



HAL
open science

Functional metagenomics of the bovine rumen microbiota to boost enzyme discovery for complex polymer breakdown

Lisa Ufarté

► **To cite this version:**

Lisa Ufarté. Functional metagenomics of the bovine rumen microbiota to boost enzyme discovery for complex polymer breakdown. Animal biology. INSA de Toulouse, 2016. English. NNT : 2016ISAT0014 . tel-02917921

HAL Id: tel-02917921

<https://theses.hal.science/tel-02917921>

Submitted on 20 Aug 2020

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

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



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par

Institut National des Sciences Appliquées de Toulouse (INSA Toulouse)

Discipline ou spécialité :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries

Présentée et soutenue par

Lisa UFARTÉ

Le 25 février 2016

Titre :

Functional metagenomics of the bovine rumen microbiota to boost enzyme discovery for complex polymer breakdown

JURY

Dr Gabrielle POTOCKI-VERONESE (LISBP, INRA Toulouse), directrice de thèse

Pr. Trevor CHARLES (University of Waterloo, Canada), rapporteur

Pr. Thierry GIARDINA (Université Aix Marseille), rapporteur, président du jury

Dr Elisabeth LAVILLE (LISBP, INRA Toulouse)

Dr Diego MORGAVI (INRA, Theix)

Dr Daniel ZALKO (Toxalim, INRA Toulouse)

Pr. Magali REMAUD-SIMEON (LISBP, INSA Toulouse), membre invité

Ecole doctorale :

ED SEVAB : Ingénieries microbienne et enzymatique

Unité de recherche :

LISBP (UMR CNRS 5504, UMR INRA 792), INSA de Toulouse

NOM : UFARTÉ

PRÉNOM : Lisa

TITRE : Métagénomique fonctionnelle du microbiote du rumen bovin pