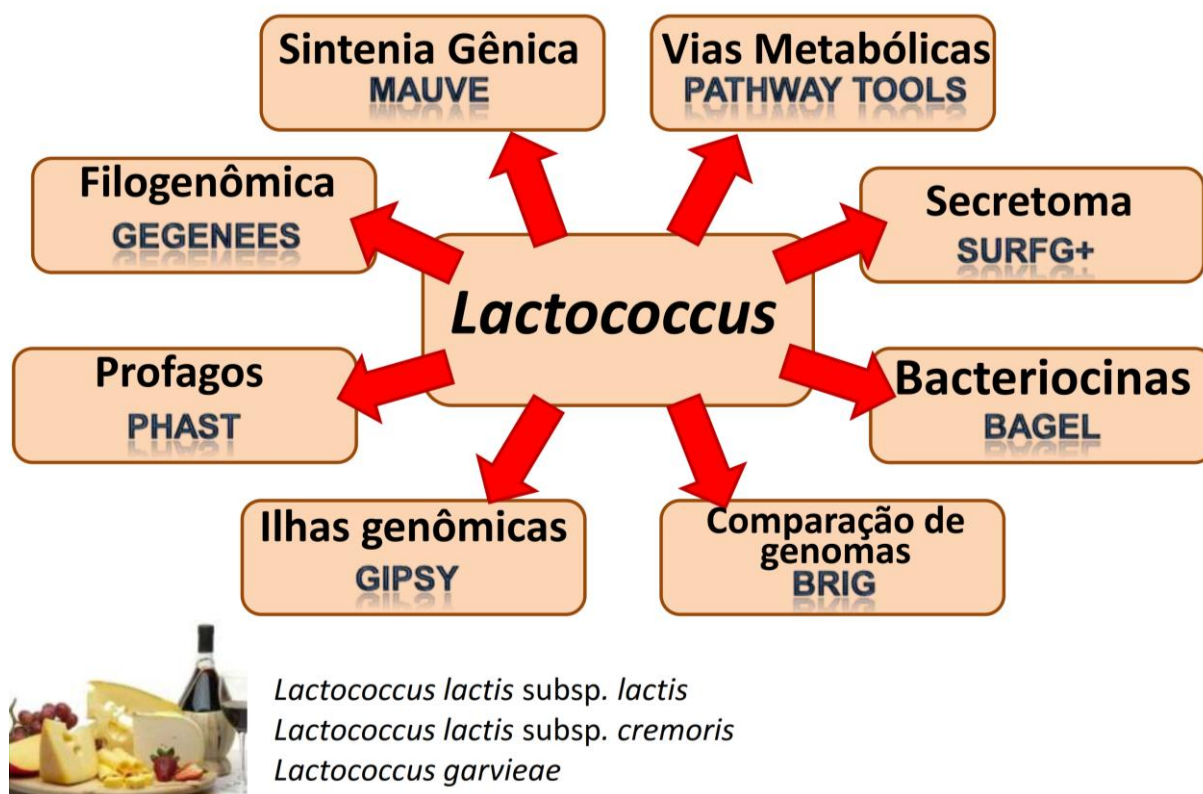
Na curadoria de *frameshifts*, as análises foram focadas nas CDS que apresentavam mudança na matriz de leitura, onde os pseudogenes gerados foram analisados um a um. Neste momento, os dados foram trabalhados conjuntamente em dois *softwares*: Artemis e CLC Genomics Workbench. Este último continha os dados do mapeamento das *reads* do sequenciamento com o genoma final fechado na montagem. Geralmente, os pseudogenes apresentavam uma repetição nucleotídica nas regiões de quebra, também chamada de homopolímeros, um erro

específico gerado pela química do sequenciamento utilizando Ion PGM™. Para corrigi-las, uma parte da sequência de nucleotídeos próxima à região de quebra dos prováveis pseudogenes foi selecionada e, em seguida, localizada no CLC. Por meio dele, foi possível observar a conservação das bases pelo alinhamento das *reads* e, da cobertura visualizada na região específica, correções como: deleção, inserção, ou troca de bases à sequência, foram feitas nos dados em análise do Artemis.

Após a curadoria de *frameshifts*, cada sequência passou por uma nova verificação de alinhamento no BLAST do Uniprot, para assegurar que as CDS estavam posicionadas corretamente após a correção na matriz de leitura.

## 4.6 COMPARAÇÃO GENÔMICA

Com a sequência de *L. lactis* NCDO 2118 curada, seu genoma foi comparado com os genomas de outras *Lactococcus* depositadas no banco de dados GENBANK do NCBI (**Tabela 2**). Destas sequências, 13 espécies de *Lactococcus* com genoma completo foram utilizadas, com o objetivo de não gerar um viés nos resultados. As comparações das espécies de *Lactococcus* foram realizadas por meio das seguintes ferramentas de bioinformática dispostas na **Figura 8**.



**Fig. 8** Pipeline das etapas de comparação genômica do gênero *Lactococcus* (Este trabalho).

**Tabela 2.** Lista dos genomas completos utilizados para comparação genômica. Características genômicas das espécies de *Lactococcus* que possuem genoma completo utilizadas neste estudo.

| Nome  | Tamanho (Mb) | GC%   | Genes | Proteínas | Isolado                            | Nº de Acesso | Plasmídeos | Pseudogenes |
|---|--------------|-------|-------|-----------|------------------------------------|--------------|------------|-------------|
| <i>Lactococcus lactis</i> subsp. <i>lactis</i> IL1403     | 2.36559      | 35.30 | 2406  | 2277      | Produção de queijo                 | AE005176     | -          | 45          |
| <i>Lactococcus lactis</i> subsp. <i>lactis</i> KF147      | 2.63565      | 34.86 | 2662  | 2473      | Feijão mungo                       | CP001834     | 1          | 93          |
| <i>Lactococcus lactis</i> subsp. <i>lactis</i> KLDS 40325 | 2.59549      | 35.39 | 2732  | 2593      | Leite fermentado                   | CP006766     | 1          | 56          |
| <i>Lactococcus lactis</i> subsp. <i>lactis</i> IO-1       | 2.42147      | 35.10 | 2318  | 2224      | Água (poço de drenagem de cozinha) | AP012281     | -          | -           |
| <i>Lactococcus lactis</i> subsp. <i>lactis</i> CV56       | 2.51874      | 35.09 | 2549  | 2408      | Vagina de mulher saudável          | CP002365     | 5          | 51          |
| <i>Lactococcus lactis</i> subsp. <i>cremoris</i> A76      | 2.5771       | 35.88 | 2845  | 2769      | Produção de queijo                 | CP003132     | 4          | -           |
| <i>Lactococcus lactis</i> subsp. <i>cremoris</i> KW2      | 2.42705      | 35.70 | 2353  | 2268      | Milho fermentado                   | CP004884     | -          | 1           |
| <i>Lactococcus lactis</i> subsp. <i>cremoris</i> MG1363   | 2.52948      | 35.70 | 2597  | 2434      | Produção de queijo                 | AM406671     | -          | 82          |
| <i>Lactococcus lactis</i> subsp. <i>cremoris</i> NZ9000   | 2.53029      | 35.70 | 2594  | 2510      | Produção de queijo                 | CP002094     | -          | -           |
| <i>Lactococcus lactis</i> subsp. <i>cremoris</i> SK11     | 2.59835      | 35.82 | 2739  | 2501      | Produção de queijo                 | CP000425     | 5          | 144         |
| <i>Lactococcus lactis</i> subsp. <i>cremoris</i> UC5099   | 2.45735      | 35.76 | 2401  | 2109      | Produção de queijo                 | CP003157     | 8          | 188         |
| <i>Lactococcus garvieae</i> ATCC49156                     | 1.95013      | 38.80 | 2024  | 1947      | <i>Alosa fallax</i>                | AP009332     | -          | 0           |
| <i>Lactococcus garvieae</i> Lg2                           | 1.96396      | 38.80 | 2045  | 1968      | <i>Alosa fallax</i>                | AP009333     | -          | 0           |



#### 4.6.1 FILOGENÔMICA

As sequências dos genomas completos das 13 linhagens (**Tabela 2**) mais a nossa linhagem de interesse (*L. lactis* NCDO 2118) foram submetidas ao *software* GENEIES (Agren *et al.*, 2012) para realização das análises filogenéticas. Para isto, foram utilizados os arquivos de entrada no formato .fna, como representado abaixo:

- Arquivo .fna: sequência fasta de nucleotídeos:

```
>Lactococcus lactis subsp. lactis NCDO_2118
gattttcattactttttattttttttagtaagataaggtgcttatttttagtc
ttctttaccttctgagaataaaattctttgattttttcacgtagaaatgatgttttaagc
ttaactaaattttattataaataatagaaaattgaattctcttttattttgtagcaaat
agtttgctctatattccgtctttatatagtagaaaacttaaaatttaacattttatttt
caaattcttgattttattacattagagataaaaatatccccaagatagaactgactgctc
taaaagctggtgtaaatttttcttaattttttctttttccacagtaaaatactataaat
ttgactttttttgatttgggataagctctgtatgcatttgatattactcattttttctg
tggataagaaaacttttacacattatccacagttttgttaataactttttcttatgggtgg
aaaactttctttttgtaatttcaacagactgtggaaaaataaatgtgaaatgatataat
gagttatcaacaaactaaaaagtaaaggagtaaatatggcatcccttaatgaaaatcaaaa
attttgggcccgtgtaaccgaactagctcagcagagatcggtaacaagcttatgattt
ttttattgacctgctcaattaatgtctgtcagcaagatactgccaatatcttgcttga
tagtggaatgaaaaaagattattggaaaaaacagctccgatttgatcacaacggcgggatt
tgaggtctttggctggatgattgattatgaactctatgcaaatgatgagttactgaact
ggaacttcatcgtttaataatcaatcttctattgaagagcagccgcttctactgctaa
accgcttctcctttgggttagtggtctcaacgagaaataaattttgaaaattttgtgca
aggctctggtaaccgttggacttttagctgcagcaattgctgttgctgataaacctgggtga
tacctataatccactatttatttatgggtggagctgggtctaggttaagacgcatttaatgca
tgcaattggaaatcagattttgactgacaatccgacggctcgaatataatgtttcttc
ggagaatttcgttaacgactatgtaaatgcaacgcgaaaaaaccaaatggagaattttga
aaatacttatcgtaatttagatttacttctattggacgatgtgcaatttttcagtataaa
agagggcagaaaaatgaatttttcaataccttcaatgcactctatgataaaggttccca
aattgttttaacttctgatcgtattccgcaagaattaataaatttggagatcgcttgggt
ttcacgcttctcttggggttggactactgatattacggctcctgactatgaaactcggat
ggctattcttttgattaaatctgaatcaagtcggttagaatttccgagtgaaactctatc
ttatattgctggctcagattgattcaaacgttcgcaactcagggcgctctaaatcgagt
```

Os genomas de *Lactococcus* foram submetidos ao banco de dados do GENEIES, por meio do diretório local. Como *outgroup*, objetivando enraizar a árvore, foi adicionado ao conjunto de dados o genoma de uma espécie filogeneticamente relacionada ao gênero *Lactococcus*, *Streptococcus thermophilus* LMD-9. A seguir, foi criado um projeto com os parâmetros *default* para a análise comparativa, utilizando o método BLASTN de alinhamento. Feita a adição das sequências, o GENEIES iniciou o processo de alinhamento de todos contra todos fragmentos gerados de cada um dos 15 genomas. Ao final, uma imagem do gráfico *heatmap* foi criada e estes resultados foram salvos na extensão .nexus, como mostrado a seguir, para criação da árvore no SplitsTree.

Para gerar a árvore, a matriz de distância na extensão .nexus (gerada no GENEIES) foi submetida ao SplitsTree e, em seguida, foram especificados os

parâmetros para a geração de um cladograma conforme agrupamento pelo método de *Neighbor-joining* (NJ).

- Arquivo .nexus: arquivo (de saída do GENEIES e de entrada para o SplitsTree) para construção de dendograma:

```
#NEXUS

[! score_0_percent_threshold_15_genomes]

BEGIN taxa;
  DIMENSIONS ntax=15;
  TAXLABELS

    L_garvieae_ATCC49156
    L_garvieae_Lg2
    S_thermophilus_LMD-9
    L_lactis_KF147
    L_lactis_NCDO2118
    L_lactis_IO_1
    L_lactis_CV56
    L_lactis_IL1403
    L_lactis_KLDS_40325
    L_lactis_cremoris_KW2
    L_lactis_cremoris_NZ9000
    L_lactis_cremoris_MG1363
    L_lactis_cremoris_SK11
    L_lactis_cremoris_A76
    L_lactis_cremoris_UC5099

;
END;

BEGIN distances;
  DIMENSIONS ntax=15;
  FORMAT triangle=LOWER diagonal labels missing=?;
  MATRIX

    L_garvieae_ATCC49156          0.0
    L_garvieae_Lg2          0.009221727017042713          0.0
    S_thermophilus_LMD-9          0.9811683939545507          0.9810914442411883          0.0
    L_lactis_KF147          0.9650613657396416          0.9630751236769132
    0.9839142962098514          0.0
    L_lactis_NCDO2118          0.9645883443040977          0.9627173842139596
    0.9842491168322871          0.02093927110071936          0.0
```

#### 4.6.2 SINTENIA GÊNICA

Nesta etapa, os genomas foram adicionados ao *software* Mauve para a análise comparativa entre as *Lactococcus*. Todas as sequências se apresentavam na extensão .fna e, como parâmetro padrão, o alinhamento foi realizado no "*progressiveMauve*", por apresentar uma abordagem com maior acurácia. Ao finalizar a comparação, o Mauve dispôs os genomas em forma de blocos gênicos. Mesmo que a ordem gênica seja diferente, os blocos são conectados de acordo com a sintenia existente entre eles, onde é possível observar características específicas dentro de cada sequência.

#### 4.6.3 PREDIÇÃO DE VIAS METABÓLICAS

Os dois principais arquivos utilizados para reconstrução das vias metabólicas das espécies de *Lactococcus* foram sequências no formato FASTA e o genoma anotado no formato GBK. Posteriormente, cada uma das 15 linhagens foram computacionalmente preditas no Pathway/Genome Database (PGDB) usando o *software* Pathway Tools versão 16.5 (Karp *et al.* 2002), desenvolvido pela SRI Internacional. O MetaCyc, um banco de dados de metabolismo de pequenas moléculas, altamente curado e não redundante foi usado como um banco de dados de referência para o PathoLogic, componente do *software* Pathway Tools (Caspi *et al.*, 2010). As vias metabólicas de *Lactococcus lactis* NCDO 2118 foram usadas como referência para análise comparativa de acordo com as seguintes estratégias: i) comparação apenas entre linhagens filogeneticamente relacionadas: *L. lactis* NCDO 2118, *L. lactis* KF147 e *L. lactis* IL1403; ii) comparação com as espécies não-patogênicas (*L. lactis* subsp. *lactis* e *L. lactis* subsp. *cremoris*); iii) comparação com todas as linhagens de *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris* e *L. garvieae*.

#### 4.6.4 PLASTICIDADE GENÔMICA

Utilizando o recurso de interface gráfica do GIPSy, as sequências de duas linhagens foram adicionadas à ferramenta no formato .gbk (GenBank), apresentado a seguir. Nesta etapa, foram criados os arquivos para cada organismo em diretório predefinido do próprio *software*. Adiante, os dados passaram pela análise de conteúdo G+C e, em seguida, foram executadas as análises para uso de codon. Na etapa quatro, foi feita uma busca por genes de transposases no genoma de interesse (*L. lactis* NCDO 2118). No passo seguinte, foi possível selecionar quais fatores específicos seriam analisados quanto à presença de ilhas genômicas.

Para este trabalho, os dados foram executados para predição de ilhas metabólicas e simbióticas com valor de *e-value* de  $10^{-6}$ . A seguir, os genomas passaram pelo BLAST de proteínas (BLASTP), onde todas as CDS do genoma de interesse foram comparadas contra todas as CDS do genoma de *L. lactis* subsp. *cremoris* A76. A etapa sete consistiu na busca por genes de tRNA no genoma de *L. lactis* NCDO 2118. E por fim, com base nas características até então identificadas,

foi realizada a predição das ilhas genômicas (<http://www.bioinformatics.org/ftp/pub/GIPSy/Manual.pdf>).

Uma análise também foi feita entre as espécies de *Lactococcus*, utilizando uma linhagem de cada subespécie para observação. Assim, nas análises de ilhas de patogenicidade, o genoma de referência utilizado foi de *L. garvieae* Lg2. Já para a análise de ilhas genômicas, foi utilizado o genoma de referência de *L. cremoris* MG1363. Sendo tais referências comparadas com todas as outras espécies de *Lactococcus* deste presente estudo.

- Arquivo .gbk: formato de arquivo para genoma completo no padrão GenBank.

```
LOCUS       pNCDO2118                37571 bp    DNA     circular     01-JUL-2014
DEFINITION  Lactococcus lactis subsp lactis pNCDO2118.
ACCESSION
VERSION
KEYWORDS
SOURCE      Lactococcus lactis subsp lactis pNCDO2118
  ORGANISM  Lactococcus lactis subsp lactis pNCDO2118
            Unclassified...
  TITLE     Genome sequence of Lactococcus lactis subsp. lactis NCDO 2118, a
            GABA-producing strain...

            CDS           complement(2078161..2078331)
                        /colour=
                        /gene=""
                        /product="Hypothetical Protein"
                        /locus_tag="NCDO_1936"
            CDS           complement(2078396..2079088)
                        /colour=
                        /gene=""
                        /product="Hypothetical protein"
                        /locus_tag="NCDO_1937"
BASE COUNT  833673 a 443529 c 448387 g 829104 t
ORIGIN
1  gattttcatt  actttttatt  ttttcttatt  ttttagtaag  ataagggtgct  tatttttagtc
61  ttctttacct  tctgagaata  aaattccttg  atttttcacg  tagaaatgat  gtttttaagc
121  ttaactaaat  ttttattata  aataatagaa  aattgaattc  tcttttattt  tgtagcaaat
181  agtttgctc  tatattccgt  ctttatatag  tagaaaactt  aaaatttaac  attttatttt
241  caaattcttg  gattttatta  cattagagat  aaaatatccc  caaagataga  actgactgtc
301  taaaagctgt  ggtaaatttt  tcttaatttt  ttctttttcc  acagtaaata  atctataaat
361  ttgacttttt  ttgattttgt  gataagctct  tatgcatttg  atattactca  tttttttctg
421  tggataagaa  aacttttaca  cattatccac  agttttgtta  ataacttttt  cttatgggtg
481  aaaactttct  tttgttaatt  tcaacagact  gtggaaaaat  ataatgtgaa  atgatataat
541  gagttatcaa  caaactaaa  agtaaaggag  taatatggca  tcccttaatg  aaaatcaaaa
601  attttgggcc  cgtgtaaccg  aactagctca  gcagagtatc  ggtaacaag  cttatgattt
661  ttttattgag  cctgctcaat  taatgtctgt  cgagcaagat  actgccaata  tcttgcttga
721  tagtggaatg  aaaaaagatt  attggaaaa  acagtccgat  ttgatcacia  cggcgggatt
```

Para estas análises, foram utilizados os valores padrão para os parâmetros pré-definidos em cada etapa, como por exemplo, valor de sensibilidade e de *e-value*. Os resultados gerados foram armazenados automaticamente dentro de uma pasta específica criada pelo próprio *software*. Nela, foram obtidos dois arquivos para cada ilha genômica predita, um em formato .txt (bloco de notas) e outro .tab (tabular). A partir do arquivo .tab, foram coletadas todas as coordenadas preditas das ilhas

metabólicas e simbióticas e estes foram adicionados ao BRIG (Blast Ring Image Generator) (Alikhan *et al.*, 2011) para posterior visualização.

#### 4.6.5 PREDIÇÃO DE PROFAGOS

O arquivo no formato .gbk foi submetido ao programa Phast e, em seguida, foram gerados os resultados em forma de tabela com detalhes das predições contendo figuras com as características e localização das regiões preditas. A partir dos dados disponibilizados sobre as regiões preditas, as coordenadas foram adicionadas aos dados para visualização no BRIG.

#### 4.6.6 PREDIÇÃO DE BACTERIOCINAS

O genoma de *L. lactis* NCDO 2118 foi adicionado ao *software* BAGEL no formato .fna por meio de diretório local. A partir daí, foram preditas bacteriocinas presentes no genoma da linhagem, por meio de dados detalhados das sequências (posição, tamanho e nomenclatura do PFAM das áreas envolvidas), além da geração de uma imagem mostrando a disposição das regiões que foram preditas.

#### 4.6.7 REPRESENTAÇÃO CIRCULAR DA COMPARAÇÃO DAS SEQUÊNCIAS GENÔMICAS

Para a visualização dos genomas, todas as sequências foram adicionadas ao *software* BRIG (*BLAST Ring Image Generator*) no formato .fna, onde *L. lactis* NCDO 2118 foi selecionada como referência (ao centro da imagem circular) em uma das comparações. Em seguida, os demais genomas contidos em um diretório local foram adicionados, onde foi criado o banco de dados utilizado para realização do BLAST.

No BRIG, após dispor corretamente os arquivos, foram criados os anéis correspondentes às respectivas sequências genômicas, juntamente com as coordenadas das ilhas genômicas preditas anteriormente no GIPSY. Somado aos resultados supracitados, o arquivo com as coordenadas das regiões de fago preditas no genoma, utilizando o *software* PHAST, foi submetido.

Na última etapa, um BLAST de todos os genomas com a referência foi realizado e os resultados foram convertidos em um formato específico, onde foi possível a criação de uma imagem circular de todos os genomas.

#### 4.6.8 IDENTIFICAÇÃO DAS PROTEÍNAS DO SECRETOMA

A provável localização subcelular de proteínas em *L. lactis* NCDO 2118 foi identificada *in silico* por meio do SurfG+. Este *software* incorpora ferramentas como *SignalP*, *LipoP* e *TMHMM* para identificação de motivos (Santos *et al.*, 2012).

Em busca de uma melhor diferenciação entre as proteínas de membrana (MEM) das proteínas expostas à superfície (PSE), foram utilizadas nestas análises as medições da parede da membrana, por meio de microscopia eletrônica com um equipamento EM10A (Zeiss) (Soares *et al.*, 2013).

*L. lactis* NCDO 2118 foi crescida em meio M17 (Difco) a 30°C, contendo 0,5% de glicose (GM17), sem agitação durante 18h (Luerce *et al.*, 2014). O precipitado (~500 mL) resultante foi colocado em um tubo Eppendorf, fixado em 2,5% de gluteraldeído a 0,1M de tampão cacodilato de sódio (pH 7,2) por 6h a 8°C e lavado três vezes a 0,1M com tampão cacodilato de sódio (pH 7,2). Depois da lavagem do padrão, a amostra foi pós fixada em 1% de tetróxido de ósmio em 0,1M de tampão cacodilato de sódio (pH 7,2) + 1,5% de ferrocianeto de potássio durante 90 minutos, lavado com 0,1M com tampão cacodilato de sódio (pH 7,2), desidratado em etanol graduado (50% EtOH, 70% EtOH, 95% EtOH, and 100% EtOH) e incorporado em resina de Eponate–Araldite. Cortes ultrafinos foram obtidos com a utilização de acetato de uranilo e citrato de chumbo e posteriormente, examinado no Zeiss-EM-10A (de Melo *et al.*, 1993). As micrografias foram obtidas por uma câmera *CCD Mega View*.

Para medição da parede, foram utilizadas pelo menos, cinco micrografias de *L. lactis* NCDO 2118 no tamanho de 50 e 100 mil vezes. Destas micrografias, ao menos 10 medições por célula foram realizadas em cada foto, chegando a um valor médio. Os valores foram calculados em regra de três simples entre, o valor em *pixels*, a medida de cobertura de campo em nanômetros (Nm) e as medições encontradas nas diferentes imagens. A média da espessura da parede obtida no cálculo total foi de ~55 aminoácidos. Este valor foi adicionado ao SurfG+ juntamente com a sequência fasta de aminoácidos (.faa) da linhagem de interesse.

Em seguida, a ferramenta online BLASTclust foi utilizada. Por meio dela, foi feita uma predição de genes ortólogos e parálogos com o objetivo de comparar as proteínas identificadas em *L. lactis* NCDO 2118 com as proteínas do genoma de *L. lactis* IL1403, confirmada por Luerce e colaboradores, em 2014, ser uma espécie que não possui capacidade probiótica (Luerce *et al.*, 2014).

## 5. RESULTADOS E DISCUSSÃO

### 5.1 MONTAGENS

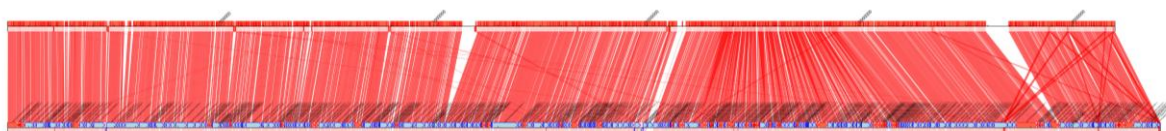
Pelo fato do genoma de *L. lactis* NCDO 2118 não ter sido fechado na primeira montagem em SOLiD 5500XL, um outro sequenciamento foi realizado utilizando a plataforma Ion PGM™. Nesta etapa, objetivando eliminar a grande quantidade de pseudogenes ainda existentes da etapa anterior (1º sequenciamento em Ion PGM™), um segundo sequenciamento nesta mesma plataforma foi realizado, conforme mostra a tabela detalhada abaixo.

| Plataformas         | SOLiD 5500XL         | 1º Ion PGM™     | 2º Ion PGM™     |
|---------------------|----------------------|-----------------|-----------------|
| <b>Cobertura</b>    | 1785x                | 71,07x          | 474x            |
| <b>Biblioteca</b>   | <i>Mate-paired</i>   | <i>Fragment</i> | <i>Fragment</i> |
| <b>Montador</b>     | CLC Genomics         | Mira            | Mira + Newbler  |
| <b>CDS</b>          | -                    | 4.387           | -               |
| <b>Anotação</b>     | -                    | 2.470           | -               |
| <b>Pseudogenes</b>  | -                    | 1.821           | -               |
| <b>Curadoria de</b> | -                    | 480 pseudogenes | 52 pseudogenes  |
| <b>Genoma</b>       | 409 <i>scaffolds</i> | -               | 52 pseudogenes  |

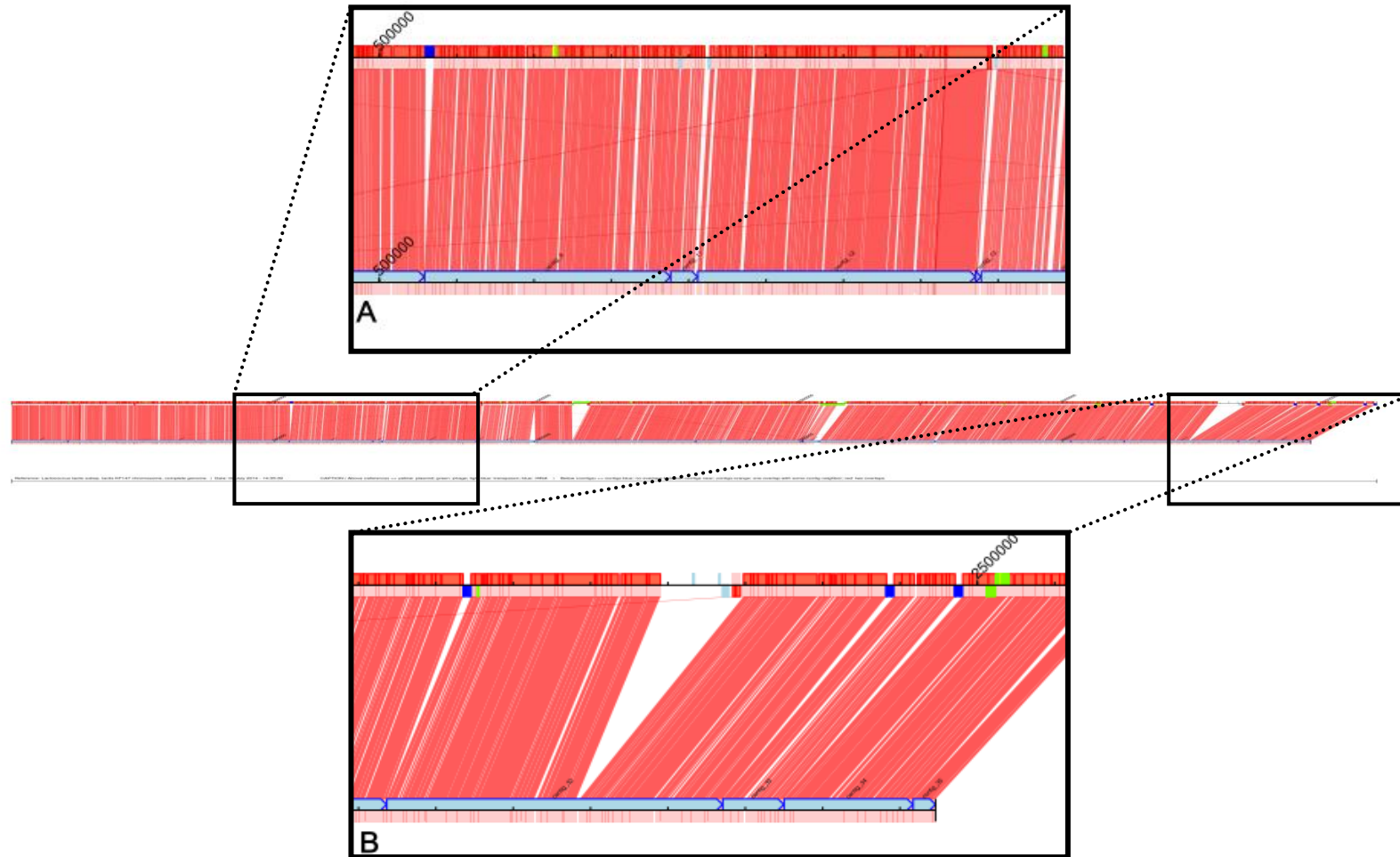
**Tabela 3.** Resultados das diferentes abordagens de sequenciamento, montagem e anotação do genoma de *L. lactis* NCDO 2118. Destaque para os parâmetros utilizados para obtenção do genoma *draft* (SOLiD 5500XL) e os processos realizados para obtenção do genoma completo (1º e 2º sequenciamento em Ion Torrent PGM). (Este trabalho, 2014).

Utilizando dados oriundos de SOLiD 5500XL, foram geradas 85.550.956 *reads* de 60 pares de base (5.133.057.360 bp), equivalente a uma sequência genômica com cobertura de 1785 vezes. No CLC Genomics Workbench, foi gerado um total de 1.641 sequências de *overlap*, sendo que destas sequências, as redundantes foram removidas com o software Simplifier. Ao final da curadoria manual, foi obtido um total de 409 *scaffolds* (2.874.854 bp).

O genoma escolhido como referência (*L. lactis* KF147) na montagem da *L. lactis* NCDO 2118, pode ser explicado com base na alta similaridade observada na pré-análise feita com o programa CONTIGuator (**Fig. 9**). Pelo fato de o genoma não ter sido fechado neste momento, foi feito um novo sequenciamento, desta vez em Ion PGM™, com o uso da mesma referência (**Fig. 10**).



**Fig. 9.** Sintonia entre *L. lactis* NCDO 2118 (*draft*) e *L. lactis* KF147.



**Fig. 10.** Sintenia entre *L. lactis* KF147 e *L. lactis* NCDO 2118 (genoma completo). Imagem representativa do alinhamento do genoma de referência (fita superior) com o genoma em estudo (fita inferior). **A)** Na fita superior, em azul escuro, estão representadas regiões de rRNA; em verde, regiões de fago e em azul claro, transposases. **B)** Na fita inferior o azul claro representa a ausência de *overlap* entre os *contigs* na sequência.



Com relação aos dados de sequenciamento em Ion PGM™, foram gerados um total de 187.303.001 bp (cobertura de ~71 vezes). Devido à grande quantidade de pseudogenes restantes (mais de 400) desta etapa, um segundo sequenciamento em Ion PGM™ foi realizado. Desta vez, um chip 318 completo foi preenchido com *L. lactis* NCDO 2118, para gerar uma quantidade maior de dados, no intuito de alcançar uma maior cobertura do genoma e, por conseguinte, facilitar a resolução dos *frameshifts*.

Durante a segunda montagem feita em Ion PGM™, foi gerado um total aproximado de 1.249.154.478 bp (cobertura de 474 vezes). Nesta última etapa, foi possível reduzir consideravelmente a quantidade de pseudogenes.

Durante as montagens feitas com a linhagem *L. lactis* NCDO 2118, foi identificado uma quantidade de *reads* que não foram alinhadas com nenhuma parte do genoma, sendo, então, comparadas com a sequência do plasmídeo do genoma de referência. Assim, foi encontrado na *L. lactis* NCDO 2118, um plasmídeo de 37.571 pb (número de acesso CP009055).

## 5.2 ANOTAÇÃO E CURADORIA DE *FRAMESHIFTS*

Inicialmente, após transferência da *L. lactis* NCDO 2118 (*draft*), a sequência completa vista no Artemis apresentou um tamanho de aproximadamente 2,5 mb. O arquivo a ser anotado possuía 4.387 CDS. Levando em consideração que genomas bacterianos são altamente compactos e tendem a apresentar um tamanho médio de 1000 pb por gene, espera-se que um genoma de 2,5 mb tenha aproximadamente 2500 genes. Fazendo correlação do tamanho do genoma com a quantidade de CDS, foi cogitada a presença de uma grande quantidade de *frameshifts* (**Fig. 11**).

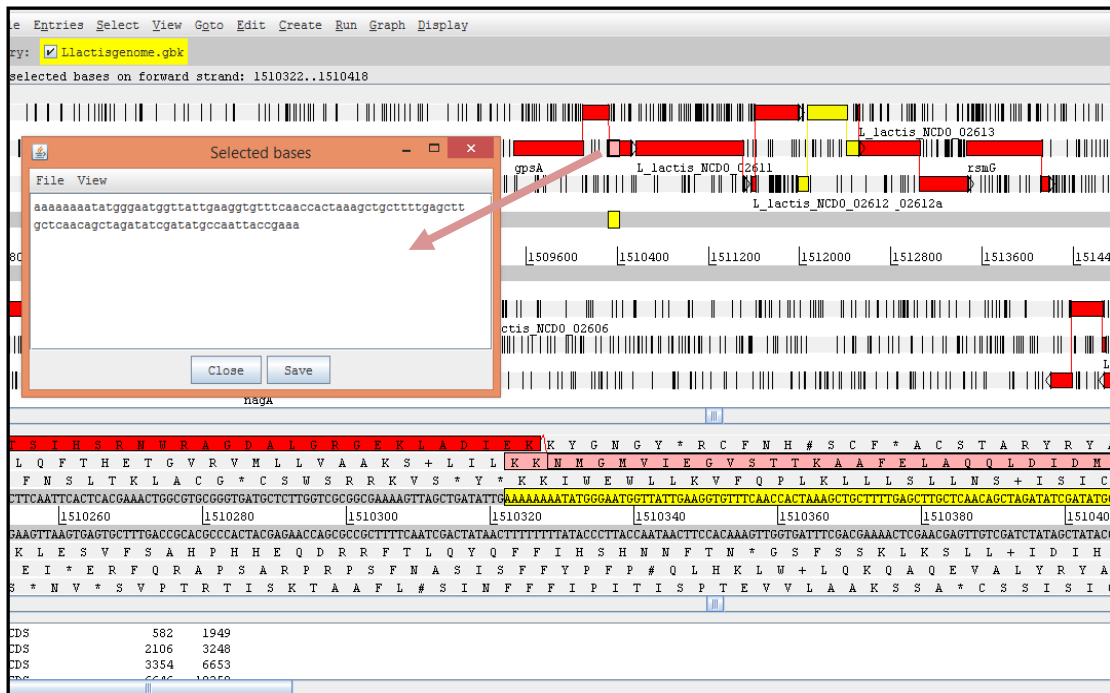
Terminada a anotação manual, a quantidade de CDS do genoma caiu de 4.387 para aproximadamente 2.470, onde 26% destas foram identificadas como proteínas hipotéticas. Além de tais características, o cromossomo de *L. lactis* NCDO 2118 possui 66 genes de tRNA, 6 operons de rRNA e um conteúdo G+C de 34,9%.

A linhagem de interesse precisou de dois processos de curadoria de *frameshifts*. No primeiro, ao final da anotação manual, o genoma continha 1821 pseudogenes. Destes, 1341 puderam ser solucionados, restando ainda 480 pseudogenes em todo genoma. Diante da dificuldade em solucionar os *frameshifts*, uma nova curadoria foi realizada por meio dos dados da montagem do segundo seqüenciamento em Ion PGM™, que apresentava uma cobertura maior do genoma. Assim, dos 480 pseudogenes remanescentes no cromossomo de *L. lactis* NCDO 2118, restaram apenas 52 no arquivo final.

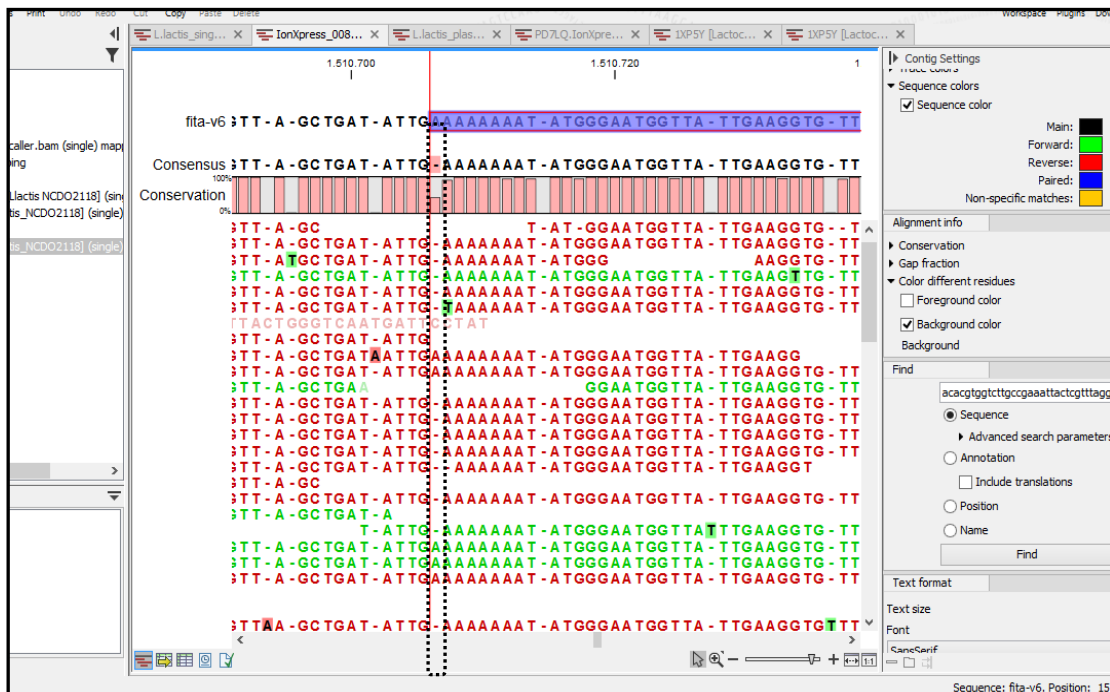
O plasmídeo de 37,5 mb, identificado nas montagens, continha 66 CDS. Após a anotação, a quantidade diminuiu para 49; aproximadamente 48% destas CDS foram identificadas como proteínas hipotéticas. Além das características já mencionadas, o plasmídeo possui um conteúdo G+C de 32,3%. Com os dados do primeiro seqüenciamento de Ion PGM™, foi possível resolver a maioria dos pseudogenes encontrados na sequência, restando apenas 4 deles no genoma final.

As três etapas de seqüenciamento, montagem e curadoria manual, mostraram a complexidade do organismo, pois ele não pôde ser fechado na primeira etapa do seqüenciamento, utilizando SOLiD 5500XL. O genoma completo foi obtido apenas com o segundo seqüenciamento em Ion PGM™, pois a quantidade de pseudogenes não puderam ser resolvidos na etapa anterior. Como resultado do trabalho do nosso grupo de pesquisa, a sequência do genoma de *L. lactis* NCDO

2118 foi submetida ao NCBI e a publicação deste estudo foi aceita na revista *Genome Announcement* (Anexo I).



A



B

**Fig. 11.** Curadoria manual de *frameshift*. **A.** Uma região de ~100 nucleotídeos em um local de quebra é selecionada e copiada no **B.** CLC. Nele, é encontrada a sequência (em azul) anteriormente destacada no Artemis. Na imagem, de acordo com a cobertura das *reads* (sequências em verde e vermelho), é indicada, pela fita "Consensus" a deleção de uma base de adenina (marcado em rosa), alterando as repetições de oito para sete adeninas. Onde "fita v6"= sequência do genoma de interesse mapeado contra as *reads* geradas no sequenciamento; *Consensus* = fita consenso gerada a partir do mapeamento; *Conservation* = porcentagem de conservação da base nucleotídica em relação a cobertura de *reads*.

### 5.3 COMPARAÇÃO GENÔMICA

As BL são caracterizadas pelo seu baixo conteúdo G+C e muitos genomas destas bactérias tiveram a sua capacidade de biossíntese reduzida resultante dos eventos de decaimento gênico, que refletem na sua adaptação a ambientes ricos em nutrientes, como o leite (Pfeiler & Klaenhammer, 2007).

De acordo com os dados apresentados (**Tabela 2**), é possível observar um padrão quanto ao perfil das espécies. O tamanho do genoma entre elas varia de 1,96 mb a 2,63 mb, onde as duas linhagens de *L. garvieae* se destacam com os menores genomas. Porém, estas duas linhagens se diferem das demais espécies de *Lactococcus* quanto à sua maior porcentagem de conteúdo G+C, 38,80%. Tratando-se deste aspecto, os menores valores são das linhagens *L. lactis* NCDO 2118 e *L. lactis* KF147, com 34,86%.

A relação filogenética entre as linhagens são refletidas nas diferenças da sequência e conteúdo cromossomal, embora sua adaptação a diferentes ambientes também esteja relacionada à aquisição ou perda de elementos móveis como fagos, plasmídeos, etc (Siezen *et al.*, 2011).

Das espécies listadas na **Tabela 2**, apenas 7 delas possuem plasmídeo, sendo elas: *L. lactis* NCDO 2118, *L. lactis* KF147, *L. lactis* KLDS 40325, *L. lactis* CV56, *L. lactis* A76, *L. cremoris* SK11 e *L. cremoris* UC5099, onde esta última se destaca com a maior quantidade de plasmídeos das espécies em análise.

#### 5.3.1 ANÁLISE COMPARATIVA ATRAVÉS DE SIMILARIDADE NUCLEOTÍDICA E FILOGENÔMICA

Uma abordagem utilizada para as análises de genômica comparativa com *Lactococcus* foi feita usando o Gegenees. Por meio dos genomas submetidos ao programa, uma matriz de distância foi disposta em um *heatmap*.

Aplicando uma abordagem de alinhamento de genomas fragmentados, esta ferramenta realiza análises comparativas entre vários genomas. Antes da disponibilização dos dados de entrada para o *software*, é necessário configurar o diretório correto do BLAST (se existente), ou instalar a versão que estiver disponível pelo próprio programa do Gegenees. Em seguida, é possível acessar os dados a serem analisados de duas maneiras diferentes: através de diretório local, ou usando

o banco de dados do NCBI através do *File Transfer Protocol* (FTP): <ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Bacteria/>.

Todas as sequências foram convertidas em fragmentos menores para a criação de um banco de dados na etapa de comparação, o qual representa todas as regiões genômicas de cada organismo de interesse. Deste banco, os alinhamentos foram calculados e os melhores resultados de BLAST das sequências foram coletados. Gegenees disponibiliza o BLASTN (comparação de nucleotídeos) ou TBLASTX (similaridade em nível de aminoácidos traduzidos a partir da sequência de nucleotídeos nas seis matrizes de leitura) para a realização dos alinhamentos. Para criação da matriz, o Gegenees apresentou os resultados de análise em forma de um gráfico *heatmap*, onde as cores variaram numa escala de verde (alta similaridade) a vermelho (baixa similaridade), equivalentes às porcentagens de similaridade existentes nos fragmentos melhor alinhados. Com a matriz de distância criada, foi possível, além de obter a figura em formato “.html”, usar também a opção de arquivo final para criação de uma árvore filogenômica no Splitstree (Huson & Bryant, 2006).

As espécies de *Lactococcus* e *Streptococcus thermophilus* LMD-9 foram organizadas em um gráfico de cores de acordo com o grau de similaridade existente entre elas, juntamente com uma árvore filogenética criada do SplitsTree (**Fig. 12**).

É possível observar que a similaridade é maior quando se trata das subespécies de *Lactococcus*. Dois blocos maiores em verde foram gerados no gráfico: um entre *Lactococcus lactis* subsp. *lactis* (posições de 2 a 7), onde a similaridade gerada foi em torno de 94% a 100% e outro entre *Lactococcus lactis* subsp. *cremoris* (posições de 8 a 13), com similaridade de 92% a 100%; um bloco menor foi formado com as duas linhagens de *Lactococcus garvieae* (posições 14 e 15), onde as duas linhagens foram 100% similares.

A espécie *L. garvieae* é um importante patógeno de peixe, mas apenas a linhagem Lg2 é virulenta. Esta última tem como característica a presença de um conjunto de genes de cápsula de 16,5 kb, ausente na linhagem ATCC 49156. Essa diferença na porcentagem de similaridade se deve a inserção do *cluster* de genes de cápsula (Morita *et al.*, 2011). Já, se comparados aos dados de semelhança entre estas duas linhagens com as demais espécies, observamos que a maior porcentagem de similaridade não passa de 65%.

Contudo, há de se considerar que o Gegenees realiza a análise baseando-se nas regiões conservadas de todos os genomas. Levando em consideração a

presença do genoma de *S. thermophilus* como referência, a parte conservada do genoma se torna mínima. Além disso, foi utilizado um *threshold* de 40% de estringência. Logo, dentre as regiões conservadas entre todos os genomas, *L. garvieae* apresenta de 59% a 65% de similaridade contra *L. lactis* subsp. *lactis* e *L. lactis* subsp. *cremoris*.

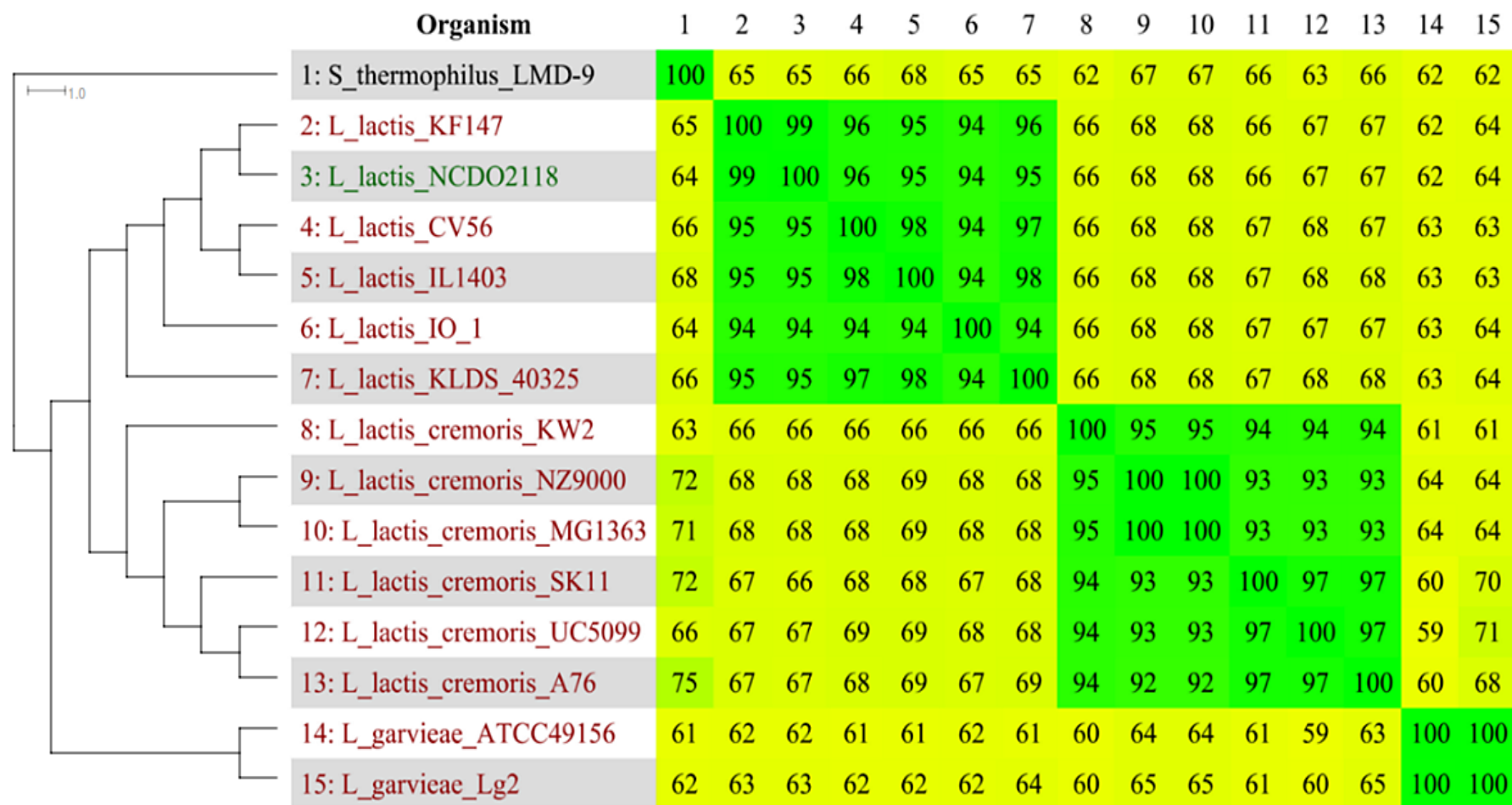
Com relação às espécies de *L. lactis cremoris*, é notada uma similaridade entre 92-100%. Vale destacar, as similaridades observadas entre *L. cremoris* A76, *L. cremoris* SK11 e *L. cremoris* UC5099, que variou de 97% a 100%.

A relação de semelhança dos dois blocos em amarelo, entre os genomas das espécies de *L. lactis* subsp. *lactis* e *L. lactis* subsp. *cremoris* é menos expressiva, apresentando uma similaridade entre 66% a 69%. Reforçando o maior grau de similaridade entre as subespécies, que no geral de 92% a 100%.

Dentre as *Lactococcus lactis* subsp. *lactis*, os resultados mais expressivos foram representados pelas linhagens isoladas de planta, no qual a similaridade de *L. lactis* NCDO 2118 com *L. lactis* KF147 foi de 99% entre elas.

Na disposição da árvore filogenômica, é constatada a formação de dois cladogramas principais, citados anteriormente como os blocos maiores, ao centro do *heatmap*. Um internó, formado abaixo da árvore é visto gerando os ramos com as linhagens de *L. garvieae*, as duas linhagens mais distintas do restante dos genomas em comparação.

Os menores graus de divergência apresentados na árvore filogenômica estão correlacionados às espécies de *Lactococcus* que apresentaram similaridade maior que 92% no *heatmap*.



**Fig. 12** Análise filogenômica do gênero *Lactococcus*. Para melhor visualização, o resultado apresentado é o conjunto dos dados obtidos no Gegenees (*heatmap*) e no SplitsTree (cladograma à esquerda). A espécie *Streptococcus thermophilus* LMD-9 (posição 1), foi adicionada para enraizar a árvore gerada. As espécies em comparação estão dispostas de 1 a 15, tanto na vertical quanto na horizontal do gráfico. Os números de 59 a 100 no gráfico mostram a porcentagem de similaridade entre as espécies, onde as cores principais: amarela e verde representam baixo e alto grau de similaridade (o tom de cada uma das três cores varia de acordo com o grau de semelhança), respectivamente.

### 5.3.2 ANÁLISE COMPARATIVA ATRAVÉS DE SINTENIA GÊNICA

Para realização das análises de sintenia gênica entre os genomas, foi utilizado o *software* Mauve. O Mauve é uma ferramenta para visualização da sintenia entre várias sequências genômicas de interesse. A comparação é baseada na homologia entre os genomas, objetivando dispor a sintenia através de blocos gênicos ajudando, dessa forma, na identificação de possíveis rearranjos entre os genomas (Darling *et al.*, 2010).

Das análises comparativas realizadas no Mauve entre o gênero *Lactococcus*, a similaridade observada entre subespécies foi bastante significativa, confirmando os dados analisados no Gegenees.

No geral, dentre as *Lactococcus lactis* subsp. *cremoris* (**Fig. 13**), foi observado muitas regiões de inverção com relação à referência (*L. lactis* subsp. *cremoris* NZ9000). Além disso, é possível notar a presença de regiões de inserção como a região em vermelho presente em *L. lactis* subsp. *cremoris* KW2, próximo à região 1200000. Neste mesmo genoma, o grau de similaridade é representado nas regiões em branco dentro dos blocos, observado na posição 800000, caracterizando a baixa similaridade. Em outras linhagens como: *L. lactis* subsp. *cremoris* SK11 e *L. lactis* subsp. *cremoris* A76, destacam-se regiões de deleção gênica vistas na posições 1700000 e 100000, respectivamente.

Quanto à comparação entre *Lactococcus garvieae* ATCC 49156 e *Lactococcus garvieae* Lg2, foi possível observar um alto grau de sintenia (**Fig. 14**). As duas linhagens compartilham de características muito semelhantes no seu genoma, reforçando a análise filogenômica vista anteriormente. Na análise sintênica, é possível observar os blocos gênicos apresentando o mesmo tamanho, mesma ordem, posições correlatas, e apenas um bloco com inversão e uma pequena região de inserção. Esta última, representada em azul na linhagem Lg2, ausente em ATCC 49156, correspondente aos genes de cápsula mencionados anteriormente.

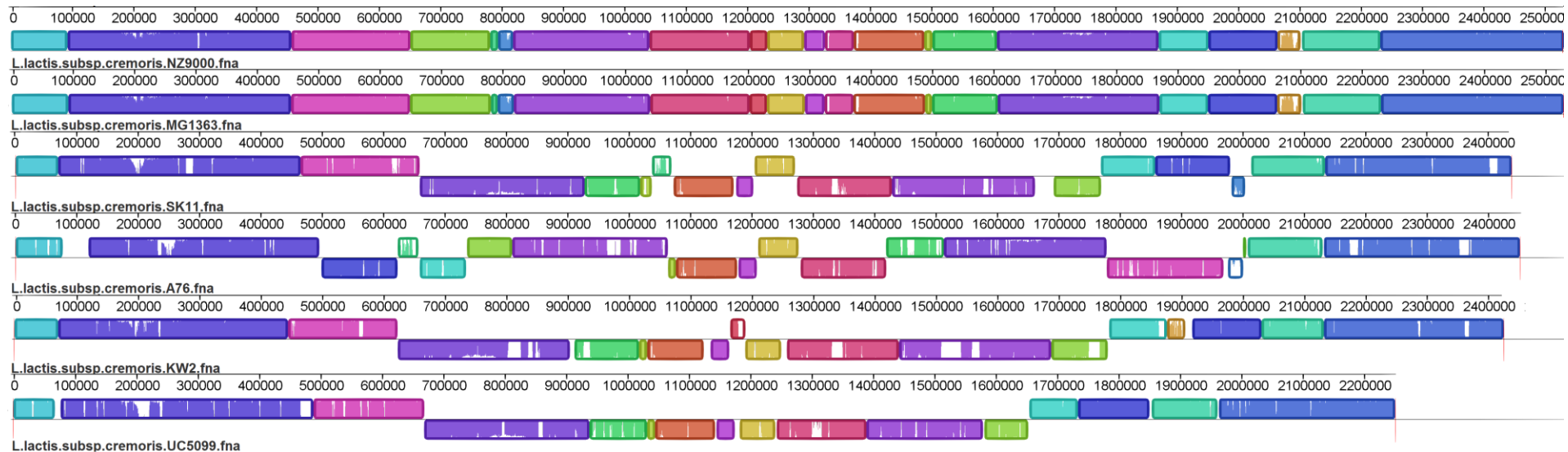
No trabalho de Roland J. Siezen e colaboradores, a montagem da *L. lactis* KF147 foi realizada com base no genoma de referência *Lactococcus. lactis* subsp. *lactis* IL1403, considerada muito sintênica, mesmo tendo sido isolada de derivados do leite (Siezen *et al.*, 2010). Na época, ainda não havia sido disponibilizado nenhum genoma de espécie de *Lactococcus* correlacionada a plantas. No entanto, a **Fig. 15** apresenta uma conservação mais significativa entre os blocos gênicos de *L. lactis* NCDO 2118 e *L. lactis* KF147 se comparado esta última à *L. lactis* IL1403. Uma de



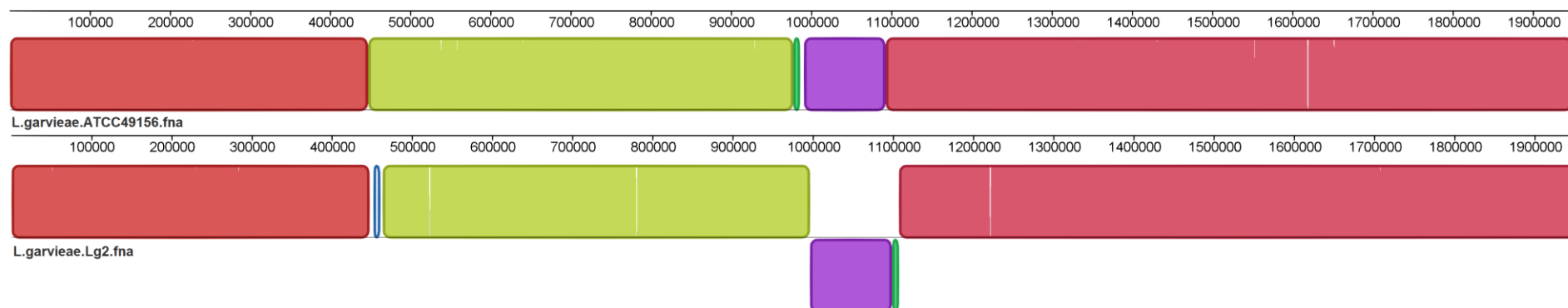
suas explicações pode estar relacionada com o seu isolamento (**Tabela 2**). Ambas foram isoladas de plantas apresentando características genômicas semelhantes como o tamanho do genoma, a quantidade de tRNA e rRNA.

Além disso, é possível verificar regiões de inserção (representadas em vermelho) entre os quatro primeiros genomas de *L. lactis* e regiões de deleção (regiões em branco entre os blocos gênicos), lembrando que as regiões em branco dentro dos blocos representam a ausência de similaridade na região.

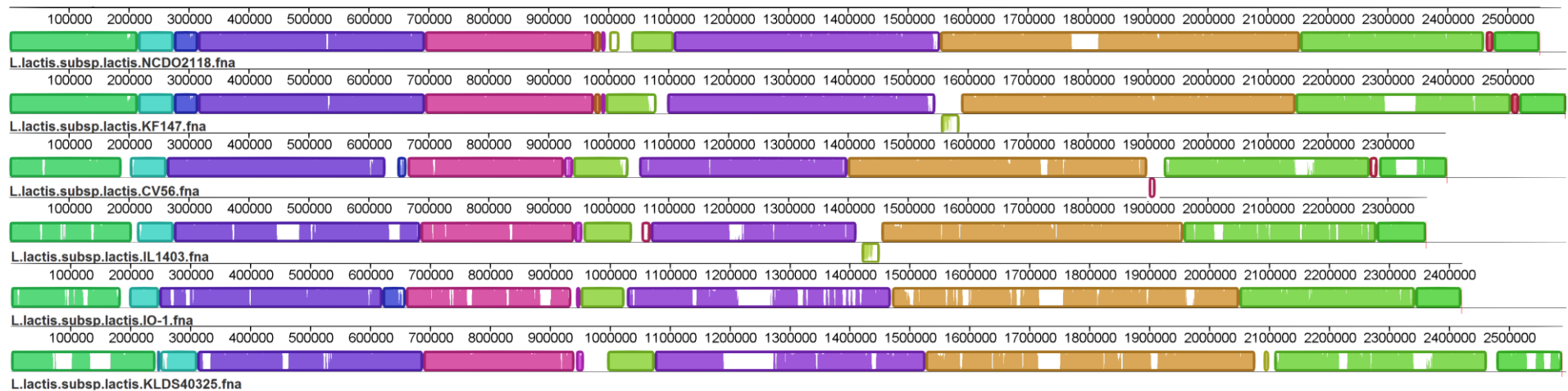
Na imagem representativa da sintonia entre os plasmídeos de *L. lactis* NCDO 2118 e *L. lactis* KF147 (**Fig. 16**), constata-se um alto grau de similaridade do início ao fim de cada sequência. Vale a pena lembrar que no processo de montagem de *L. lactis* NCDO 2118, o genoma de referência utilizado foi da linhagem *L. lactis* KF147. Como já mencionado na metodologia, praticamente metade do plasmídeo da linhagem *L. lactis* NCDO 2118 foi predito como proteínas hipotéticas, mostrando quanta informação ainda é desconhecida.



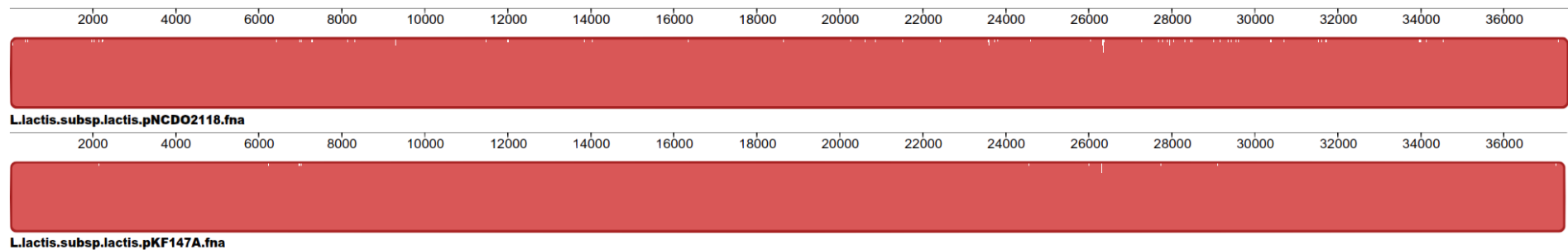
**Fig. 13.** Análise de sintenia gênica entre *Lactococcus lactis* subsp. *cremoris*. Na análise de sintenia, onde *L. lactis cremoris* NZ9000 foi selecionada como genoma de referência, observa-se um grande presença de regiões de inversão nos quatro últimos genomas, diferença no tamanho dos genomas, além da presença de alguns blocos gênicos específicos a algumas espécies (regiões de inserção).



**Fig. 14.** Análise de sintenia gênica entre *Lactococcus garvieae*. *Lactococcus garvieae* ATCC 49156 foi usada como genoma de referência. Na análise sintênica entre as linhagens, é observado um alto grau de semelhança entre os genomas, com apenas a inversão de apenas dois blocos em verde e roxo na posição de 1000000. Uma região em azul entre as posições 400000 e 500000 é presente apenas na linhagem Lg2 e ausente em ATCC49156.



**Fig. 15.** Análise de sintenia gênica entre *Lactococcus lactis* subsp. *lactis*. *L. lactis* NCDO 2118 foi usada como genoma de referência. Na representação sintênica do Mauve, a conservação expressiva é representada pelas mesmas cores preenchendo os blocos gênicos, ao contrário dos espaços em branco dentro destes, onde a similaridade é menor, ou ausente.



**Fig. 16.** Análise de sintenia gênica entre os plasmídeos de *L. lactis* NCDO 2118 e *L. lactis* KF147. É notável o grau de sintenia apresentado entre os plasmídeos das duas linhagens, os blocos gênicos são idênticos.

### 5.3.3 PREDIÇÃO E ANÁLISE DE FAGOS DE *L. lactis* NCDO 2118

O PHAST (Zhou *et al.*, 2011) é uma ferramenta *web* de anotação, identificação e visualização de sequências de profagos dentro de um genoma bacteriano, onde profagos são os fagos integrados ao genoma que não causam lise à célula hospedeira (Casjens, 2003). Sua análise é realizada por meio da integração de dados que combinam *softwares* como Glimmer (*Gene Locator and Interpolated Markov Modeler*), que faz a predição e tradução da ORF no genoma; identifica proteínas por homologia através do BLAST; reconhece sítios de ligação e mede a densidade de genes agrupados pelo DBSCAN (*Density-Based Spatical Clustering of Applications with Noise*). Além disso, o PHAST pode também, identificar tRNA, avaliar a integridade de prováveis profagos, dispor informações sobre os dados preditos em formato tabular, ou características semelhantes a estes, bem como gerar gráficos e tabelas em cores com os resultados da busca (Zhou *et al.*, 2011).

Os genomas bacterianos podem apresentar proporções maiores que 20% de genes de bacteriófagos, funcionais ou não. A presença destas regiões permite que as bactérias sejam capazes de adquirir resistência a antibióticos, a capacidade para sobreviver em novos ambientes, melhorar sua capacidade de adesão ou mesmo, tornarem-se patogênicas (Casjens, 2003). Devido à alta especificidade e ao grande potencial, os genes de fago têm sido estudados como candidatos para novos antibióticos (Coates & Hu, 2007), ou até mesmo em terapias contra o câncer (Bar *et al.*, 2008).

Na predição feita pelo PHAST, foram identificados 5 profagos (**Fig. 17**). Destes, 2 são incompletos e 3 foram preditos como intactos (**Tabela 4**). Posteriormente, as posições destas regiões serão representadas e discutidas na imagem gerada pelo BRIG, juntamente com as ilhas genômicas preditas. Os genes identificados no genoma estão listados a seguir, segundo o tipo de profago predito.

O fago intacto, correspondente a região 1, apresentou os genes *rusA* e *arsC1* preditos. O gene *rusA* está associado à sequências de profagos de vários gêneros de bactérias como: *Bacillus*, *Streptococcus*, *Staphylococcus*, *Enterococcus*, e está presente inclusive, em *Lactococcus lactis* fago r1t (Sharples *et al.*, 2002), já o gene *arsC1* está relacionado à resistência ao arsenato em *Corynebacterium glutamicum* (Mateos *et al.*, 2006). A região de fago intacto 2, apresenta os genes: *arsC3*, correlacionado ao arsenato redutase tioredoxina-dependente de *Microbacterium* sp.

A33 (Achour-Rokbani *et al.*, 2010); *amtB* é um gene da família dos transportadores de amônia encontrados em eubacteria, arqueias, fungos, plantas e animais; em procariotos, o homólogo AmtB é co-transcrito com o parálogo PII, GlnK, em resposta a limitação de nitrogênio (Blauwkamp & Ninfa, 2003); *rpmE2*, gene ligado a codificação da proteína ribossomal L31; os genes *carA*, *pyrB* e *pyrP* constituem um operon em *Lactococcus lactis* MG1363, onde *pyrP* codifica uma ligação de membrana com alta afinidade pela uracil permease, *pyrB* e *carA* codificam uma pirimidina (Martinussen *et al.*, 2001). Na região 3 do profago, o gene *pepT* é responsável pela coficação da tripeptidase.

**Tabela 4.** Genes e proteínas presentes nas regiões de fagos preditas pelo PHAST.

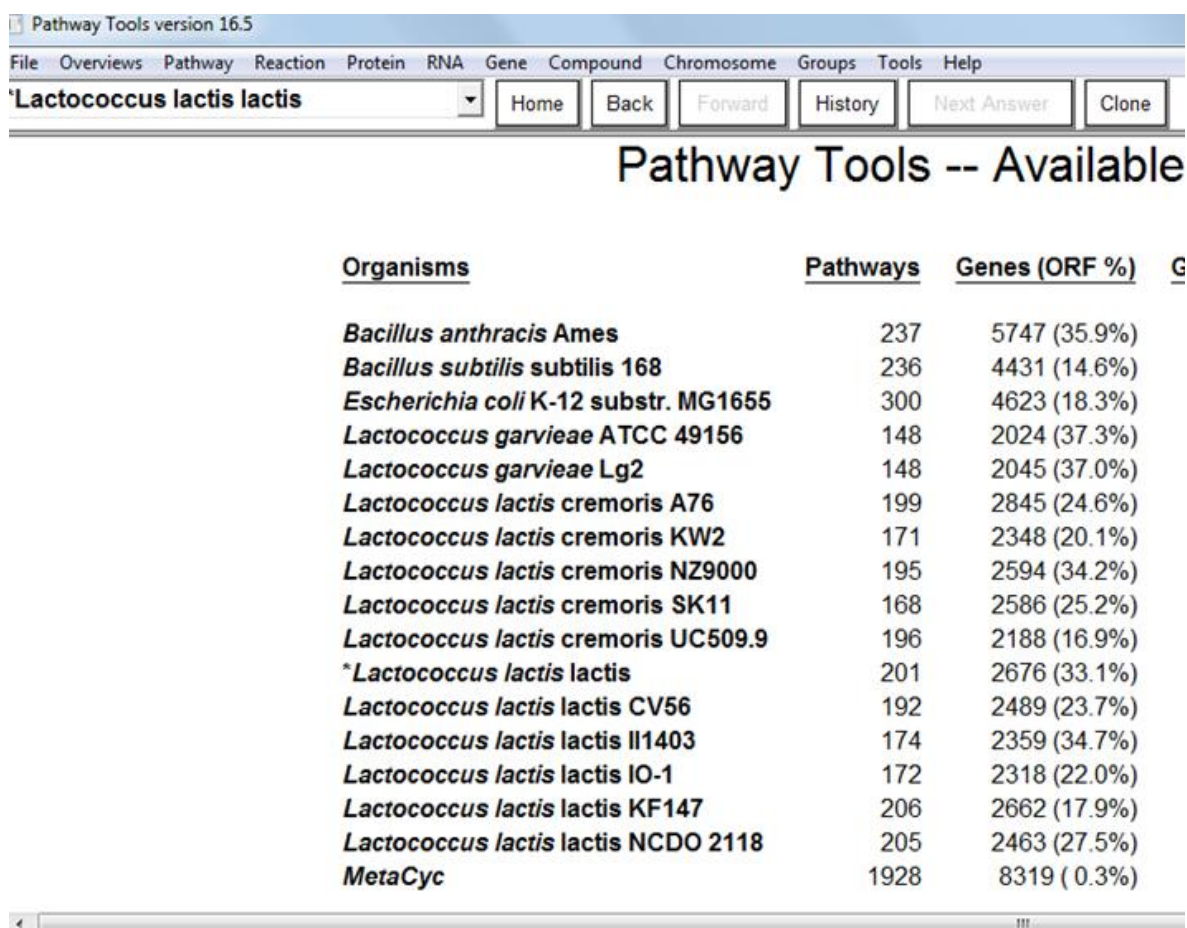
| Fagos intactos           | Genes  | Proteínas   |
|--------------------------|--|---|
| Região 1                 | <i>rusA</i> e <i>arsC1</i>                             | <i>Integrase, Prophage, Phage antirepressor, Transcriptional regulator, Recombinase, Endodeoxyribonuclease, Aminotransferase, Phage terminase small subunit, Peptidase, Bacteriophage lysine, Arsenate reductase</i>  |
| Região 2                 | <i>amtB, kinA, lra, rpmE2, arsC3, carA, pyrB, pyrP</i> | <i>Ammonium transporter, Sensor protein kinase, Two-component system regulator, 50S ribosomal protein L31 type B, Universal stress protein, Arsenate reductase, Bacteriophage lysine, Phage tail protein, Head-tail joining protein, Capsid protein, Phage ATP-dependent endopeptidase, Phage terminase small subunit, Endonuclease, Terminase, Replisome organizer, BRO-like protein, DNA binding protein, Phage integrase, Carbamoyl-phosphate synthase small chain, Aspartate carbamoyltransferase, Uracil transporter</i> |
| Região 3                 | <i>pepT, ppaC, pflA, ysiA, ysiB</i>                    | <i>Amino Acid permease, Peptidase T, Manganese-dependent inorganic pyrophosphatase, Pyruvate-formate lyase activating enzyme, Permease, Phage protein, Integrase</i>  |
| <b>Fagos incompletos</b> |  |   |
| Região 1                 | <i>ardA, ecfA1, ecfA2, ecfT, dapH e yciA</i>           | <i>Peptidoglycan hydrolase, Antirestriction protein, Integrase, ATPase, Energy-coupling factor transporter, Thiol-disulfide isomerase, N-acetyldiaminopimelate deacetylase</i>  |
| Região 2                 | <i>glnA</i>  | <i>Integrase, Bacteriocin, DNA primase, Glutamine synthetase</i>  |



### 5.3.4 PREDIÇÃO DE VIAS METABÓLICAS

Com o objetivo de identificar características metabólicas conservadas ou não, as análises foram realizadas de três formas diferentes: três linhagens filogeneticamente relacionadas de *L. lactis* subsp. *lactis*; linhagens de *L. lactis* subsp. *lactis* e *L. lactis* subsp. *cremoris* e, uma outra análise com todas as linhagens (como mencionado na metodologia).

O número de vias metabólicas abrigada por cada genoma varia de 148 a 206 (**Fig. 18**), apresentando uma média 182.7. As duas linhagens de *L. garvieae* apresentaram 148 vias metabólicas; *L. lactis* subsp. *cremoris* apresentou 185.8 e finalmente, *L. lactis* subsp. *lactis* apresentou uma média de 191.6.



The screenshot shows the Pathway Tools version 16.5 interface. The search bar contains 'Lactococcus lactis lactis'. Below the search bar, there are navigation buttons: Home, Back, Forward, History, Next Answer, and Clone. The main content area displays a table titled 'Pathway Tools -- Available' with the following data:

| Organisms                                   | Pathways | Genes (ORF %) | G |
|---|----------|---------------|---|
| <i>Bacillus anthracis</i> Ames              | 237      | 5747 (35.9%)  |   |
| <i>Bacillus subtilis</i> subtilis 168       | 236      | 4431 (14.6%)  |   |
| <i>Escherichia coli</i> K-12 substr. MG1655 | 300      | 4623 (18.3%)  |   |
| <i>Lactococcus garvieae</i> ATCC 49156      | 148      | 2024 (37.3%)  |   |
| <i>Lactococcus garvieae</i> Lg2             | 148      | 2045 (37.0%)  |   |
| <i>Lactococcus lactis</i> cremoris A76      | 199      | 2845 (24.6%)  |   |
| <i>Lactococcus lactis</i> cremoris KW2      | 171      | 2348 (20.1%)  |   |
| <i>Lactococcus lactis</i> cremoris NZ9000   | 195      | 2594 (34.2%)  |   |
| <i>Lactococcus lactis</i> cremoris SK11     | 168      | 2586 (25.2%)  |   |
| <i>Lactococcus lactis</i> cremoris UC509.9  | 196      | 2188 (16.9%)  |   |
| * <i>Lactococcus lactis</i> lactis          | 201      | 2676 (33.1%)  |   |
| <i>Lactococcus lactis</i> lactis CV56       | 192      | 2489 (23.7%)  |   |
| <i>Lactococcus lactis</i> lactis II1403     | 174      | 2359 (34.7%)  |   |
| <i>Lactococcus lactis</i> lactis IO-1       | 172      | 2318 (22.0%)  |   |
| <i>Lactococcus lactis</i> lactis KF147      | 206      | 2662 (17.9%)  |   |
| <i>Lactococcus lactis</i> lactis NCDO 2118  | 205      | 2463 (27.5%)  |   |
| MetaCyc                                     | 1928     | 8319 (0.3%)   |   |

**Fig. 18** Vias metabólicas previstas para cada espécie analisada no Pathway Tools.

Nas análises com as três linhagens filogeneticamente relacionadas de *L. lactis* subsp. *lactis* (**Fig. 19**), foram observadas vias metabólicas conservadas:

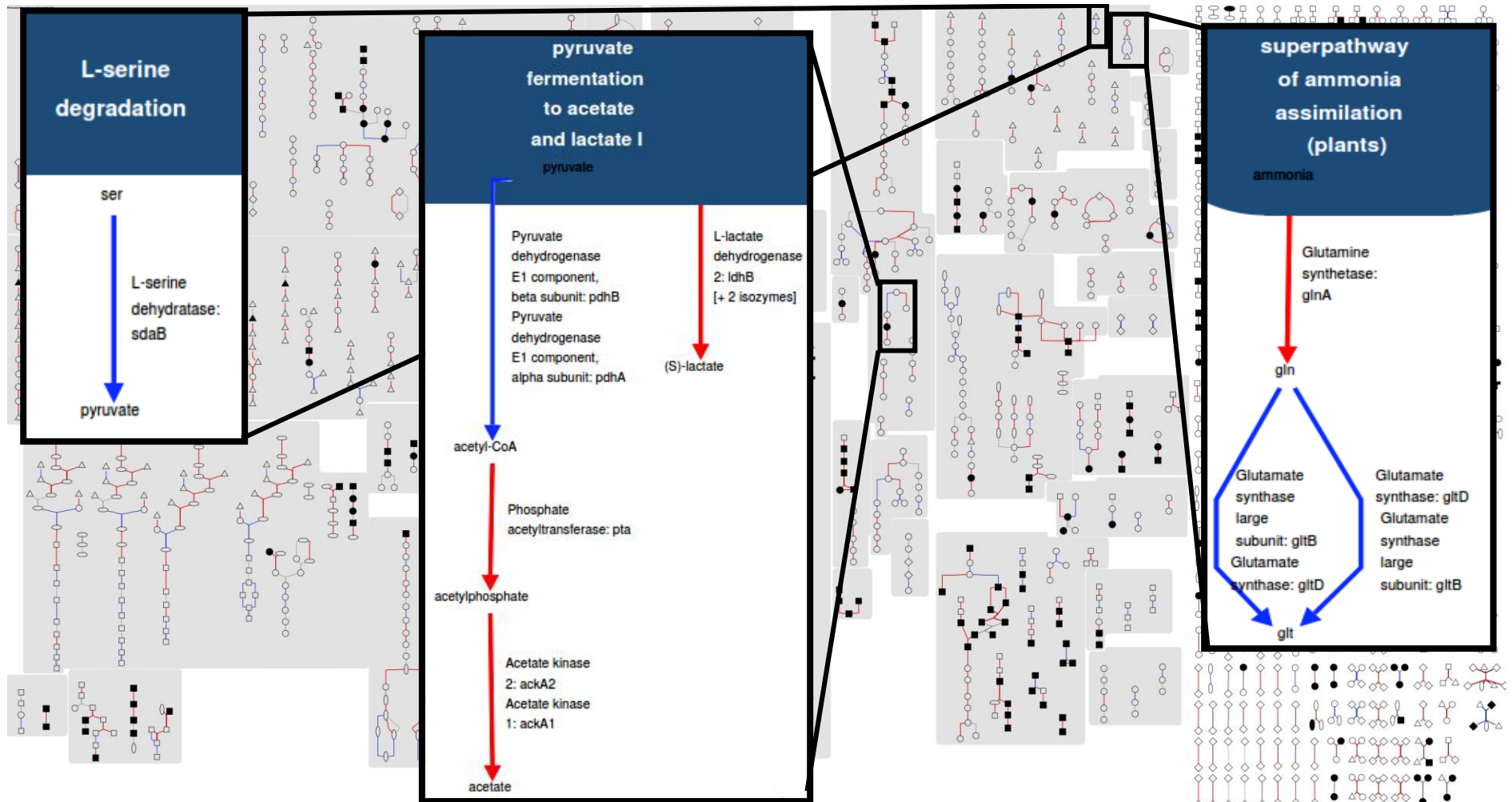
biossíntese de aminoácidos, utilização e assimilação de componentes C1, degradação de carboxilase, detoxificação, degradação de nucleosídeos e nucleotídeos, degradação de metabólitos secundários, a maior parte de degradação de carboidratos (lactose, galactose, D-manose e maltose), glicólise III, pentose fosfato, conversão de fosfoenolpiruvato à etanol, acetato e 2-oxaglutarato, fermentação de piruvato à (S)-lactato, biossíntese de (R) acetoína, síntese de coenzima A, vias alternativas I de: adenina e adenosina, guanina e guanosina, ribonucleotídeos pirimidina; assimilação de amônia e metabolismo de purina e pirimidina.

A principal diferença observada foi na linhagem *L. lactis* NCDO2118, que possui mais biossíntese de peptideoglicanos se comparada com as linhagens *L. lactis* KF147 and IL1403 (**linhas azuis, Anexo II**). Outras características exclusivas de *L. lactis* NCDO 2118 são: vias completas de respiração anaeróbia, fermentação do piruvato ao acetato, fermentação do fumarato, fermentação completa heterolática, degradação de valina e L-serina, assimilação de amônia a glutamato, supervia completa da utilização e formação do acetato, degradação de proteína, via inicial I da degradação da sucrose, degradação de valina e lisina e via acil-ACP tioesterase.



***Lactococcus lactis* subsp. *lactis* NCDO 2118**  
cellular overview

▲ Aminoácidos □ Carboidratos ◇ Proteínas ○ Purinas ○ Pirimidinas ▼ Cofatores T tRNAs ○ Outros  
● (Preenchido) Fosforilado ■ Compartilhado com *L. lactis* KF147 e *L. lactis* IL1403



**Fig. 19** Comparação de vias metabólicas entre três linhagens de *Lactococcus*. Vias metabólicas previstas no comparativo entre as linhagens de *L. lactis* NCDO 2118, *L. lactis* KF147 e *L. lactis* IL1403. Em vermelho, estão representadas as vias compartilhadas entre as três linhagens e em azul, as vias exclusivas de *L. lactis* NCDO 2118. Em zoom, destacam-se algumas das vias desta linhagem (Da esquerda para direita: degradação de L-serina, fermentação de piruvato à acetato e lactato I, supervia de assimilação de amônia (plantas)).

### 5.3.5 PREDIÇÃO E ANÁLISE DE ILHAS GENÔMICAS

Na predição de possíveis eventos de transferência horizontal de genes nos genomas de *Lactococcus lactis*, *Lactococcus cremoris* e *Lactococcus garvieae*, foi utilizado o GIPSy. Esta ferramenta é uma versão atualizada das metodologias utilizadas, até então, no PIPS (*Pathogenicity Island Prediction Software*) (Soares *et al.*, 2012). Este último segue um *pipeline*, onde informações como (i) uso de codon e (ii) desvio G+C, (iii) tRNA flanqueadores, (iv) genes de transposases e (v) fatores de virulência são utilizados para predição das ilhas de patogenicidade (PAI). Tais características são analisadas em conjunto com o genoma de uma bactéria não patogênica.

O GIPSy dispõe do mesmo *pipeline* de sua versão anterior; porém, além de prever possíveis PAI, a sua nova versão permite identificar ilhas simbióticas (SI), ilhas metabólicas (MI) e ilhas de resistência (RI).

Sendo *L. lactis* NCDO 2118 não só, uma bactéria não patogênica, como também considerada uma bactéria segura para consumo, segundo o status GRAS, as análises de ilhas de patogenicidade para este organismo foram descartadas, focando assim, na predição das outras possíveis ilhas. Para tanto, a linhagem supracitada foi comparada com *Lactococcus lactis* subsp. *cremoris* A76. Sabendo da característica virulenta de *Lactococcus garvieae*, foi acrescentado a este estudo, a predição de ilhas de patogenicidade, com o intuito de enriquecer as informações de comparação para o gênero *Lactococcus*.

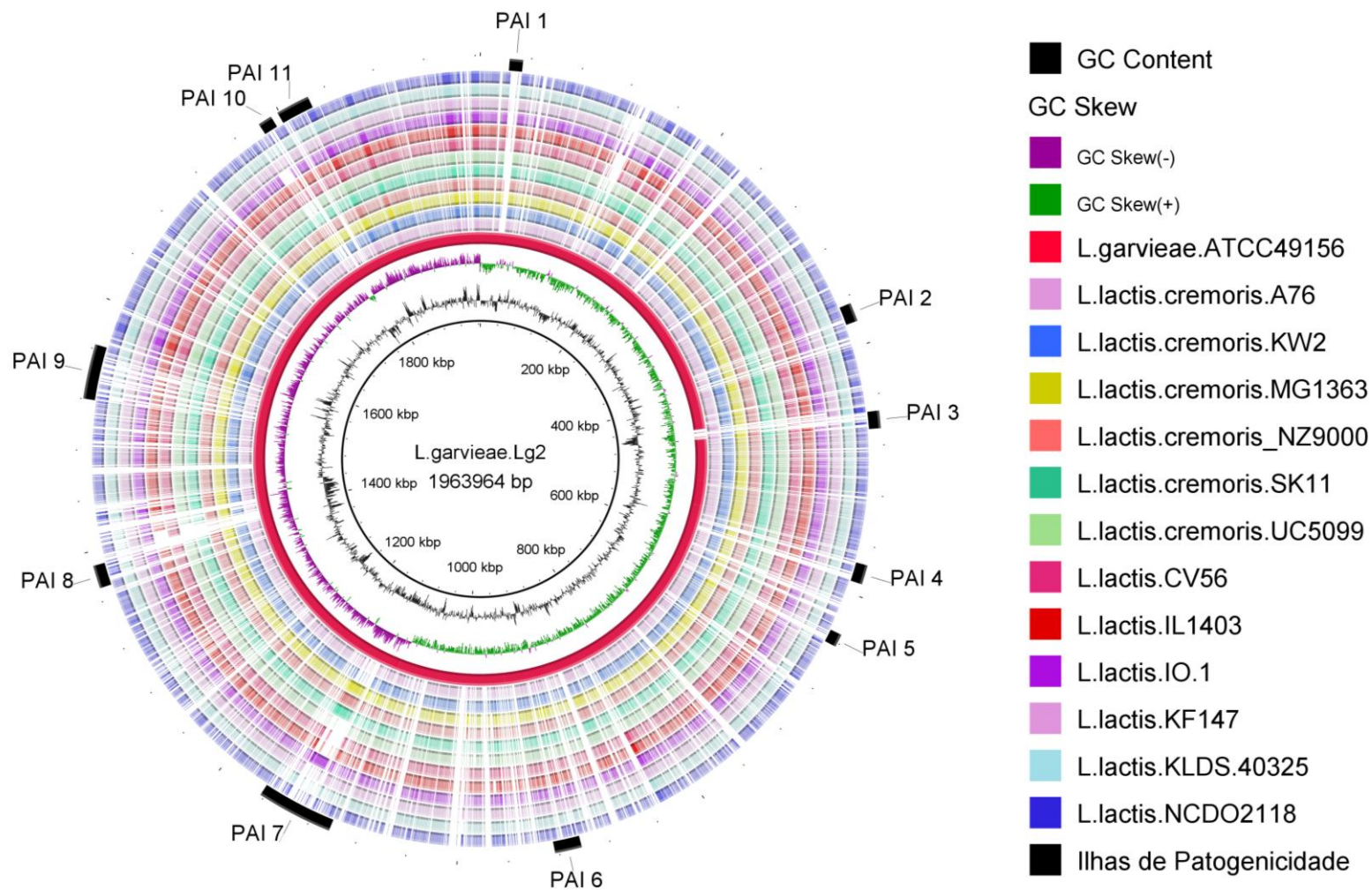
O GIPSy conta com 8 etapas para a análise de ilhas genômicas. Na primeira, etapa, são adicionados os arquivos a serem analisados (*query* e *subject*) em formatos específicos, em seguida são feitos os cálculos para desvio dos conteúdos G+C de cada genoma. Na terceira, utilizando o *software* Colombo Sigi-HMM, é feito o cálculo para o uso de desvio de codon; após a avaliação anterior, apenas o genoma de interesse (*query*), passa por uma busca de genes de transposase por meio do HMMER. Na quinta, por meio do BLAST, realiza-se uma busca por fatores específicos presentes na *query* (fatores de virulência, genes relacionados à resistência a antibióticos, metabolismo e simbiose). Em seguida, um BLASTP (aminoácidos) é iniciado, comparando as CDS dos genomas da *query* contra a *subject*. Adiante, o genoma de interesse entra em uma busca por genes de tRNA, seguindo para a última etapa de predição de ilhas genômicas baseadas nas

características identificadas em todo processo até então. Da etapa 2 a 8 é possível visualizar todos os resultados estatísticos identificados pelo GIPSY.

Para visualizar eventos de plasticidade genômica entre as linhagens, juntamente com a disposição das regiões de ilhas genômicas e fagos preditos, foi utilizado o *software* BRIG. Por meio desta ferramenta é possível a visualização circular de genomas, que facilita a identificação das diferenças no genótipo de procariotos filogeneticamente próximos. A similaridade existente entre as sequências genômicas, baseada em uma referência, é disposta na intensidade das cores dos anéis (sequência do genoma), que varia de acordo com a identidade resultante do BLAST. O BRIG usa o CGView para criar uma imagem circular comparativa dos genomas.

Na predição de ilhas genômicas, as comparações foram realizadas como feito no Mauve, entre as subespécies. As ilhas de patogenicidade (**Fig. 20**) foram identificadas na espécie *L. garvieae* Lg2 (referência). Esta última foi comparada com a *L. garvieae* ATCC 49156 e com as demais *Lactococcus*. Pela intensidade de cores observada no comparativo entre os 13 genomas contra a *L. garvieae* Lg2, observamos que apenas o primeiro anel central apresenta similaridade significativa com a referência, sendo esta, a linhagem *L. garvieae* ATCC 49156. Entre todas as *L. lactis* e *L. cremoris*, observa-se baixa similaridade, representado pelas cores claras. Algumas regiões apresentam uma ausência de genes expressiva. Na PAI 1, por exemplo, é possível identificar regiões de deleção em todas *L. lactis* e *L. cremoris*, significando a presença desta ilha genômica apenas em *L. garvieae* ATCC 49156 e *L. garvieae* Lg2. O mesmo é observado nas PAI 4, 5, 6, 10 E 11. Na PAI 7, por exemplo, existem outras espécies, entre *L. lactis* e *L. cremoris*, que possuem os mesmos genes da ilha genômica predita para *L. garvieae* Lg2.

Corroborando com os resultados apresentados no Mauve (**Fig. 14**), juntamente com os dados da literatura, foi confirmado uma região ausente em todas *L. lactis* e *L. cremoris*, além de estar ausente também em *L. garvieae* ATCC 49156, a PAI 3. Esta ilha corresponde ao *cluster* gênico presente em *L. garvieae* Lg2 e ausente em *L. garvieae* ATCC 49156, caracterizando a virulência do genoma de referência utilizado nesta análise comparativa (Morita *et al.*, 2011).



**Fig. 20** O grau de similaridade é representado pela intensidade das cores entre as sequências. Pela imagem, observa-se que a única espécie com alto grau de similaridade é o genoma do primeiro círculo de dentro para fora, relacionado à *L. garvieae* ATCC49156. A maioria das PAI são exclusivas da referência e a única ilha ausente na linhagem *L. garvieae* ATCC49156 é a PAI 3, localizada logo após a posição de 400 kb. A ausência desta ilha no primeiro genoma circular reforça a presença do *cluster* de genes específicos do genoma de referência apresentado no Mauve.

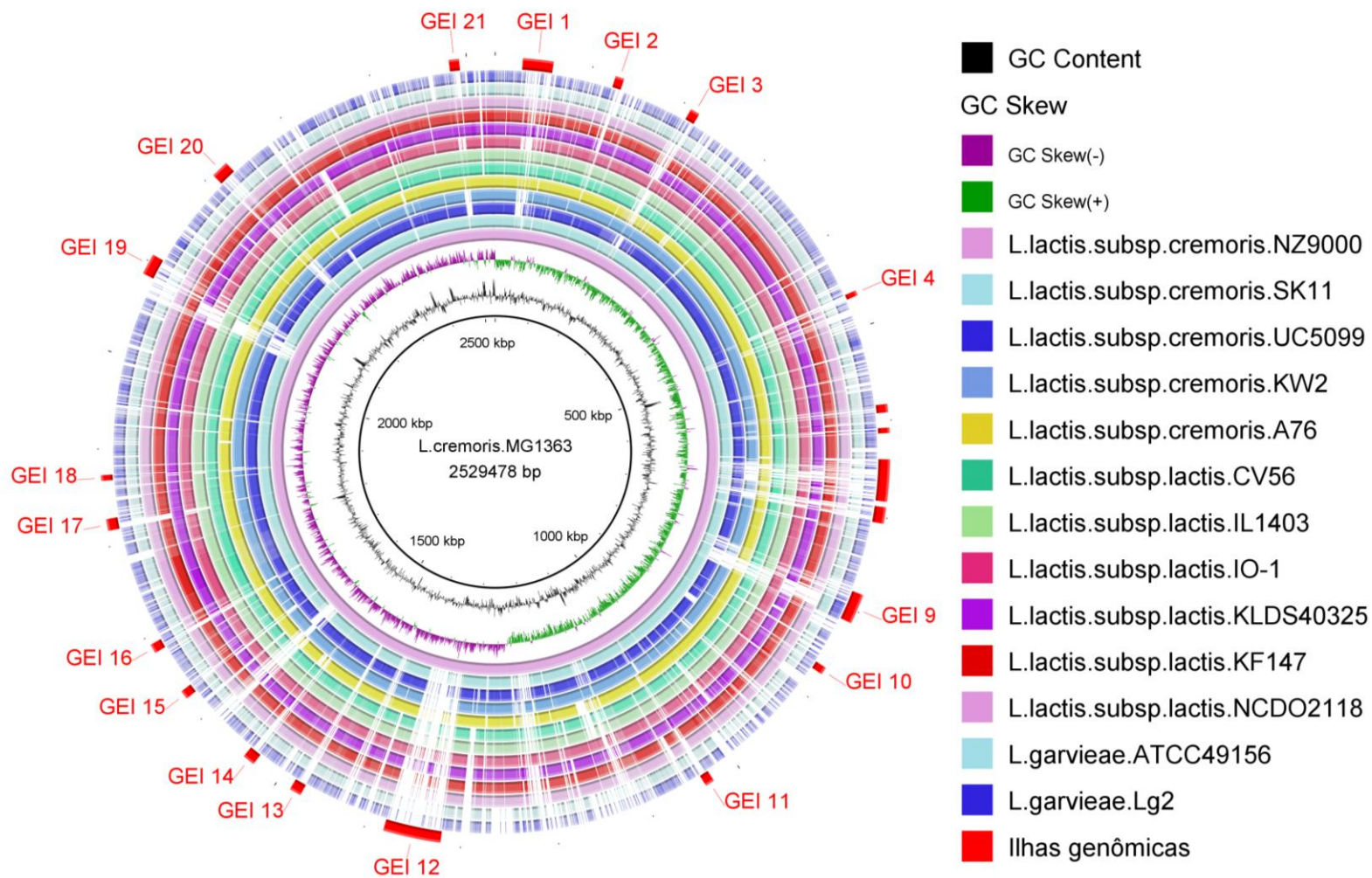
Para *L. cremoris*, foram preditas apenas ilhas genômicas, sem identificação entre metabólicas ou simbióticas (**Fig. 21**) no genoma de referência (*L. cremoris* MG1363). As regiões preditas foram observadas em linhagens específicas, como por exemplo, as GEI 1, 2, 3, 4, 9, 10, 14, 17, 18, 19, 20 e 21. Os genes da GEI 12 e 13, por exemplo, só estão presentes na referência *L. cremoris* MG1363 e em sua derivada, *L. cremoris* NZ9000, chamando a atenção para possíveis regiões que venham possuir características específicas a estas bactérias.

Em *L. lactis* NCDO 2118, foram identificadas regiões de fago, ilhas metabólicas e simbióticas (**Fig. 22**). As ilhas genômicas estão listadas conforme **Tabela 5**.

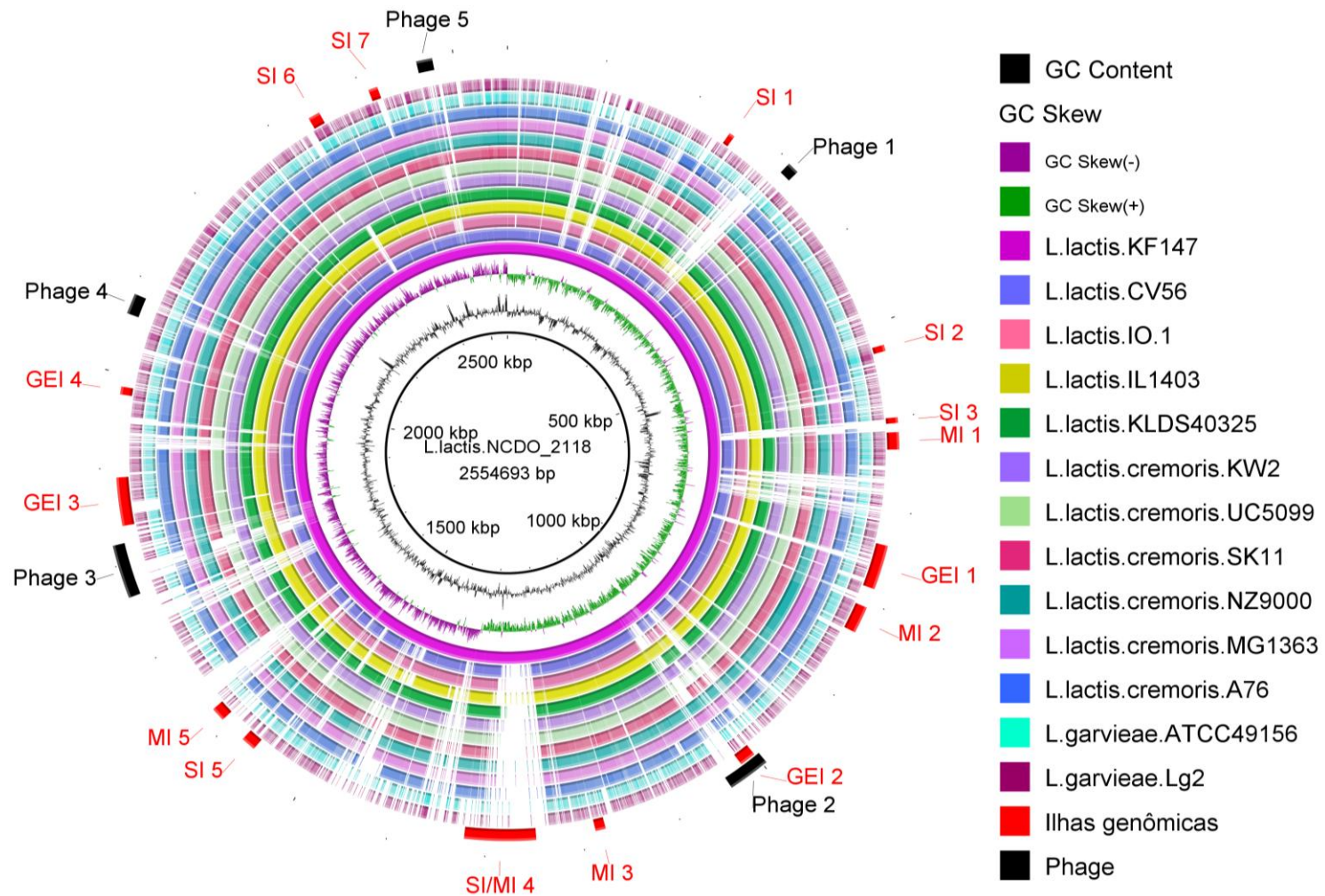
Na **figura 22**, é possível observar ilhas simbióticas que estão presentes em todas as espécies, como a SI1, SI2 e SI6 e parcialmente presentes, como a SI5. Além de ilhas metabólicas, também presentes em todos os genomas comparados, representados pelas MI1, MI2, MI3 e ilhas metabólicas, parcialmente presentes: MI5. As predições de ilhas genômicas que se destacam são: SI3, ausente em todas *Lactococcus*, com exceção das duas únicas linhagens isoladas de planta: *L. lactis* NCDO 2118 (referência) e *L. lactis* KF147. Outra ilha, identificada como metabólica ou simbiótica, está representada pelas: SI4 e MI4, esta apresenta a maior região de genes ausentes em todas *Lactococcus*, com exceção, novamente de *L. lactis* NCDO 2118 e *L. lactis* KF147.

Em relação aos fagos preditos, o fago 1, conforme identificado na **figura 22**, apresentou-se incompleto em todas as espécies, com exceção da referência e da *L. lactis* KF147. O fago 2 e 3, preditos como intactos e presentes na referência, também aparecem em *L. lactis* KF147 e *L. lactis* IL1403 e nas espécies *L. cremoris* NZ9000 e *L. cremoris* MG1363, onde o primeiro também é encontrado em *L. cremoris* KW2. O fago 4, intacto, está presente completamente em *L. lactis* KF147 e *L. lactis* IL1403, e parcialmente presente nas outras espécies. Já o fago 5, predito como incompleto, está ausente em *L. lactis* IO-1, *L. cremoris* KW2, *L. cremoris* UC5099 e parcialmente presente nas duas espécies de *L. garvieae*.





**Fig. 21.** Comparação entre as sequências do genoma de *L. cremoris* MG1363 e as outras espécies de *Lactococcus*. A similaridade é expressiva com o primeiro genoma circular de dentro para fora, relacionado à *L. cremoris* NZ9000 (sua derivada), porém o grau de similaridade de um modo geral é maior, se comparado à similaridade observada no comparativo de *L. garvieae*. As ilhas genômicas, neste caso, não foram identificadas em SI ou MI, sendo apresentadas apenas como ilhas genômicas presentes ou ausentes nos genomas de acordo com a referência.



**Fig. 22.** Comparação entre as sequências do genoma de *L. lactis* NCDO 2118 e as outras espécies de *Lactococcus*. O grau de similaridade é expressivo com o primeiro genoma circular de dentro para fora, relacionado à *L. lactis* KF147 e menos significativo com os dois últimos genomas dos anéis mais externos, representados pelas *L. garvieae*. As ilhas genômicas que foram possíveis ser preditas, foram divididas em MI e SI, onde foi identificado uma mesma ilha tanto pra ilha metabólica, quanto para a ilha simbiótica (SI/MI4). Além da predição de ilhas, foram acrescentadas as predições das 5 regiões de fago (em preto, na parte exterior ao genoma circular).

**Tabela 5.** Ilhas simbióticas e metabólicas preditas para *L. lactis* NCDO 2118:

| <b>Ilhas Simbióticas</b> | <b>Locus tag e genes</b> | <b>Produtos</b>                                   |
|--------------------------|--------------------------|---|
| <b>SI 1</b>              | pepD                     | Dipeptidase                                       |
|                          | NCDO2118_0248            | Transcriptional regulator                         |
|                          | NCDO2118_0249            | ABC transporter ATP binding protein               |
|                          | NCDO2118_0250            | ABC transporter permease protein                  |
|                          | NCDO2118_0251            | SAM-dependent methyltransferase                   |
|                          | NCDO2118_0252            | Fructose-1,6-bisphosphatase class 3               |
| <b>SI 2</b>              | recU                     | Holliday junction resolvase                       |
|                          | NCDO2118_0528            | Hypothetical protein                              |
|                          | spxA1                    | Regulatory protein Spx 1                          |
|                          | NCDO2118_0530            | Hypothetical protein                              |
|                          | NCDO2118_0531            | Inositol-1-monophosphatase                        |
|                          | murA2                    | UDP-N-acetylglucosamine1carboxyvinyltransferase   |
|                          | NCDO2118_0533            | Hypothetical protein                              |
|                          | tig                      | Trigger factor                                    |
| <b>SI 3</b>              | gnd                      | 6-phosphogluconate dehydrogenase, decarboxylating |
|                          | NCDO2118_0606            | DNA methylase                                     |
|                          | NCDO2118_0607            | Hypothetical protein                              |
|                          | NCDO2118_0608            | Hypothetical protein                              |
|                          | NCDO2118_0609            | DNA helicase                                      |
| <b>SI 4</b>              | nrpA                     | Non-ribosomal peptide sythetase                   |
|                          | pksA                     | Acyl-CoA transferase                              |
|                          | pksB                     | Biosynthesis protein                              |
|                          | pksC                     | Biosynthesis protein                              |
|                          | NCDO2118_1191            | ABC transporter                                   |
|                          | NCDO2118_1192            | ABC transporter                                   |
|                          | NCDO2118_1193            | Hypothetical protein                              |
|                          | npkS                     | Hybrid nonribosomal peptide                       |
|                          | nrpB                     | Thioesterase                                      |
|                          | nrpC                     | 4'-phosphopantetheinyl transferase                |



|  |               |   |
|--|---------------|---|
|  | NCDO2118_1197 | Hypothetical protein  |
|  | NCDO2118_1198 | Macrolide biosynthetic protein  |
|  | NCDO2118_1199 | FAD dependent oxidoreductase  |
|  | NCDO2118_1200 | Hypothetical protein  |
|  | NCDO2118_1201 | Hypothetical protein  |
|  | NCDO2118_1202 | Hypothetical protein  |
|  | NCDO2118_1204 | Hypothetical protein  |
|  | NCDO2118_1205 | Hypothetical protein  |
|  | NCDO2118_1206 | Hypothetical protein  |
|  | kdpD          | Osmosensitive K <sup>+</sup> channel sensor histidine kinase            |
|  | kdpE          | Osmosensitive K <sup>+</sup> channel response regulator                 |
|  | kdpA          | Potassium-transporting ATPase A chain                                   |
|  | kdpB          | Potassium-transporting ATPase B chain                                   |
|  | kdpC          | Potassium-transporting ATPase C chain                                   |
|  | NCDO2118_1212 | Transcriptional regulator   |
|  | NCDO2118_1213 | Hypothetical protein  |
|  | rnr           | Ribonuclease R  |
|  | hisC          | Histidinol-phosphate aminotransferase                                   |
|  | hisZ          | ATP phosphoribosyltransferase regulatory subunit                        |
|  | hisG          | ATP phosphoribosyltransferase   |
|  | hisD          | Histidinol dehydrogenase  |
|  | NCDO2118_1219 | SAM-dependent methyltransferase   |
|  | hisB          | Imidazoleglycerol-phosphate dehydratase                                 |
|  | ymdC          | Aminoglycoside 3-phosphotransferase                                     |
|  | hisH          | Imidazole glycerol phosphate synthase                                   |
|  | hisA          | Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase |
|  | hisF          | Imidazole glycerol phosphate synthase                                   |
|  | hisI          | Histidine biosynthesis bifunctional protein                             |
|  | hisK          | Histidinol-phosphatase  |
|  | NCDO2118_1227 | Hypothetical protein  |
|  | leuA          | 2-isopropylmalate synthase  |

|             |               |  |
|-------------|---------------|--|
|             | leuB          | 3-isopropylmalate dehydrogenase                                |
|             | leuC          | 3-isopropylmalate dehydratase large subunit                    |
| <b>SI 5</b> | coaA          | Pantothenate kinase  |
|             | NCDO2118_1453 | Phosphate-starvation-inducible protein                         |
|             | NCDO2118_1454 | Cation transporter   |
|             | NCDO2118_1455 | Transcriptional regulator                                      |
|             | NCDO2118_1456 | Chromosome partitioning protein ParB                           |
|             | yohC          | Transcriptional regulator                                      |
|             | NCDO2118_1459 | Hypothetical protein   |
|             | opuABC        | Betaine ABC transporter permease and substrate binding protein |
|             | opuAA         | Glycine betaine transport ATP-binding protein                  |
|             | busR          | Transcriptional regulator                                      |
|             | gtf1          | Glycosyltransferase Gtf1                                       |
|             | NCDO2118_1464 | Lipopolysaccharide biosynthesis protein                        |
|             | bglH          | Beta-glucosidase   |
| <b>SI 6</b> | rpsM          | 30S ribosomal protein S13                                      |
|             | infA          | Translation initiation factor IF-1                             |
|             | adk           | Adenylate kinase   |
|             | secY          | Protein translocase subunit                                    |
|             | rplO          | 50S ribosomal protein L15                                      |
|             | rpmD          | 50S ribosomal protein L30                                      |
|             | rpsE          | 30S ribosomal protein S5                                       |
|             | rplR          | 50S ribosomal protein L18                                      |
|             | rplF          | 50S ribosomal protein L6                                       |
|             | rpsH          | 30S ribosomal protein S8                                       |
|             | NCDO2118_2195 | Hypothetical protein   |
|             | rpsZ          | 30S ribosomal protein S14 type Z                               |
|             | rplE          | 50S ribosomal protein L5                                       |
|             | rplX          | 50S ribosomal protein L24                                      |
|             | rplN          | 50S ribosomal protein L14                                      |
|             | rpsQ          | 30S ribosomal protein S17                                      |

|             |               |                           |
|-------------|---------------|---------------------------|
|             | rpmC          | 50S ribosomal protein L29 |
|             | rplP          | 50S ribosomal protein L16 |
|             | rpsC          | 30S ribosomal protein S3  |
|             | rplV          | 50S ribosomal protein L22 |
|             | rpsS          | 30S ribosomal protein S19 |
|             | rplB          | 50S ribosomal protein L2  |
|             | rplW          | 50S ribosomal protein L23 |
|             | rplD          | 50S ribosomal protein L4  |
| <b>SI 7</b> | rpsB          | 30S ribosomal protein S2  |
|             | NCDO2118_2256 | Hypothetical protein      |
|             | adhE          | Alcohol dehydrogenase     |
|             | NCDO2118_2258 | Hypothetical protein      |
|             | NCDO2118_2259 | Hypothetical protein      |
|             | ywdC          | Fe-S oxidoreductase       |
|             | ywdD          | Fe-S oxidoreductase       |
|             | NCDO2118_2262 | Transcriptional regulator |
|             | NCDO2118_2263 | Hypothetical protein      |

**Tabela 6.** Ilhas metabólicas preditas para *L. lactis* NCDO 2118:

| <b>Ilhas Metabólicas</b> | <b>Locus tag e genes</b> | <b>Produtos</b>                          |
|--------------------------|--------------------------|--|
| <b>MI 1</b>              | kup1                     | Potassium transport system protein kup 1 |
|                          | kup2                     | Potassium transport system protein kup 2 |
|                          | NCDO2118_0621            | Hypothetical protein                     |
|                          | NCDO2118_0622            | Hypothetical protein                     |
|                          | miaA                     | tRNA dimethylallyltransferase            |
|                          | NCDO2118_0624            | Hypothetical protein                     |
|                          | NCDO2118_0625            | Hypothetical protein                     |
|                          | NCDO2118_0626            | Hypothetical protein                     |
|                          | rnz                      | Ribonuclease Z                           |
|                          | NCDO2118_0628            | Oxidoreductase                           |

|             |               |   |
|-------------|---------------|---|
|             | recJ          | Single-stranded-DNA-specific exonuclease                        |
|             | apt           | Adenine phosphoribosyltransferase                               |
|             | rpoE          | DNA-directed RNA polymerase subunit delta                       |
|             | NCDO2118_0632 | Aminodeoxychorismate lyase                                      |
|             | greA          | Transcription elongation factor                                 |
|             | ctsR          | Transcriptional regulator                                       |
|             | clpC          | ATP-dependent Clp protease, ATP-binding subunit                 |
| <b>MI 2</b> | rimP          | Ribosome maturation factor                                      |
|             | nusA          | Transcription termination/antitermination protein               |
|             | NCDO2118_0781 | Hypothetical protein  |
|             | NCDO2118_0782 | Hypotetical protein   |
|             | infB          | Translation initiation factor IF-2                              |
|             | rbfA          | Ribosome-binding factor A                                       |
|             | NCDO2118_0785 | Mannose-6-phosphate isomerase                                   |
|             | rmaG          | Transcriptional regulator                                       |
|             | fabH          | 3-oxoacyl-acyl-carrier-protein synthase 3                       |
|             | acpP          | Acyl carrier protein  |
|             | fabD          | Malonyl CoA-acyl carrier protein transacylase                   |
|             | NCDO2118_0790 | 3-Oxoacyl-acyl-carrier protein reductase                        |
|             | fabF          | 3-oxoacyl-acyl-carrier-protein synthase 2                       |
|             | accB          | Acetyl-CoA carboxylase, biotin carboxyl carrier protein         |
|             | fabZ2         | 3-hydroxyacyl-acyl-carrier-protein dehydratase                  |
|             | accC          | Acetyl-CoA carboxylase, biotin carboxylase                      |
|             | accD          | Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta |
|             | accA          | Acetyl-CoA carboxylase, carboxyl transferase subunit alpha      |
|             | metC          | Cystathionine beta-lyase  |
|             | cysK2         | Cysteine synthase   |
|             | NCDO2118_0799 | Hypothetical protein  |
|             | NCDO2118_0800 | Hypothetical protein  |
|             | noxC          | NADH dehydrogenase  |
|             | NCDO2118_0802 | Hypothetical protein  |

|             |   |  |
|-------------|---|--|
|             | NCDO2118_0803   | Arsenate reductase                                 |
|             | rdrA  | Transcriptional regulator                          |
|             | NCDO2118_0805   | Acetyltransferase                                  |
|             | exoA  | Exodeoxyribonuclease                               |
|             | metG  | Methionine tRNA ligase                             |
|             | NCDO2118_0808   | Hipotetical Protein                                |
|             | NCDO2118_0809   | Short chain dehydrogenase                          |
| <b>MI 3</b> | NCDO2118_1126   | Hypothetical protein                               |
|             | truB  | tRNA pseudouridine synthase B                      |
|             | ribC  | Riboflavin kinase                                  |
|             | ldhX  | L-lactate dehydrogenase                            |
|             | NCDO2118_1130   | Hypothetical protein                               |
|             | murQ  | N-acetylmuramic acid 6-phosphate etherase          |
|             | NCDO2118_1132   | PTS system, sucrose-specific                       |
|             | NCDO2118_1133   | Transcriptional regulator                          |
|             | tpiA  | Triosephosphate isomerase                          |
|             | pacA  | Penicillin acylase                                 |
|             | NCDO2118_1136   | Hypothetical protein                               |
| <b>MI 4</b> | <b>(Os dados estão na tabela de ilhas simbióticas – mesma região predita)</b> |  |
| <b>MI 5</b> | NCDO2118_1495   | Hypothetical protein                               |
|             | NCDO2118_1497   | Hypothetical protein                               |
|             | NCDO2118_1498   | Sugar ABC transporter substrate-binding protein    |
|             | NCDO2118_1499   | Oligosaccharider ABC transporter, permease protein |
|             | NCDO2118_1500   | Sugar ABC transporter permease protein             |
|             | guaA  | GMP synthase - glutamine-hydrolyzing               |
|             | fruK  | Fructokinase                                       |
|             | glcK  | Glucokinase  |
|             | NCDO2118_1504   | Beta-glucosidase                                   |
|             | NCDO2118_1505   | Hypothetical protein                               |
|             | NCDO2118_1506   | Endo-beta-N-acetylglucosaminidase                  |
|             | dexB  | Alpha 1-6-glucosidase                              |

## 5.4 ANÁLISES PROBIÓTICAS DE *L.lactis* NCDO 2118

Nas análises a seguir, daremos atenção aos três pontos principais que caracterizam um organismo como sendo probiótico: (i) a sobrevivência através do trânsito GIT e adesão às células do epitélio intestinal; (ii) a atividade antimicrobiana e a exclusão por competitividade; e, (iii) a modulação do sistema imune do GIT do hospedeiro (Johnson & Klaenhammer, 2014). Cada um deles será descrito nas próximas seções.

### 5.4.1 IDENTIFICAÇÃO DE GENES RELACIONADOS À RESISTÊNCIA AO ESTRESSE ÁCIDO E À ADESÃO

O número de mecanismos de resistência ao estresse em *L. lactis* vem sendo identificado e caracterizado. O primeiro mecanismo para a sobrevivência da espécie está no controle do baixo pH intracelular (pHi) pela ATPase de membrana  $F_0F_1$ , a qual transloca prótons para o ambiente às custas de ATP (Even *et al.*, 2003; O'Sullivan & Condon, 1999). Outros mecanismos incluem a geração de substâncias alcalinas por meio do catabolismo de aminoácidos (por exemplo: deaminação) (Christensen *et al.*, 1999; Sanders *et al.*, 1999). *L. lactis* também desenvolve uma resposta adaptativa complexa ao estresse ácido, onde este é dependente da síntese de proteínas como a proteína de choque térmico e proteinases (Even *et al.*, 2003).

Pela **Tabela 7**, é possível observar os genes codificantes de proteínas envolvidas na resistência ao estresse ácido de *L. lactis* NCDO 2118.

Segundo o trabalho de Carvalho e colaboradores, em 2003, vários genes relacionados ao estresse foram expressos de forma diferente em células cultivadas em baixo e ótimo pH (5,1 e 6,5, respectivamente), como os genes responsáveis pela codificação de chaperonas: *dnaK* e *groEL* e o gene de resposta estrigente, *relA* (Carvalho *et al.*, 2013). Estes mesmos genes foram identificados em *L. lactis* NCDO 2118.

**Tabela 7.** Genes codificantes de proteínas envolvidas na resistência ao estresse ácido:

| Locus_tag     | Gene | Produto                    |
|---------------|------|----------------------------|
| NCDO2118_1870 | atpC | ATP synthase epsilon chain |
| NCDO2118_1871 | atpD | ATP synthase subunit beta  |
| NCDO2118_1872 | atpG | ATP synthase gamma chain   |
| NCDO2118_1873 | atpA | ATP synthase subunit alpha |

|               |       |   |
|---------------|-------|---|
| NCDO2118_1874 | atpH  | ATP synthase subunit delta**                              |
| NCDO2118_1875 | atpF  | ATP synthase subunit b                                    |
| NCDO2118_1876 | atpB  | ATP synthase subunit a                                    |
| NCDO2118_1877 | atpE  | ATP synthase subunit C                                    |
| NCDO2118_1384 | ldh   | L-lactate dehydrogenase                                   |
| NCDO2118_0475 | ptcC  | PTS system, cellobiose-specific IIC component             |
| NCDO2118_0542 | gapA  | Glyceraldehyde-3-phosphate dehydrogenase                  |
| NCDO2118_0637 | eno   | Enolase   |
| NCDO2118_0399 | gpmA  | 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase |
| NCDO2118_2272 | pgi   | Glucose-6-phosphate isomerase                             |
| NCDO2118_0096 | pyk1  | Pyruvate kinase   |
| NCDO2118_1385 | pyk2  | Pyruvate kinase**   |
| NCDO2118_0240 | pgk   | Phosphoglycerate kinase                                   |
| NCDO2118_0417 | recA1 | Protein RecA  |
| NCDO2118_1251 | recA2 | Protein RecA  |
| NCDO2118_1019 | dnaK  | Chaperone protein   |
| NCDO2118_0540 | clpE  | ATP-dependent Clp protease ATP-binding subunit            |
| NCDO2118_0453 | groL  | 60 kDa chaperonin   |
| NCDO2118_1545 | clpB  | Chaperone protein   |
| NCDO2118_0467 | sodA  | Superoxide dismutase                                      |
| NCDO2118_0073 | relA  | GTP pyrophosphokinase                                     |

\*\* genes associados na resposta ao estresse ácido que não foram identificados em *L. lactis* subsp. *lactis* MG1363, segundo trabalho de Carvalho e colaboradores, 2003.

Com relação às proteínas de adesão preditas em *L. lactis* NCDO 2118, merece destaque o gene *chiA* e a proteína codificada *Chitin binding protein*. A quitina é degradada por quitinases que pertencem aos membros da família 18 de glicosídeo hidrolase (Kawada *et al.*, 2008).

Um exemplo de bactéria que produz quitinases é a *Serratia marcescens*, um dos organismos mais eficientes em degradar quitina (Kamba *et al.*, 2013). Quando *E. coli* foi substituída com uma proteína que se liga à quitina (CBP *chitin-binding protein*) de *Serratia marcescens*, houve um aumento significativo na habilidade da *E. coli* em aderir às células do cólon humano (Kawada *et al.*, 2008). Durante anos, a quitina tem sido usada de forma segura e efetiva como ingrediente anti-inflamatório na medicina chinesa (Zhou *et al.*, 2005). Os produtos suplementares da dieta de glucosamina (sulfato de glucosamina, N-acetil-glucosamina, dentre outros), do qual a quitina é derivada, aparecem como substâncias anti-inflamatórias quem têm sido utilizadas em doenças cardiovasculares, artrite e na prevenção do câncer (Fouladbakhsh, 2012; Greenlee *et al.*, 2009; Yeh *et al.*, 2006).

Genes que codificam a família 33 CBP podem ocorrer em bactérias que contém, de algum modo, maquinaria quitinolítica, como a que está presente em *L. lactis* subsp. *lactis* IL1403 (Vaaje-Kolstad *et al.*, 2009). A sequência do genoma de *L.*

*Lactis* IL1403 mostra três genes que estão potencialmente envolvidos no *turnover* de quitina: uma família 18 quitinase (*chiA*; proteína referida como LIChi18A), uma família 33 CBP (*yucG*; referida como LICBP33A) e a família 20 *N*-acetilhexosaminidase (*LnbA*) (Bolotin *et al.*, 2001).

Um organismo que mostra grande afinidade por componentes da matriz extracelular (ECM) é *Borrelia burgdorferi*, uma bactéria extracelular associada com o processo de ligação a tecidos do hospedeiro (Cabello *et al.*, 2007; Cadavid *et al.*, 2003), ela apresenta afinidades ao colágeno do tipo I, fibronectina e decorina. Verma e colaboradores, em 2009, determinaram que este mesmo organismo é capaz de se ligar à laminina de mamíferos, um componente importante no ECM (Verma *et al.*, 2009).

Com base nos dados da literatura acima descritos, observou-se as predições das proteínas que podem estar envolvidas nos mecanismos de adesão de *L. lactis* NCDO 2118, na **Tabela 8**, o gene *chiA* e a proteína codificante da *locus tag* NCDO2118\_2054 (*Chitin binding protein*), também presentes em *L. lactis* IL1403.

**Tabela 8** - Proteínas potencialmente envolvidas nos mecanismos de adesão de *L. lactis* ao hospedeiro:

| Locus_tag     | Gene | Produto  |
|---------------|------|--|
| NCDO2118_0315 |      | Hypothetical protein   |
| NCDO2118_0552 |      | Hypothetical protein   |
| NCDO2118_0647 | pycA | Pyruvate carboxylase   |
| NCDO2118_0684 |      | ChW repeat-/cell adhesion domain-containing transglutaminase-like protease |
| NCDO2118_0727 |      | Hypothetical protein   |
| NCDO2118_0774 |      | Flagellar hook-length control protein FliK                                 |
| NCDO2118_0776 |      | Hypothetical protein   |
| NCDO2118_0806 | exoA | Exodeoxyribonuclease   |
| NCDO2118_0857 |      | Hipotetical protein  |
| NCDO2118_1205 |      | Hypothetical protein   |
| NCDO2118_1365 |      | Hypothetical protein   |
| NCDO2118_1446 | bmpA | Basic membrane protein A (laminin-binding protein)                         |
| NCDO2118_1515 | ypdD | Sugar hydrolase  |
| NCDO2118_1627 |      | Hypothetical protein   |
| NCDO2118_2053 | chiA | Chitinase  |
| NCDO2118_2054 |      | Chitin binding protein   |
| NCDO2118_2211 |      | Hypothetical protein   |
| NCDO2118_2278 |      | Fibronectin-binding protein  |
| NCDO2118_2284 |      | Hypothetical protein   |



#### 5.4.2 PREDIÇÃO E ANÁLISE DE BACTERIOCINAS

Na identificação de prováveis bacteriocinas, utilizamos o BAGEL (de Jong *et al.*, 2010), uma ferramenta web capaz de identificá-las através de domínios conservados, presença de biossíntese, transporte e genes de imunidade e propriedades físicas no contexto genômico.

Além da identificação, o BAGEL classifica as bacteriocinas em três classes: (i) bacteriocinas que contém lantioninas (ii) bacteriocinas que não possuem lantioninas e (iii) as bacteriolisinas (Cotter *et al.*, 2005); não há classificação para a classe (iv) previamente mencionada no início deste trabalho.

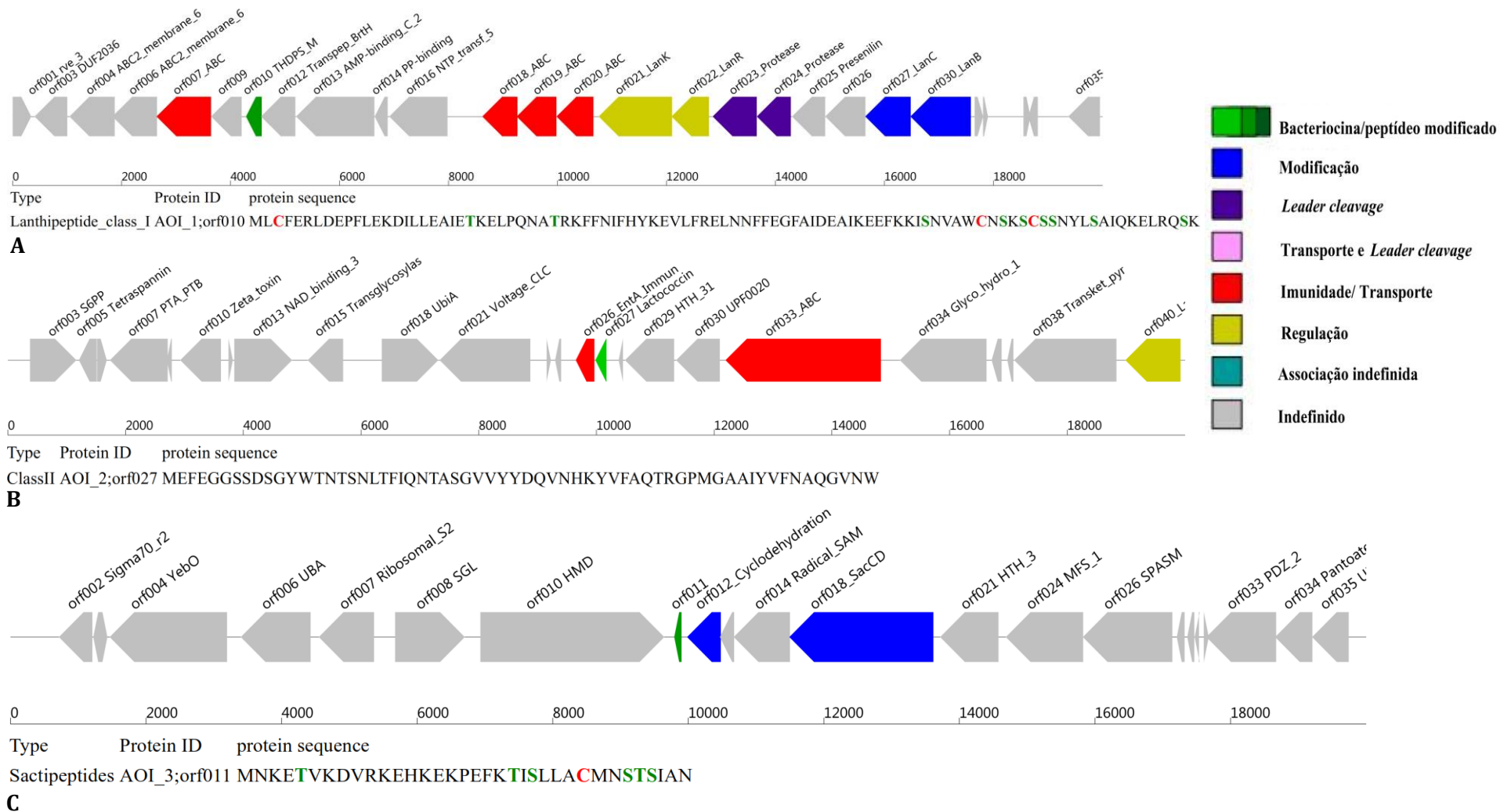
O *software* trabalha com um banco de dados curado de bacteriocinas, sendo que os dados são avaliados com base no Modelo Oculto de Markov (*Hidden Markov Model*). As informações genéticas são analisadas com base em combinações de domínios do PFAM (*Protein Families Database*).

Foram preditas para *L. lactis* NCDO 2118 (**Fig. 23**), através do BAGEL, uma bacteriocina correspondente a cada uma das três classes que a ferramenta disponibiliza. Uma bacteriocina classe I: Lanthipeptide, uma de classe II e outra da classe III, Sacceptidase. Utilizando as posições de cada bacteriocina, buscou-se no genoma de *L. lactis* NCDO 2118 a localização de cada um delas. As bacteriocinas de classe I e III, mesmo tendo sido preditas, não foram localizadas no genoma de *L. lactis* NCDO 2118, sendo possivelmente regiões que não foram preditas durante a anotação manual, porém foi observado que a Sacceptidase/Classe III se encontrava na mesma região em que foi predita uma ilha simbiótica em *L. lactis* NCDO 2118, a ilha SI7. Os genes *acmA*, *acmB*, *acmC* e *acmD* e a proteína *Lysozyme M1*, foram adicionadas à tabela após busca manual no genoma de *L. lactis* NCDO 2118.

*Lactococcus lactis* subsp. *cremoris* MG1363 possui uma autolisina, *AcmA*, necessária para separação celular durante o crescimento e lise na fase estacionária (Buist *et al.*, 1995; Buist *et al.*, 1997). Ela é modular e contém um sítio ativo N-terminal e um domínio de ligação a peptídeoglicanos na porção C-terminal, onde este primeiro é homólogo a muramidase-2 de *Enterococcus hirae* (Steen *et al.*, 2005). Em um trabalho publicado em 2000, Marínez-Cuesta e colaboradores demonstraram que a lise de *L. lactis* após a adição de bacteriocinas, foi causada devido aos danos acometidos na célula, o qual foi causado pela degradação incontrolada da parede celular causada pela *AcmA* (Martínez-Cuesta *et al.*, 2000).

**Tabela 9** Genes codificantes de bacteriocinas, muramidases e macrolídeos.

| <b>Locus tag</b>                      | <b>Gene</b> | <b>Proteína</b>                           |
|---------------------------------------|-------------|---|
| NCDO2118_0270                         | acmA        | N-acetylmuramidase                        |
| NCDO2118_0524                         | acmD        | N-acetylmuramidase                        |
| NCDO2118_1416                         | acmC        | N-acetyl-muramidase                       |
| NCDO2118_1435                         | -           | Lysozyme M1 (1,4-beta-N-acetylmuramidase) |
| NCDO2118_2028                         | acmB        | N-acetylmuramidase                        |
| NCDO2118_1198                         | -           | Macrolide biosynthetic protein            |
| <i>Não foi identificado no genoma</i> | -           | Lanthipeptide / Class I                   |
| NCDO2118_1768                         | -           | Class II                                  |
| <i>Não foi identificado no genoma</i> | -           | Sacpeptidase / Classe III                 |



**Fig. 23** Predição de bacteriocinas em *L. lactis* NCDO 2118. **A.** Bacteriocina de classe I, representada pela orf010, **B.** Bacteriocina de classe II, representada pela orf027 e **C.** Bacteriocina de classe III, representada pela orf011. A figura apresenta os elementos presentes na região onde foram preditas as bacteriocinas, de acordo com a legenda à direita.

### 5.4.3 IDENTIFICAÇÃO *IN SILICO* DE PROVÁVEIS PROTEÍNAS SECRETADAS

O SurfG+ aplica uma classificação das proteínas identificando-as de acordo com a presença ou ausência de peptídeo sinal, retenção de sinal e hélice transmembrana (Barinov *et al.*, 2009). Caso nenhum destes motivos tenha sido encontrado na sequência proteica, ele então, as caracteriza como citoplasmáticas (CYT) (Santos *et al.*, 2012).

Para *L. lactis* NCDO 2118, foram preditas mais de 90 proteínas caracterizadas como secretadas.

De acordo com o trabalho de Luerce e colaboradores, na comparação dos efeitos anti-inflamatórios entre as linhagens NCDO 2118, IL1403 E MG1363, apenas o sobrenadante da linhagem NCDO 2118 foi capaz de diminuir a produção de IL-8 (em 45%) mostrando sua capacidade imunomoduladora diante a inflamação (Luerce *et al.*, 2014).

Por meio destes dados, as proteínas secretadas de *L. lactis* NCDO 2118 foram preditas e comparadas com as proteínas identificadas em *L. lactis* IL1403. Nesta comparação, das mais de 90 proteínas preditas, restaram 34, exclusivas de *L. lactis* NCDO 2118 conforme tabela abaixo.

**Tabela 10** Predição de proteínas secretadas exclusivas de *L. lactis* NCDO 2118:

| Locus tag     | Gene          | Start  | Stop   | Product  |
|---------------|---------------|--------|--------|--|
| NCDO2118_0052 | NCDO2118_0052 | 57803  | 58270  | Hypothetical protein   |
| NCDO2118_0128 | epsX          | 133945 | 134712 | Polysaccharide biosynthesis protein  |
| NCDO2118_0139 | epsK          | 144750 | 145652 | Polysaccharide biosynthesis protein  |
| NCDO2118_0140 | epsL          | 145677 | 146600 | Transcriptional regulator  |
| NCDO2118_0212 | NCDO2118_0212 | 214606 | 215988 | Hypothetical protein   |
| NCDO2118_0256 | NCDO2118_0256 | 255719 | 256297 | Hypothetical protein   |
| NCDO2118_0291 | NCDO2118_0291 | 285998 | 287113 | Endoglucanase  |
| NCDO2118_0294 | NCDO2118_0294 | 288612 | 289397 | Hypothetical protein   |
| NCDO2118_0322 | NCDO2118_0322 | 311982 | 313025 | Peptidoglycan hydrolase  |
| NCDO2118_0483 | NCDO2118_0483 | 478392 | 479351 | Hypothetical protein   |
| NCDO2118_0682 | NCDO2118_0682 | 696943 | 697533 | Peptidoglycan hydrolase  |
| NCDO2118_0683 | NCDO2118_0683 | 697697 | 698158 | Hypothetical protein   |
| NCDO2118_0684 | NCDO2118_0684 | 698177 | 701176 | ChW repeat-/cell adhesion domain-containing transglutaminase-like protease |
| NCDO2118_0693 | NCDO2118_0693 | 708825 | 709256 | Hypothetical protein   |
| NCDO2118_0904 | NCDO2118_0904 | 939391 | 940704 | Hypothetical protein   |

|               |               |         |         |  |
|---------------|---------------|---------|---------|--|
| NCDO2118_0907 | NCDO2118_0907 | 944910  | 946208  | Transcriptional regulator              |
| NCDO2118_0932 | tagL          | 971010  | 971921  | Exopolysaccharide biosynthesis protein |
| NCDO2118_0942 | NCDO2118_0942 | 985414  | 986700  | Hypothetical protein                   |
| NCDO2118_0991 | NCDO2118_0991 | 1034860 | 1035321 | Hypothetical protein                   |
| NCDO2118_1361 | NCDO2118_1361 | 1468537 | 1469364 | Hypothetical protein                   |
| NCDO2118_1362 | NCDO2118_1362 | 1469396 | 1474342 | Hypothetical protein                   |
| NCDO2118_1363 | NCDO2118_1363 | 1474372 | 1475115 | Hypothetical protein                   |
| NCDO2118_1364 | NCDO2118_1364 | 1475137 | 1475901 | Hypothetical protein                   |
| NCDO2118_1420 | NCDO2118_1420 | 1537567 | 1538400 | Hypothetical protein                   |
| NCDO2118_1625 | NCDO2118_1625 | 1755880 | 1756671 | Hypothetical protein                   |
| NCDO2118_1795 | NCDO2118_1795 | 1927992 | 1929140 | Transcriptional regulator              |
| NCDO2118_2048 | NCDO2118_2048 | 2199962 | 2201083 | Aspartate protease                     |
| NCDO2118_2077 | NCDO2118_2077 | 2227730 | 2228593 | Hypothetical protein                   |
| NCDO2118_2086 | yufC          | 2235671 | 2236618 | Alpha/beta hydrolase                   |
| NCDO2118_2130 | NCDO2118_2130 | 2279925 | 2280251 | Hypothetical protein                   |
| NCDO2118_2151 | NCDO2118_2151 | 2304776 | 2305051 | Hypothetical protein                   |
| NCDO2118_2232 | NCDO2118_2232 | 2371307 | 2372062 | Hypothetical protein                   |
| NCDO2118_2330 | NCDO2118_2330 | 2482143 | 2482712 | Hypothetical protein                   |
| NCDO2118_2350 | usp45         | 2503404 | 2504741 | Secreted 45 kDa protein                |

---

## 6. CONCLUSÕES

---

Diante das características já conhecidas de *L. lactis* NCDO 2118 e de tantas outras, preditas neste estudo, a integração das mais variadas formas de análises de bioinformática e a evolução das abordagens de sequenciamento, chegam com o intuito de desvendar características específicas, que conferem tanto potencial a esta linhagem. Áreas como a genômica comparativa, proteômica, além das diversas ferramentas disponíveis para predições de genes que possam estar relacionados aos processos exclusivos desta linhagem, somam meios de descobrirmos novas aplicações, sejam elas nas áreas alimentícias, médias ou biotecnológicas.

A partir das análises realizadas neste trabalho, nos deparamos com características específicas dentro das subespécies de *Lactococcus*, como o alto grau de sintenia gênica, quando comparado seus genomas em relação a todas as espécies. O fato de *L. lactis* NCDO 2118 e *L. lactis* KF147 terem sido isoladas de planta, pode ter contribuído fortemente para a semelhança vista entre elas, além das ilhas genômicas preditas apenas nos genomas destas duas linhagens de *L. lactis*.

As chances de se identificar as proteínas-alvo que podem estar relacionadas ao efeito probiótico de *L. lactis* NCDO 2118, crescem ainda mais, pois apenas 34, das mais de 90 proteínas são exclusivas da linhagem de interesse se comparado com a não probiótica, *L. lactis* IL1403.

Este trabalho abre vários caminhos a serem investigados. Imaginando nas possibilidades de aplicação já conhecidas desta linhagem, as tantas características ainda ocultas nos faz crer em vários estudos promissores que estão por vir.

---

## 7. PERSPECTIVAS

---

Os dados gerados neste trabalho nos levam a investigar profundamente, vários pontos relacionados à *L. lactis* NCDO 2118, como por exemplo:

- Com o genoma de *L. lactis* NCDO 2118 sequenciado e anotado, um novo espaço se abre para os estudos em novas áreas ômicas;
- Possíveis mecanismos que possuem correlação ao local de isolamento desta linhagem;
- Aprofundar o estudo do plasmídeo de *L. lactis* NCDO 2118, pois muitas regiões codificantes são proteínas hipotéticas e existe uma semelhança expressiva com o plasmídeo da espécie *L. lactis* KF147;
- Entender a ação dos profagos encontrados nesta espécie, intactos e incompletos;
- Descobrir a ação das bacteriocinas preditas na funcionalidade da ação probiótica;
- Realizar ensaios experimentais com os genes preditos, possivelmente relacionados à resistência ao estresse ácido; competição por exclusão e adesão;
- Identificar, dentre as proteínas exclusivas de *L. lactis* NCDO 2118, as que possam estar relacionadas diretamente com a capacidade probiótica desta linhagem.

---

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

---

1. Achour-Rokbani A, Cordi A, Poupin P, Bauda P, Billard P (2010) **Characterization of the ars gene cluster from extremely arsenic-resistant Microbacterium sp. strain A33.** *Appl Environ Microbiol* 76: 948-955.
2. Adeghate E, Ponery A (2002) **GABA in the endocrine pancreas: cellular localization and function in normal and diabetic rats.** *Tissue & Cell* 34 (1): 1-6.
3. Agren J, Sundström A, Håfström T, Segerman B (2012) **Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups.** *PLoS One* 7: e39107.
4. Alikhan N, Petty N, Zakour N, Beatson S (2011) **BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons.** *BMC Genomics* 12: 402.
5. Altermann E, Russell WM, Azcarate-Peril MA, Barrangou R, Buck BL et al. (2005) **Complete genome sequence of the probiotic lactic acid bacterium Lactobacillus acidophilus NCFM.** *Proc Natl Acad Sci U S A* 102: 3906-3912.
6. Amann RI, Ludwig W, Shleifer KH (1995) **Phylogenetic identification and in situ detection of individual microbial cells without cultivation.** *Microbiol Ver* 59: 143-169.
7. Ataie-Jafari A, Larijani B, Alavi Majd H, Tahbaz F (2009) **Cholesterol-lowering effect of probiotic yogurt in comparison with ordinary yogurt in mildly to moderately hypercholesterolemic subjects.** *Ann Nutr Metab* 54: 22-27.
8. Bar H, Yacoby I, Benhar I (2008) **Killing cancer cells by targeted drug-carrying phage nanomedicines.** *BMC Biotechnol* 8: 37.
9. Barinov A, Loux V, Hammani A, Nicolas P, Langella P et al. (2009) **Prediction of surface exposed proteins in Streptococcus pyogenes, with a potential application to other Gram-positive bacteria.** *Proteomics* 9: 61-73.
10. Bermudez LG, Corthier G, Langella P (2004) **Recent advances in the use of Lactococcus lactis as live recombinant vector for the development of new safe mucosal vaccines.** *Recent research developments in microbiology* 8: 147-160.



11. Bibel D (1988) **Elie Metchnikoff's Bacillus of Long Life**. *ASM News* 54: 661-665.
12. Bingham RJ, Rudiño-Piñera E, Meenan NAG, Schwarz-Linek U, Turkenburg JP et al. (2008) **Crystal structures of fibronectin-binding sites from Staphylococcus aureus FnBPA in complex with fibronectin domains**. *Proc Natl Acad Sci U S A* 105: 12254-12258.
13. Blauwkamp TA, Ninfa AJ (2003) **Antagonism of PII signalling by the AmtB protein of Escherichia coli**. *Mol Microbiol* 48: 1017-1028.
14. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) **Scaffolding pre-assembled contigs using SSPACE**. *Bioinformatics* 27: 578-579.
15. Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD et al. (2004) **Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus**. *Nat Biotechnol* 22: 1554-1558.
16. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarne K et al. (2001) **The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403**. *Genome Res* 11: 731-753.
17. Bowers J, Mitchell J, Beer E, Buzby PR, Causey M et al. (2009) **Virtual terminator nucleotides for next-generation DNA sequencing**. *Nat Methods* 6: 593-595.
18. Buist G, Karsens H, Nauta A, van Sinderen D, Venema G et al. (1997) **Autolysis of Lactococcus lactis caused by induced overproduction of its major autolysin, AcmA**. *Appl Environ Microbiol* 63: 2722-2728.
19. Buist G, Kok J, Leenhouts KJ, Dabrowska M, Venema G et al. (1995) **Molecular cloning and nucleotide sequence of the gene encoding the major peptidoglycan hydrolase of Lactococcus lactis, a muramidase needed for cell separation**. *J Bacteriol* 177: 1554-1563.
20. Cabello FC, Godfrey HP, Newman SA (2007) **Hidden in plain sight: Borrelia burgdorferi and the extracellular matrix**. *Trends Microbiol* 15: 350-354.
21. Cadavid D, Bai Y, Dail D, Hurd M, Narayan K et al. (2003) **Infection and inflammation in skeletal muscle from nonhuman primates infected with**

**different genospecies of the Lyme disease spirochete *Borrelia burgdorferi*.** *Infect Immun* 71: 7087-7098.

22. Carr FJ, Chill D, Maida N (2002) **The lactic acid bacteria: a literature survey.** *Crit Rev Microbiol* 28: 281-370.
23. Carvalho AL, Turner DL, Fonseca LL, Solopova A, Catarino T et al. (2013) **Metabolic and transcriptional analysis of acid stress in *Lactococcus lactis*, with a focus on the kinetics of lactic acid pools.** *PLoS One* 8: e68470.
24. Casjens S (2003) **Prophages and bacterial genomics: what have we learned so far?** *Mol Microbiol* 49: 277-300.
25. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD (2010) **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.** *Nucleic Acids Res* 38:D473-9.
26. Chaillou S, Champomier-Vergès M, Cornet M, Crutz-Le Coq A, Dudez A et al. (2005) **The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K.** *Nat Biotechnol* 23: 1527-1533.
27. Chauvière G, Coconnier MH, Kernéis S, Fourniat J, Servin AL (1992) **Adhesion of human *Lactobacillus acidophilus* strain LB to human enterocyte-like Caco-2 cells.** *J Gen Microbiol* 138 Pt 8: 1689-1696.
28. Chevreux B, Wetter T, Suhai S (1999) **Genome sequence assembly using trace signals and additional sequence information.** *German Conference on Bioinformatics GCB'99*: 45-46.
29. Christensen JE, Dudley EG, Pederson JA, Steele JL (1999) **Peptidases and amino acid catabolism in lactic acid bacteria.** *Antonie Van Leeuwenhoek* 76: 217-246.
30. Coates ARM, Hu Y (2007) **Novel approaches to developing new antibiotics for bacterial infections.** *Br J Pharmacol* 152: 1147-1154.

31. Cogan TM, Beresford TP, Steele J, Broadbent J, Shah NP et al. (2007) **Invited review: Advances in starter cultures and cultured foods.** *J Dairy Sci* 90: 4005-4021.
32. Cohendy M (1906) **Déscription d'un ferment lactique puissant capable de s'acclimater dans l'intestin d'homme.** *Comptes rendus des séances et mémoires de la société de biologie* 60: 558-560.
33. Cohendy M (1906) **Essai de traitement de l'enterite mucomembraneuse aigue par l'acclimation d'un ferment lactique dans le gros intestine.** *Comptes rendus des séances et mémoires de la société de biologie* 60: 363.
34. Corcoran BM, Stanton C, Fitzgerald G, Ross RP (2008) **Life under stress: the probiotic stress response and how it may be manipulated.** *Curr Pharm Des* 14: 1382-1399.
35. Corr SC, Li Y, Riedel CU, O'Toole PW, Hill C et al. (2007) **Bacteriocin production as a mechanism for the antiinfective activity of Lactobacillus salivarius UCC118.** *Proc Natl Acad Sci U S A* 104: 7617-7621.
36. Cotter P, Hill C, Ross P (2005) **Bacteriocins: developing innate immunity for food..** *Nat Rev Microbiol* 3 (10): 777-788.
37. Cronin M, Ventura M, Fitzgerald GF, van Sinderen D (2011) **Progress in genomics, metabolism and biotechnology of bifidobacteria.** *Int J Food Microbiol* 149: 4-18.
38. Darling AE, Mau B, Perna NT (2010) **progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement.** *PLoS One* 5: e11147.
39. Delves-Broughton J, Blackburn P, Evans RJ, Hugenholtz J (1996) **Applications of the bacteriocin, nisin.** *Antonie Van Leeuwenhoek* 69: 193-202.
40. Du Y, Gisselberg JE, Johnson JD, Lee PJ, Prigge ST et al. (2010) **Lactococcus lactis fabH, encoding beta-ketoacyl-acyl carrier protein synthase, can be functionally replaced by the Plasmodium falciparum congener.** *Appl Environ Microbiol* 76: 3959-3966.
41. Díaz-Muñiz I, Banavara DS, Budinich MF, Rankin SA, Dudley EG et al. (2006) **Lactobacillus casei metabolic potential to utilize citrate as an energy**

**source in ripening cheese: a bioinformatics approach.** *J Appl Microbiol* 101: 872-882.

42. Euzéby J. (2007). **List of Prokaryotic names with Standing in Nomenclature.** online.
43. Even S, Lindley ND, Coccagn-Bousquet M (2003) **Transcriptional, translational and metabolic regulation of glycolysis in *Lactococcus lactis* subsp. cremoris MG 1363 grown in continuous acidic cultures.** *Microbiology* 149: 1935-1944.
44. FAO, WHO (2002) **Guidelines for the Evaluation of Probiotics in Food.** *Food and Agriculture Organization* : 1-11.
45. Fleischmann R, Adams M, White O, Clayton R, Kirkness E et al. (1995) **Whole-genome random sequencing and assembly of *Haemophilus influenzae*.** *Science* 269: 496-512.
46. Foligné B, Daniel C, Pot B (2013) **Probiotics from research to market: the possibilities, risks and challenges.** *Curr Opin Microbiol* 16: 284-292.
47. Fouladbakhsh J (2012) **Complementary and alternative modalities to relieve osteoarthritis symptoms: a review of the evidence on several therapies often used for osteoarthritis management.** *Orthop Nurs* 31: 115-121.
48. Fozo EM, Kajfasz JK, Quivey RGJ (2004) **Low pH-induced membrane fatty acid alterations in oral bacteria.** *FEMS Microbiol Lett* 238: 291-295.
49. Galardini M, Biondi EG, Bazzicalupo M, Mengoni A (2011) **CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes.** *Source Code Biol Med* 6: 11.
50. Gao XW, Mubasher M, Fang CY, Reifer C, Miller LE (2010) **Dose-response efficacy of a proprietary probiotic formula of *Lactobacillus acidophilus* CL1285 and *Lactobacillus casei* LBC80R for antibiotic-associated diarrhea and *Clostridium difficile*-associated diarrhea prophylaxis in adult patients.** *Am J Gastroenterol* 105: 1636-1641.
51. Garneau JE, Moineau S (2011) **Bacteriophages of lactic acid bacteria and their impact on milk fermentations.** *Microb Cell Fact* 10 Suppl 1: S20.

52. Garrity GM, Liburn TG, Cole JR *et al.* (2007). Part 1 - The "Archea", Phyla "Crenarchaeota" and "Euryarchaeota". **Taxonomic Outline of the Bacteria and Archaea**, Release 7.7, 6 Mar 2007: 6-31.
53. Gillor O, Giladi I, Riley MA (2009). **Persistence of colicinogenic Escherichia coli in the mouse gastrointestinal tract.** *BMC Microbiol* 9:165.
54. Goh YJ, Klaenhammer TR (2010) **Functional roles of aggregation-promoting-like factor in stress tolerance and adherence of Lactobacillus acidophilus NCFM.** *Appl Environ Microbiol* 76: 5005-5012.
55. Gratia A (1925) **Sur un remarquable exemple d'antagonisme entre deux souches de Colibacille.** *C. R. Soc. Biol* 93: 1040–1041.
56. Greenlee H, Kwan ML, Ergas IJ, Sherman KJ, Krathwohl SE *et al.* (2009) **Complementary and alternative therapy use before and after breast cancer diagnosis: the Pathways Study.** *Breast Cancer Res Treat* 117: 653-665.
57. He F, Ouwehan AC, Hashimoto H, Isolauri E, Benno Y *et al.* (2001) **Adhesion of Bifidobacterium spp. to human intestinal mucus.** *Microbiol Immunol* 45: 259-262.
58. Henson J, Tischler G, Ning Z (2012) **Next-generation sequencing and large genome assemblies.** *Pharmacogenomics* 13: 901-915.
59. Hill DA, Artis D (2010) **Intestinal bacteria and the regulation of immune cell homeostasis.** *Annu Rev Immunol* 28: 623-667.
60. Hirsch A, Grinsted E, Chapman H, Mattick A (1951) **A note on the inhibition of an anaerobic sporeformer in Swiss-type cheese by a nisin-producing Streptococcus.** *J. Dairy Sci.* 18: 205–206.
61. Holzapfel WH, Haberer P., Geisen R, Björkroth J, Schillinger U (2001) **Taxonomy and important features of probiotic microorganisms in food and nutrition.** *Am J Clin Nutr* 73: 365S-373S.
62. Hosono A (1992) **Fermented milk in the orient.** *Functions of fermented milk. Challenges for the health sciences* Elsevier Applied Science: 61-78.

63. Howell TH (1988) **Metchnikoff and prolongation of life.** *Age Ageing* 17: 420-421.
64. Huson DH, Bryant D (2006) **Application of phylogenetic networks in evolutionary studies.** *Mol Biol Evol* 23: 254-267.
65. Inoue K, Shirai T, Ochiai H, Kasao M, Hayakawa K et al. (2003) **Blood-pressure-lowering effect of a novel fermented milk containing gamma-aminobutyric acid (GABA) in mild hypertensives.** *Eur J Clin Nutr* 57: 490-495.
66. Jack RW, Tagg JR, Ray B (1995) **Bacteriocins of gram-positive bacteria.** *Microbiol Rev* 59: 171-200.
67. Jakobs C, Jaeken J, Gibson K (1993) **Inherited disorders of GABA metabolism.** *J Inher Metab Dis* 16: 704-715.
68. Janeway CA, Travers P, Walport M, Shlomchik MJ (2007) **Imunologia - O sistema imune na saúde e na doença.** *Artmed* 6: 848.
69. Johnson BR, Klaenhammer TR (2014) **Impact of genomics on the field of probiotic research: historical perspectives to modern paradigms.** *Antonie Van Leeuwenhoek* 106: 141-156.
70. Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U et al. (2013) **Updating benchtop sequencing performance comparison.** *Nat Biotechnol* 31: 294-296.
71. Kamba A, Shibata Y, Mizoguchi E (2013) **Potential roles of chitin in mucosal inflammation.** *Microbial pathogens and strategies for combating them: science, technology and education* FORMATEX: 1853-1863.
72. Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J et al. (2009) **Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein.** *Proc Natl Acad Sci U S A* 106: 17193-17198.
73. Karp PD, Paley S, Romero P (2002) **The Pathway Tools software.** *Bioinformatics* 18 Suppl 1:S225-32.

74. Karpiński TM, Szkaradkiewicz AK (2013) **Characteristic of bacteriocines and their application.** *Pol J Microbiol* 62: 223-235.
75. Kaur R, Malik CP (2013) **NEXT GENERATION SEQUENCING: A REVOLUTION IN GENE SEQUENCING.** *CIBTech Journal of Biotechnology* 2 (4): 1-20.
76. Kawada M, Chen C, Arihiro A, Nagatani K, Watanabe T et al. (2008) **Chitinase 3-like-1 enhances bacterial adhesion to colonic epithelial cells through the interaction with bacterial chitin-binding protein.** *Lab Invest* 88: 883-895.
77. Khalid K (2011) **An overview of lactic acid bacteria.** *International Journal of Biosciences* 1: 1-13.
78. Klaenhammer T, Altermann E, Arigoni F, Bolotin A, Breidt F et al. (2002) **Discovering lactic acid bacteria by genomics.** *Antonie Van Leeuwenhoek* 82: 29-58.
79. Klaenhammer TR (1993) **Genetics of bacteriocins produced by lactic acid bacteria.** *FEMS Microbiol Rev* 12: 39-85.
80. Kohn AB, Moroz TP, Barnes JP, Netherton M, Moroz LL (2013) **Single-cell semiconductor sequencing.** *Methods Mol Biol* 1048: 247-284.
81. Kuipers OP, Beerthuyzen MM, de Ruyter PG, Luesink EJ, de Vos WM (1995) **Autoregulation of nisin biosynthesis in *Lactococcus lactis* by signal transduction.** *J Biol Chem* 270: 27299-27304.
82. Kutter E, Sulakvelidze A (2005) **Bacteriophages: Biology and Applications.** *CRC Press* 2: 1-528.
83. Langella P, Loir Y (1999) **Heterologous protein secretion in *Lactococcus lactis*: a novel antigen delivery system.** *Brazilian Journal of Medical and Biological Research* 32(2): 191-198.
84. Lebeer S, Claes IJJ, Verhoeven TLA, Shen C, Lambrichts I et al. (2008) **Impact of luxS and suppressor mutations on the gastrointestinal transit of *Lactobacillus rhamnosus* GG.** *Appl Environ Microbiol* 74: 4711-4718.

85. Lebeer S, Vanderleyden J, De Keersmaecker SCJ (2008) **Genes and molecules of lactobacilli supporting probiotic action**. *Microbiol Mol Biol Rev* 72: 728-64, Table of Contents.
86. Lee S, Kim D, Bae K, Byun S, Chung J (2000) **Enhancement of secretion and extracellular stability of staphylokinase in *Bacillus subtilis* by wprA gene disruption**. *Appl. Environ. Microbiol* 66: 476-480.
87. Linares DM, Kok J, Poolman B (2010) **Genome sequences of *Lactococcus lactis* MG1363 (revised) and NZ9000 and comparative physiological studies**. *J Bacteriol* 192: 5806-5812.
88. Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M et al. (2012) **High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity**. *Nat Rev Microbiol* 10: 599-606.
89. Luerce T (2013) **Caracterização e Avaliação das Propriedades Probióticas de *Lactococcus lactis* NCDO 2118**. PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA Tese de Doutorado: 1-167.
90. Luerce TD, Gomes-Santos AC, Rocha CS, Moreira TG, Cruz DN et al. (2014) **Anti-inflammatory effects of *Lactococcus lactis* NCDO 2118 during the remission period of chemically induced colitis**. *Gut Pathog* 6: 33.
91. Machado-Vieira R, Bressan R, Frey B, Soares J (2005) **As bases neurobiológicas do transtorno bipolar**. *Rev. Psiq. Clín* 32, supl 1: 28-33.
92. Mardis ER (2008) **Next-generation DNA sequencing methods**. *Annu Rev Genomics Hum Genet* 9: 387-402.
93. Martinussen J, Schallert J, Andersen B, Hammer K (2001) **The pyrimidine operon pyrRPB-carA from *Lactococcus lactis***. *J Bacteriol* 183: 2785-2794.
94. Martínez-Cuesta MC, Kok J, Herranz E, Peláez C, Requena T et al. (2000) **Requirement of autolytic activity for bacteriocin-induced lysis**. *Appl Environ Microbiol* 66: 3174-3179.
95. Mateos LM, Ordóñez E, Letek M, Gil JA (2006) ***Corynebacterium glutamicum* as a model bacterium for the bioremediation of arsenic**. *Int Microbiol* 9: 207-215.



96. Mazzoli R, Pessione E, Dufour M, Laroute V, Giuffrida MG et al. (2010) **Glutamate-induced metabolic changes in *Lactococcus lactis* NCDO 2118 during GABA production: combined transcriptomic and proteomic analysis.** *Amino Acids* 39: 727-737.
97. Meijerink M, van Hemert S, Taverne N, Wels M, de Vos P et al. (2010) **Identification of genetic loci in *Lactobacillus plantarum* that modulate the immune response of dendritic cells using comparative genome hybridization.** *Plos One* 5: e10632.
98. Mierau I, Kleerebezem M (2005) **10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*.** *Appl Microbiol Biotechnol* 68: 705-717.
99. Miyoshi A, Jamet E, Commissaire J, Renault P, Langella P et al. (2004) **A xylose-inducible expression system for *Lactococcus lactis*.** *FEMS Microbiol Lett* 239: 205-212.
100. Moineau S, Lévesque C (2005) **The control of bacteriophages in industrial fermentations.** . In *Bacteriophages: Biology and applications* CRC Press: 285-296.
101. Morita H, Toh H, Oshima K, Yoshizaki M, Kawanishi M et al. (2011) **Complete genome sequence and comparative analysis of the fish pathogen *Lactococcus garvieae*.** *PLoS One* 6: e23184.
102. Nadalin F, Vezzi F, Policriti A (2012) **GapFiller: a de novo assembly approach to fill the gap within paired reads.** *BMC Bioinformatics* 13 Suppl 14: S8.
103. Nouaille S, Ribeiro LA, Miyoshi A, Pontes D, Le Loir Y et al. (2003) **Heterologous protein production and delivery systems for *Lactococcus lactis*.** *Genet Mol Res* 2: 102-111.
104. O'Flaherty SJ, Klaenhammer TR (2010) **Functional and phenotypic characterization of a protein from *Lactobacillus acidophilus* involved in cell morphology, stress tolerance and adherence to intestinal cells.** *Microbiology* 156: 3360-3367.

105. O'Neill E, Pozzi C, Houston P, Humphreys H, Robinson DA et al. (2008) **A novel *Staphylococcus aureus* biofilm phenotype mediated by the fibronectin-binding proteins, FnBPA and FnBPB.** *J Bacteriol* 190: 3835-3850.
106. O'Sullivan E, Condon S (1999) **Relationship between acid tolerance, cytoplasmic pH, and ATP and H<sup>+</sup>-ATPase levels in chemostat cultures of *Lactococcus lactis*.** *Appl Environ Microbiol* 65: 2287-2293.
107. Ohnishi N, Maruyama F, Ogawa H, Kachi H, Yamada S et al. (2014) **Genome Sequence of a *Bacillus anthracis* Outbreak Strain from Zambia, 2011.** *Genome Announc* 2: .
108. Ouwehand AC, Salminen S, Isolauri E (2002) **Probiotics: an overview of beneficial effects.** *Antonie Van Leeuwenhoek* 82: 279-289.
109. Patti JM, Allen BL, McGavin MJ, Höök M (1994) **MSCRAMM-mediated adherence of microorganisms to host tissues.** *Annu Rev Microbiol* 48: 585-617.
110. Pfeiler EA, Klaenhammer TR (2007) **The genomics of lactic acid bacteria.** *Trends Microbiol* 15: 546-553.
111. Piro VC, Faoro H, Weiss VA, Steffens MBR, Pedrosa FO et al. (2014) **FGAP: an automated gap closing tool.** *BMC Res Notes* 7: 371.
112. Porter CK, Gormley R, Tribble DR, Cash BD, Riddle MS (2011) **The Incidence and gastrointestinal infectious risk of functional gastrointestinal disorders in a healthy US adult population.** *Am J Gastroenterol* 106: 130-138.
113. Rafter J, Bennett M, Caderni G, Clune Y, Hughes R et al. (2007) **Dietary synbiotics reduce cancer risk factors in polypectomized and colon cancer patients.** *Am J Clin Nutr* 85: 488-496.
114. Ramos RT, Carneiro AR, Baumbach J, Azevedo V, Schneider MP, Silva, A (2011) **Analysis of quality raw data of second generation sequencers with Quality Assessment Software.** *BMC Res Notes* 4: 130.
115. Rettger L, Cheplin H (1921) **A treatise on the transformation of the intestinal flora, with special reference to implantation of *Bacillus acidophilus*.** *Yale University Press* : .

116. Rettger LF (1915) **THE INFLUENCE OF MILK FEEDING ON MORTALITY AND GROWTH, AND ON THE CHARACTER OF THE INTESTINAL FLORA.** *J Exp Med* 21: 365-388.
117. Rocha C (2011) **Propriedades imuno modulatórias de *Lactobacillus delbrueckii*** . *PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA* Tese de Doutorado: 1-188.
118. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W et al. (2011) **An integrated semiconductor device enabling non-optical genome sequencing.** *Nature* 475: 348-352.
119. Sanders J, Venema G, Kok J (1999) **Environmental stress responses in *Lactococcus lactis*.** *FEMS Microbiology Reviews* 23: 483-501.
120. Sanders M, in't Huis-Veld J (1999) **Bringing a probiotic-containing functional food to the market: microbiological, product, regulatory and labeling issues.** *Antonie van Leeuwenhoek* 76: 293-315.
121. Santos AR, Carneiro A, Gala-García A, Pinto A, Barh D et al. (2012) **The *Corynebacterium pseudotuberculosis* in silico predicted pan-exoproteome.** *BMC Genomics* 13 Suppl 5: S6.
122. Schleifer K, Kraus J, Dvorak C, Kilpper-Bälz R, Collins M et al. (1985) **Transfer of *Streptococcus lactis* and Related Streptococci to the Genus *Lactococcus* gen. nov..** *Systematic and Applied Microbiology* 6: 183-195.
123. Sharples GJ, Bolt EL, Lloyd RG (2002) **RusA proteins from the extreme thermophile *Aquifex aeolicus* and lactococcal phage r1t resolve Holliday junctions.** *Mol Microbiol* 44: 549-559.
124. Shen YA, Nahas R (2009) **Complementary and alternative medicine for treatment of irritable bowel syndrome.** *Can Fam Physician* 55: 143-148.
125. Shortt C (1999) **The probiotic century: historical and current perspectives.** *Trends in Food Science & Technology* 10: 411-417.
126. Siezen RJ, Bayjanov J, Renckens B, Wels M, van Hijum SAFT et al. (2010) **Complete genome sequence of *Lactococcus lactis* subsp. *lactis* KF147, a plant-associated lactic acid bacterium.** *J Bacteriol* 192: 2649-2650.

127. Siezen RJ, Bayjanov JR, Felis GE, van der Sijde MR, Starrenburg M et al. (2011) **Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays.** *Microb Biotechnol* 4: 383-402.
128. Smit BA, van Hylckama Vlieg JET, Engels WJM, Meijer L, Wouters JTM et al. (2005) **Identification, cloning, and characterization of a *Lactococcus lactis* branched-chain alpha-keto acid decarboxylase involved in flavor formation.** *Appl Environ Microbiol* 71: 303-311.
129. Soares SC, Abreu VAC, Ramos RTJ, Cerdeira L, Silva A et al. (2012) **PIPS: pathogenicity island prediction software.** *PLoS One* 7: e30848.
130. Soares SC, Trost E, Ramos RTJ, Carneiro AR, Santos AR et al. (2013) **Genome sequence of *Corynebacterium pseudotuberculosis* biovar equi strain 258 and prediction of antigenic targets to improve biotechnological vaccine production.** *J Biotechnol* 167: 135-141.
131. Steen A, Buist G, Horsburgh GJ, Venema G, Kuipers OP et al. (2005) **AcmA of *Lactococcus lactis* is an N-acetylglucosaminidase with an optimal number of LysM domains for proper functioning.** *FEBS J* 272: 2854-2868.
132. Tian J, Dang H, Kaufman DL (2011) **Combining antigen-based therapy with GABA treatment synergistically prolongs survival of transplanted  $\beta$ -cells in diabetic NOD mice.** *PLoS One* 6: e25337.
133. Tuomola EM, Salminen SJ (1998) **Adhesion of some probiotic and dairy *Lactobacillus* strains to Caco-2 cell cultures.** *Int J Food Microbiol* 41: 45-51.
134. Ushiyama A, Tanaka K, Aiba Y, Shiba T, Takagi A et al. (2003) ***Lactobacillus gasseri* OLL2716 as a probiotic in clarithromycin-resistant *Helicobacter pylori* infection.** *J Gastroenterol Hepatol* 18: 986-991.
135. Vaaje-Kolstad G, Bunaes AC, Mathiesen G, Eijsink VGH (2009) **The chitinolytic system of *Lactococcus lactis* ssp. *lactis* comprises a nonprocessive chitinase and a chitin-binding protein that promotes the degradation of alpha- and beta-chitin.** *FEBS J* 276: 2402-2415.
136. Varuzza L (2013) **Introdução à análise de dados de sequenciadores de nova geração.** 2.0.1: 1-76.

137. Verma A, Brissette CA, Bowman A, Stevenson B (2009) **Borrelia burgdorferi BmpA is a laminin-binding protein.** *Infect Immun* 77: 4940-4946.
138. Vélez MP, De Keersmaecker SCJ, Vanderleyden J (2007) **Adherence factors of Lactobacillus in the human gastrointestinal tract.** *FEMS Microbiol Lett* 276: 140-148.
139. Whitehead H, Cox G (1935) **The occurrence of bacteriophage in lactic streptococci.** *N Z J Dairy Sci Technol* 16: 319-320.
140. Wommack KE, Colwell RR (2000) **Virioplankton: viruses in aquatic ecosystems.** *Microbiol Mol Biol Rev* 64: 69-114.
141. Wong CGT, Bottiglieri T, Snead OC3 (2003) **GABA, gamma-hydroxybutyric acid, and neurological disease.** *Ann Neurol* 54 Suppl 6: S3-12.
142. Wood B, Warner P (2003) **Genetics of Lactic Acid Bacteria.** *KLUWER ACADEMIC PUB* 3: 160.
143. Wu S, Ye R, Wu X, Ng S, Wong S (1998) **Enhanced secretory production of a single-chain antibody fragment from Bacillus subtilis by coproduction of molecular chaperones.** *J. Bacteriol.* 180: 2830-2835.
144. Wu X, Lee W, Tran L, Wong S (1991) **Engineering a Bacillus subtilis expression-secretion system with a strain deficient in six extracellular proteases.** *J. Bacteriol* 173: 4952-4958.
145. Yeh GY, Davis RB, Phillips RS (2006) **Use of complementary therapies in patients with cardiovascular disease.** *Am J Cardiol* 98: 673-680.
146. Zhou B, Gao W, Zhang T, Chen H (2005) **[Application of chitin and its derivatives to pharmaceutical industry for traditional Chinese medicine].** *Zhongguo Zhong Yao Za Zhi* 30: 1313-1316.
147. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011) **PHAST: a fast phage search tool.** *Nucleic Acids Res* 39: W347-52.
148. de Jong A, van Heel AJ, Kok J, Kuipers OP (2010) **BAGEL2: mining for bacteriocins in genomic data.** *Nucleic Acids Res* 38: W647-51.

149. de Melo AL, Machado CR, Pereira LH (1993) **Host cell adhesion to *Schistosoma mansoni* larvae in the peritoneal cavity of naive mice. Histological and scanning electron microscopic studies.** *Rev Inst Med Trop Sao Paulo* 35: 17-22.
150. de Moreno de Leblanc A, Del Carmen S, Zurita-Turk M, Santos Rocha C, van de Guchte M et al. (2011) **Importance of IL-10 modulation by probiotic microorganisms in gastrointestinal inflammatory diseases.** *ISRN Gastroenterol* 2011: 892971.
151. de la Plaza M, Fernández de Palencia P, Peláez C, Requena T (2004) **Biochemical and molecular characterization of alpha-ketoisovalerate decarboxylase, an enzyme involved in the formation of aldehydes from amino acids by *Lactococcus lactis*.** *FEMS Microbiol Lett* 238: 367-374.
152. van Belkum MJ, Stiles ME (2000) **Nonantibiotic antibacterial peptides from lactic acid bacteria.** *Nat Prod Rep* 17: 323-335.
153. van de Guchte M, Ehrlich SD, Maguin E (2001) **Production of growth-inhibiting factors by *Lactobacillus delbrueckii*.** *J Appl Microbiol* 91: 147-153.

---

## 9. ANEXOS

---

|               |     |
|---------------|-----|
| ANEXO I.....  | 100 |
| ANEXO II..... | 102 |

# Genome Sequence of *Lactococcus lactis* subsp. *lactis* NCDO 2118, a GABA-Producing Strain

Letícia C. Oliveira,<sup>a</sup> Tessália D. L. Saraiva,<sup>a</sup> Siomar C. Soares,<sup>a\*</sup> Rommel T. J. Ramos,<sup>b</sup> Pablo H. C. G. Sá,<sup>b</sup> Adriana R. Carneiro,<sup>a</sup> Fábio Miranda,<sup>b</sup> Matheus Freire,<sup>b</sup> Wendel Renan,<sup>b</sup> Alberto F. O. Júnior,<sup>a</sup> Anderson R. Santos,<sup>a\*</sup> Anne C. Pinto,<sup>a</sup> Bianca M. Souza,<sup>a</sup> Camila P. Castro,<sup>a</sup> Carlos A. A. Diniz,<sup>a</sup> Clarissa S. Rocha,<sup>a</sup> Diego C. B. Mariano,<sup>a</sup> Edgar L. de Aguiar,<sup>a</sup> Edson L. Folador,<sup>a</sup> Eudes G. V. Barbosa,<sup>a</sup> Flavia F. Aburjaile,<sup>a</sup> Lucas A. Gonçalves,<sup>a</sup> Luís C. Guimarães,<sup>a</sup> Marcela Azevedo,<sup>a</sup> Pamela C. M. Agresti,<sup>a</sup> Renata F. Silva,<sup>a</sup> Sandeep Tiwari,<sup>a</sup> Sintia S. Almeida,<sup>a</sup> Syed S. Hassan,<sup>a</sup> Vanessa B. Pereira,<sup>a</sup> Vinicius A. C. Abreu,<sup>a</sup> Ulisses P. Pereira,<sup>a\*</sup> Fernanda A. Dorella,<sup>c</sup> Alex F. Carvalho,<sup>c</sup> Felipe L. Pereira,<sup>c</sup> Carlos A. G. Leal,<sup>c</sup> Henrique C. P. Figueiredo,<sup>c</sup> Artur Silva,<sup>b</sup> Anderson Miyoshi,<sup>a</sup> Vasco Azevedo<sup>a</sup>

Laboratory of Cellular and Molecular Genetics, Institute of Biologic Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil<sup>†</sup>; Institute of Biologic Sciences, Federal University of Pará, Belém, PA, Brazil<sup>‡</sup>; AQUACEN, National Reference Laboratory for Aquatic Animal Diseases, Ministry of Fisheries and Aquaculture, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil<sup>§</sup>

\* Present address: Siomar C. Soares, AQUACEN, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil; Anderson R. Santos, Federal University of Uberlândia, Minas Gerais, MG, Brazil; Ulisses P. Pereira, Federal University of Uberlândia, Minas Gerais, MG, Brazil.

***Lactococcus lactis* subsp. *lactis* NCDO 2118 is a nondairy lactic acid bacterium, a xylose fermenter, and a gamma-aminobutyric acid (GABA) producer isolated from frozen peas. Here, we report the complete genome sequence of *L. lactis* NCDO 2118, a strain with probiotic potential activity.**

Received 21 August 2014 Accepted 26 August 2014 Published 2 October 2014

**Citation** Oliveira LC, Saraiva TDL, Soares SC, Ramos RTJ, Sá PHCG, Carneiro AR, Miranda F, Freire M, Renan W, Júnior AFO, Santos AR, Pinto AC, Souza BM, Castro CP, Diniz CAA, Rocha CS, Mariano DCB, de Aguiar EL, Folador EL, Barbosa EGV, Aburjaile FF, Gonçalves LA, Guimarães LC, Azevedo M, Agresti PCM, Silva RF, Tiwari S, Almeida SS, Hassan SS, Pereira VB, Abreu VAC, Pereira UP, Dorella FA, Carvalho AF, Pereira FL, Leal CAG, Figueiredo HCP, Silva A, Miyoshi A, Azevedo V. 2014. Genome sequence of *Lactococcus lactis* subsp. *lactis* NCDO 2118, a GABA-producing strain. *Genome Announc.* 2(5):e00980-14. doi:10.1128/genomeA.00980-14.

**Copyright** © 2014 Oliveira et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Vasco Azevedo, [vasco@icb.ufmg.br](mailto:vasco@icb.ufmg.br).

Lactic acid bacteria (LAB), in general, acquire energy from the conversion of sugars into lactic acid (1) and are used for production of many fermented products, such as cheese, yogurt, butter, and wine. Food conservation is due to the medium acidification and production of molecules that inhibit the growth of undesirable microbiota, contributing to the development of desirable organoleptic properties in the final product (2). Moreover, some specific LAB strains produce bioactive molecules such as gamma-aminobutyric acid (GABA) (3), a product of glutamate decarboxylation by the glutamic acid decarboxylase (GAD) enzyme. Usually, GABA acts by modulating the central nervous system, contributing to smooth muscle relaxation and presenting hypotensor activity (4). Also, GABA can immunomodulate the immune system (5). Therefore, GABA-producing bacteria generally present probiotic properties (6). *Lactococcus lactis* NCDO 2118 is a nondairy strain, a xylose fermenter (a common trait of plant-associated strains), and a GABA producer isolated from frozen peas (6, 7).

*L. lactis* NCDO 2118 was sequenced three times, due to assembling complexity. First, the genome was decoded with the SOLiD 5500 platform with mate-paired libraries, generating a total of 5,133,057,360 bp, (coverage of 2,053 times). The reads were subjected to a Phred 20 quality filter using Quality Assessment software (8) and assembled with the CLC Genomics Workbench, generating a total of 1,641 overlapping sequences. These sequences were removed with the Simplifier (9), ordered and oriented based on the reference *L. lactis* KF147 genome sequence (a plant-

associated strain, accession number CP001834). Then manual curation was performed using Artemis (10), and SSPACE (11) and Gapfiller (12) were used to generate the scaffold and resolve gaps, respectively. At the end of curation and sequence assembly, a total of 409 scaffolds (2,874,854 bp) were obtained.

*L. lactis* NCDO 2118 was then decoded with the Ion PGM platform with fragment libraries generating a total of 187,303,001 bp (coverage of ~71 times). Genome assembly was performed using Mira 3.9 (13), and the assembled genome sequence was reference aligned with CONTIGuator (14). The redundant overlapping sequences were removed with “in-house scripts,” closing the remnant gaps. Annotation and frameshifts curation were then performed using Artemis and CLC, reducing the initial 1821 pseudogenes to 480.

Finally, the DNA was sequenced using the Ion Torrent PGM with fragment libraries, yielding a total of ~1,249,154,478 bp (coverage of 474 times). Assembly was performed with Mira 4.0.1 and Newbler 2.9 (15). We used CONTIGuator and FGAP 1.7 (16) to perform the alignment and gap closure steps, respectively. We followed the same previously explained steps for annotation and frameshift curation, reducing the pseudogenes to 52.

The complete genome of *L. lactis* NCDO 2118 consists of a single circular chromosome of 2,554,693 bp, containing 2,386 coding sequences (CDS), which had 52 pseudogenes, 66 tRNA genes, and 6 rRNA operons, with a G+C content of 34.9%. There is one plasmid, pNCDO2118 (37,571 bp), with 48 CDS, from which 4 are pseudogenes with a G+C content of 32.33%.



**Nucleotide sequence accession numbers.** The *Lactococcus lactis* NCDO 2118 chromosome and the plasmid were deposited at DDBJ/EMBL/GenBank under the accession numbers CP009054 and CP009055, respectively.

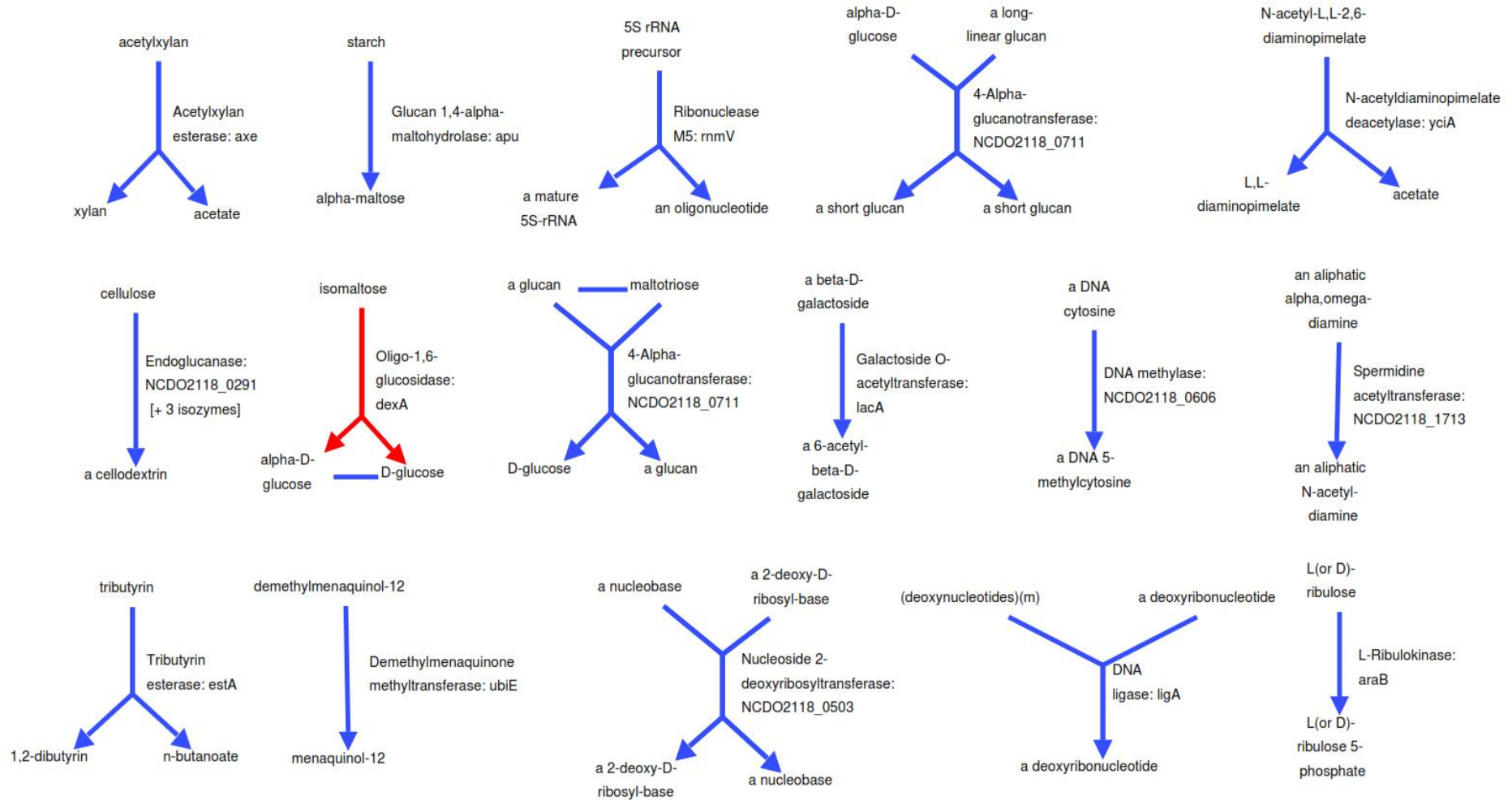
## ACKNOWLEDGMENTS

This work was supported by Rede Paraense de Genômica e Proteômica, Ministério da Pesca e Aquicultura (MPA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). We also acknowledge the support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

## REFERENCES

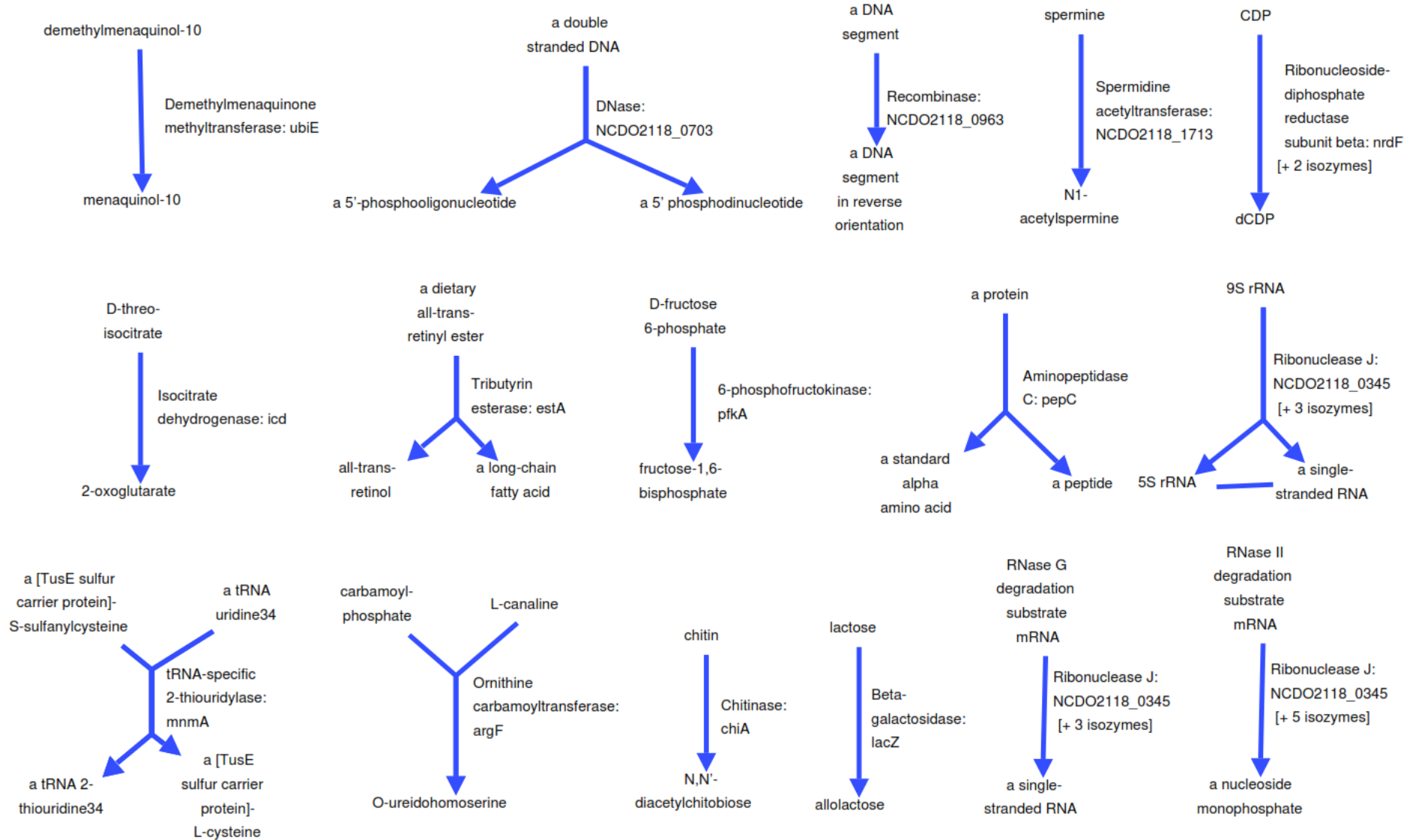
- Carr FJ, Chill D, Maida N. 2002. The lactic acid bacteria: a literature survey. *Crit. Rev. Microbiol.* 28:281–370. <http://dx.doi.org/10.1080/1040-840291046759>.
- van de Guchte M, Ehrlich SD, Maguin E. 2001. Production of growth-inhibiting factors by *Lactobacillus delbrueckii*. *J. Appl. Microbiol.* 91:147–153. <http://dx.doi.org/10.1046/j.1365-2672.2001.01369.x>.
- Zareian M, Ebrahimpour A, Bakar FA, Mohamed AK, Forghani B, Ab-Kadir MS, Saari N. 2012. A glutamic acid-producing lactic acid bacteria isolated from Malaysian fermented foods. *Int. J. Mol. Sci.* 13:5482–5497. <http://dx.doi.org/10.3390/ijms13055482>.
- Inoue K, Shirai T, Ochiai H, Kasao M, Hayakawa K, Kimura M, Sansawa H. 2003. Blood-pressure-lowering effect of a novel fermented milk containing gamma-aminobutyric acid (GABA) in mild hypertensives. *Eur. J. Clin. Nutr.* 57:490–495. <http://dx.doi.org/10.1038/sj.ejcn.1601555>.
- Jin Z, Mendu SK, Birnir B. 2013. GABA is an effective immunomodulatory molecule. *Amino Acids* 45:87–94. <http://dx.doi.org/10.1007/s00726-011-1193-7>.
- Mazzoli R, Pessione E, Dufour M, Laroute V, Giuffrida MG, Giunta C, Coccagn-Bousquet M, Loubière P. 2010. Glutamate-induced metabolic changes in *Lactococcus lactis* NCDO 2118 during GABA production: combined transcriptomic and proteomic analysis. *Amino Acids.* 39:727–737. <http://dx.doi.org/10.1007/s00726-010-0507-5>.
- Siezen RJ, Starrenburg MJ, Boekhorst J, Renckens B, Molenaar D, van Hylckama Vlieg JE. 2008. Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl. Environ. Microbiol.* 74:424–436. <http://dx.doi.org/10.1128/AEM.01850-07>.
- Ramos RT, Carneiro AR, Baumbach J, Azevedo V, Schneider MP, Silva A. 2011. Analysis of quality raw data of second generation sequencers with Quality Assessment Software. *BMC Res. Notes* 4:130. <http://dx.doi.org/10.1186/1756-0500-4-130>.
- Ramos RT, Carneiro AR, Azevedo V, Schneider MP, Barh D, Silva A. 2012. Simplifier: a web tool to eliminate redundant NGS contigs. *Bioinformatics* 8:996–999. <http://dx.doi.org/10.6026/97320630008996>.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945. <http://dx.doi.org/10.1093/bioinformatics/16.10.944>.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27:578–579. <http://dx.doi.org/10.1093/bioinformatics/btq683>.
- Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13:S8. <http://dx.doi.org/10.1186/1471-2105-13-S14-S8>.
- Chevreur B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *Comput. Sci. Biol.: Proc. German Conference on Bioinformatics GCB'99 GCB*. Hannover, Germany.
- Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol. Med.* 6:11. <http://dx.doi.org/10.1186/1751-0473-6-11>.
- Ohnishi N, Maruyama F, Ogawa H, Kachi H, Yamada S, Fujikura D, Nakagawa I, Hang'ombe MB, Thomas Y, Mweene AS, Higashi H. 2014. Genome sequence of a *Bacillus anthracis* outbreak strain from Zambia, 2011. *Genome Announc.* 2(2):e00116-14. <http://dx.doi.org/10.1128/genomeA.00116-14>.
- Piro VC, Faoro H, Weiss VA, Steffens MB, Pedrosa FO, Souza EM, Raittz RT. 2014. FGAP: an automated gap closing tool. *BMC Res. Notes* 7:371. <http://dx.doi.org/10.1186/1756-0500-7-371>.

***Lactococcus lactis* subsp. *lactis* NCDO 2118 cellular overview**



ANEXO II

**Figura adicional 1.** Vías exclusivas de *L. lactis* NCDO 2118.



**Figura adicional 2.** Vias exclusivas de *L. lactis* NCDO 2118.

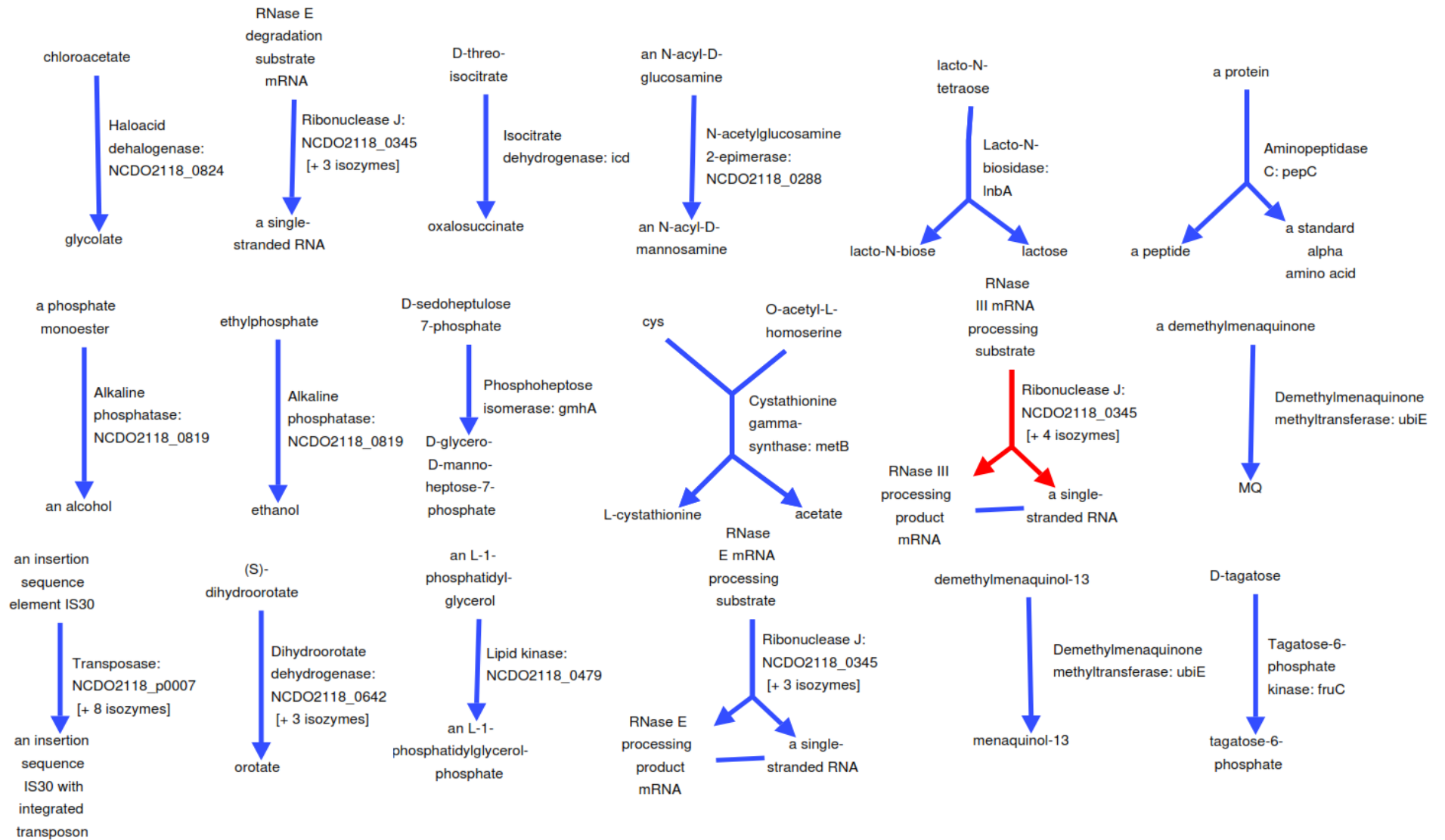


Figura adicional 3. Vias exclusivas de *L. lactis* NCDO 2118.

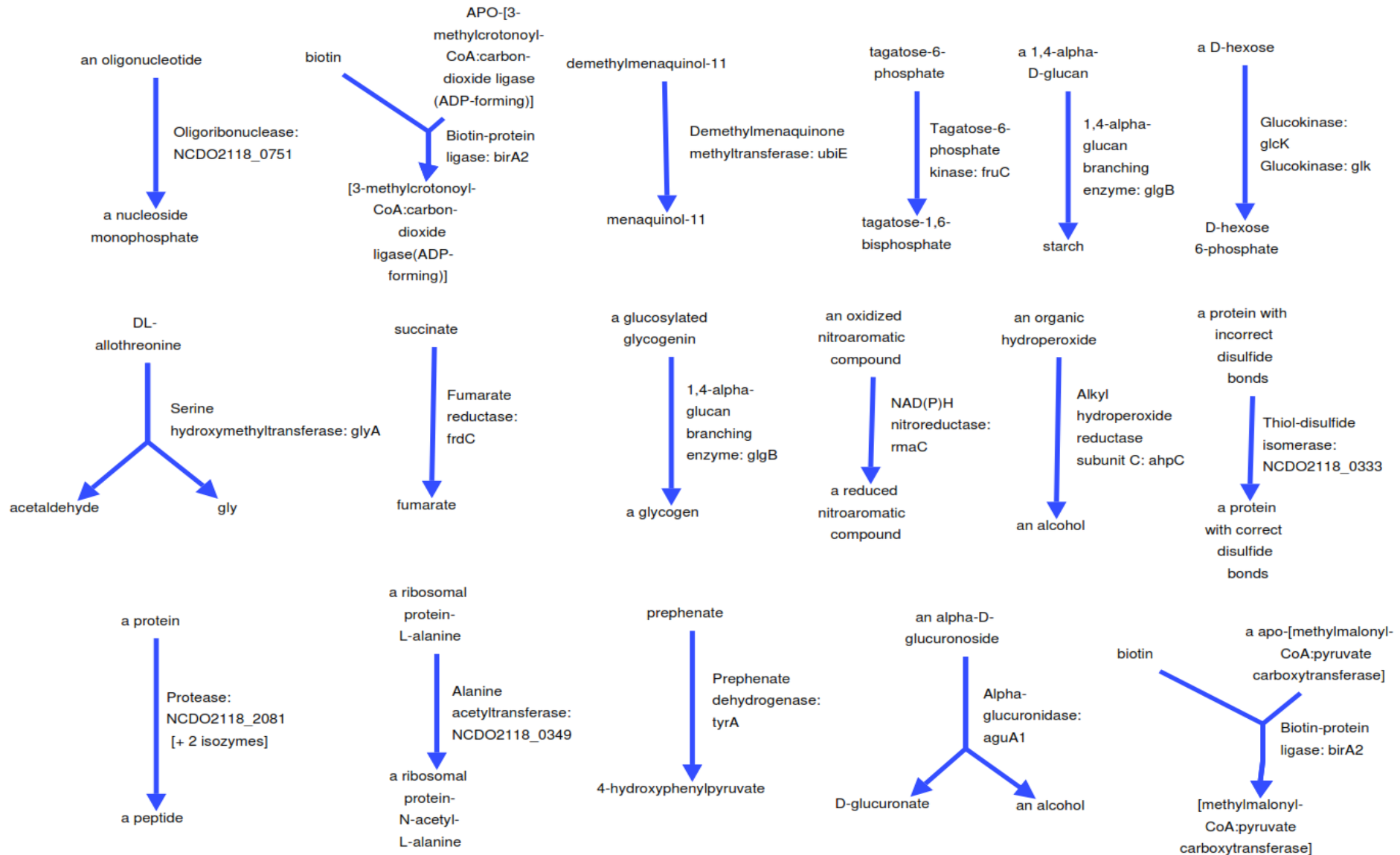


Figura adicional 4. Vias exclusivas de *L. lactis* NCDO 2118.

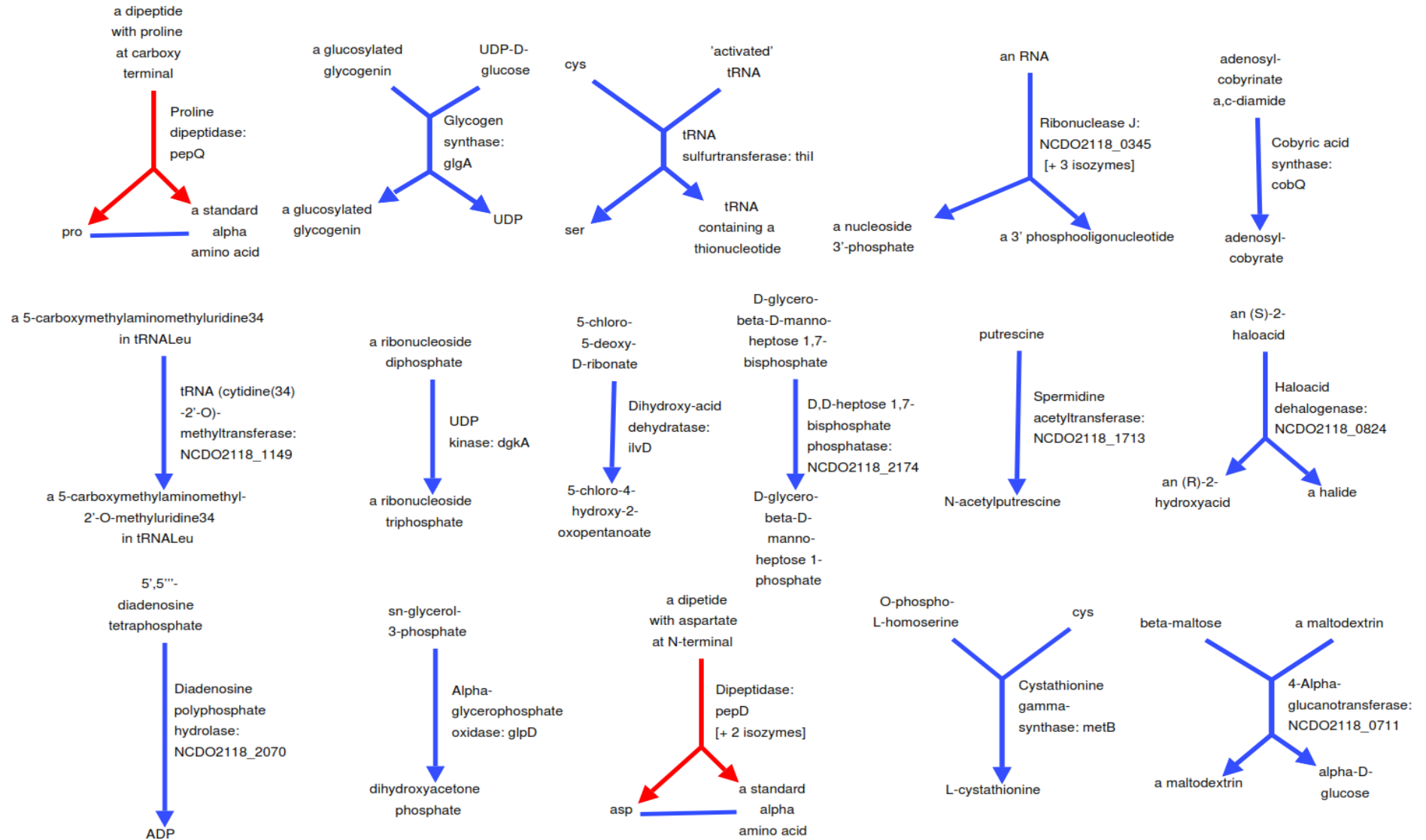


Figura adicional 5. Vías exclusivas de *L. lactis* NCDO 2118.



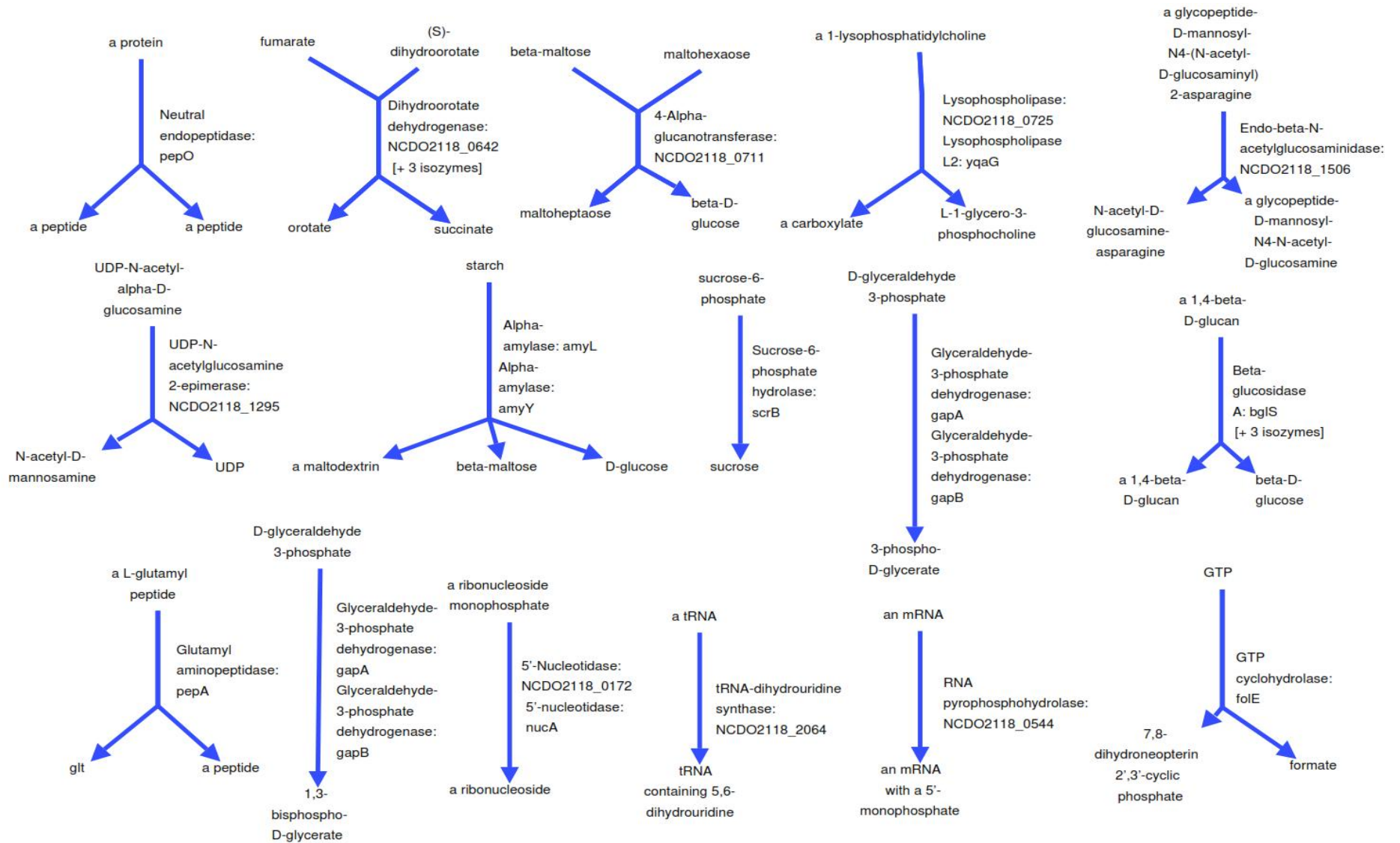


Figura adicional 6. Vias exclusivas de *L. lactis* NCDO 2118.

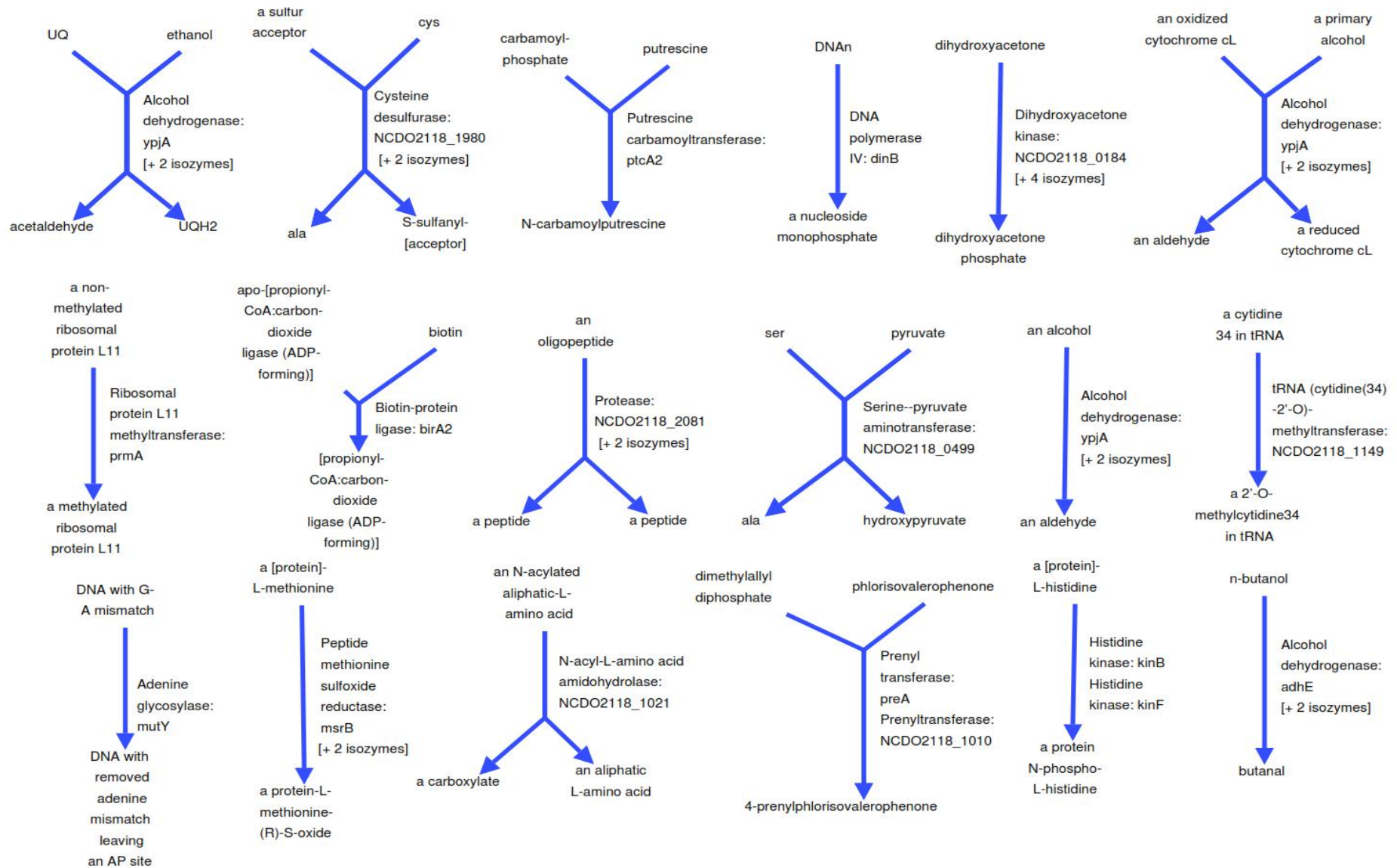


Figura adicional 7. Vias exclusivas de *L. lactis* NCDO 2118.



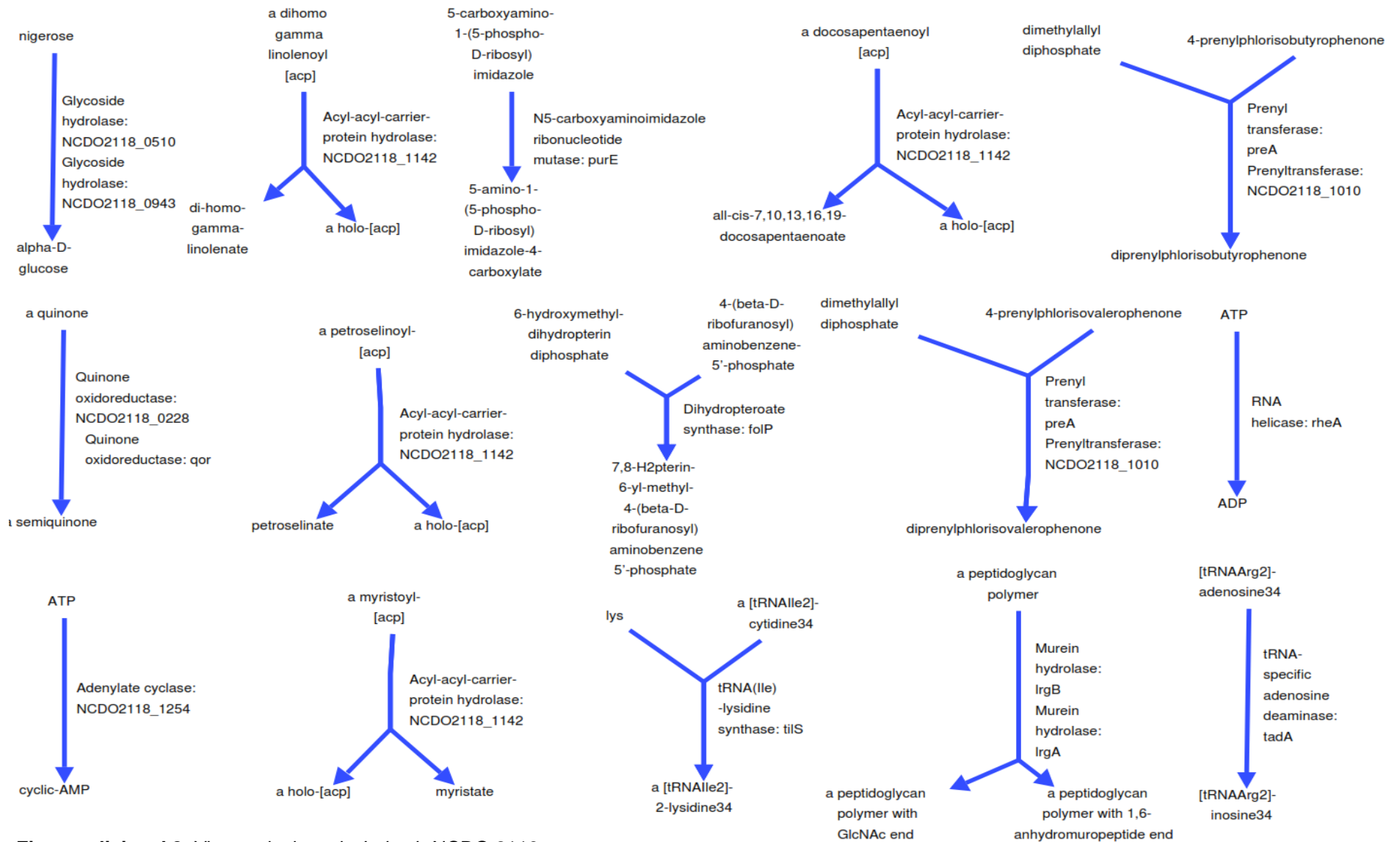


Figura adicional 8. Vias exclusivas de *L. lactis* NCDO 2118.

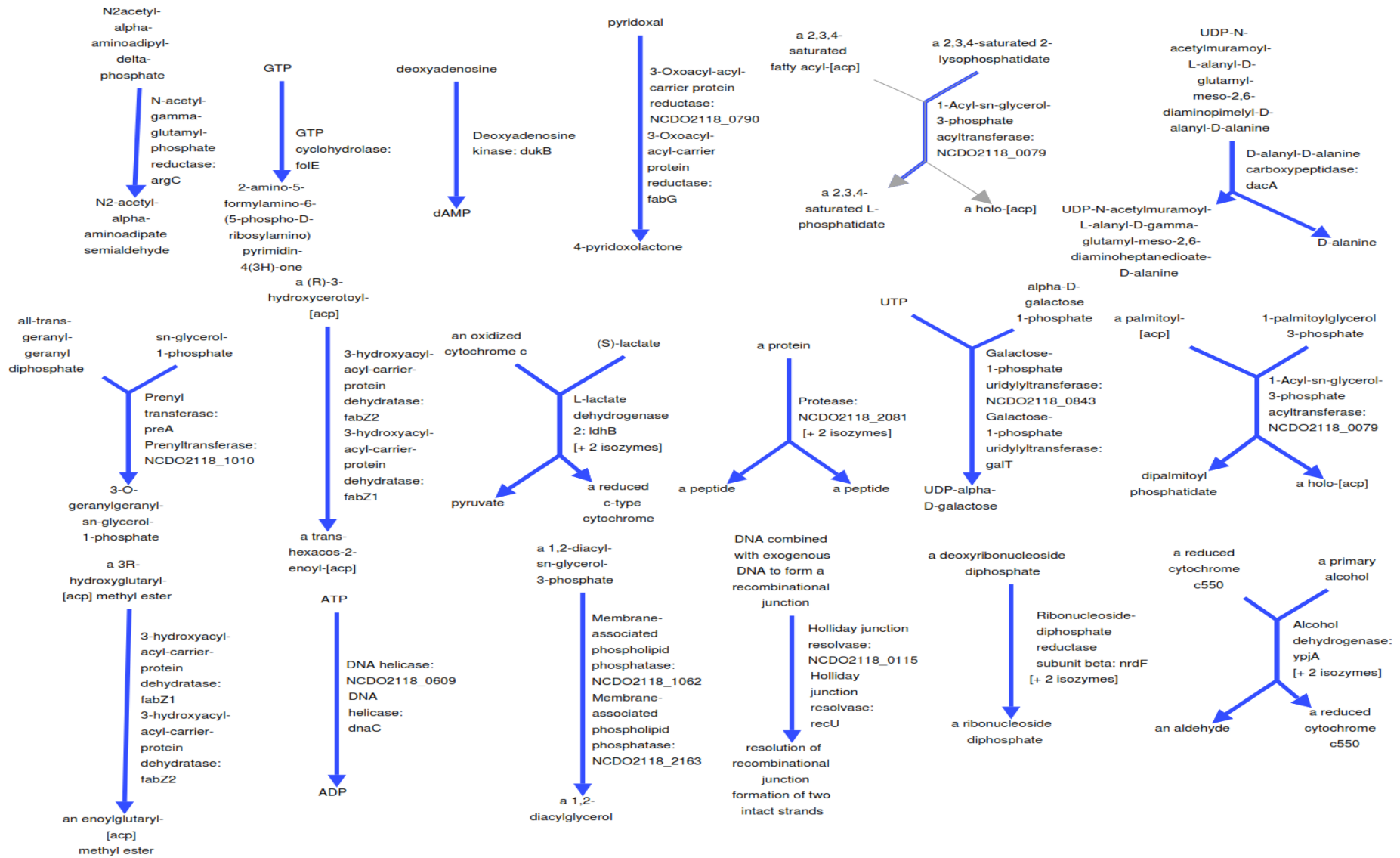


Figura adicional 9. Vias exclusivas de *L. lactis* NCDO 2118.

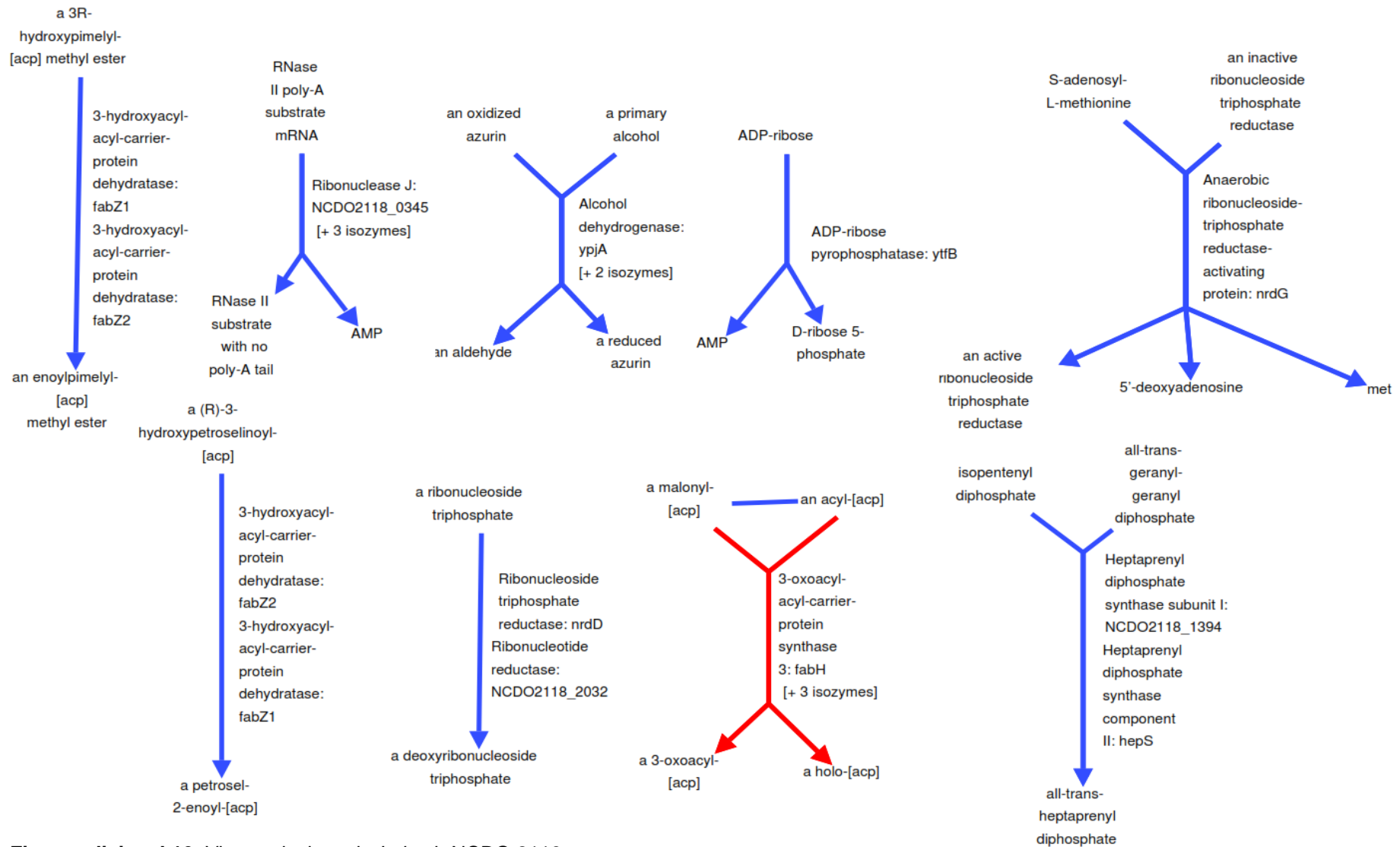


Figura adicional 10. Vias exclusivas de *L. lactis* NCDO 2118.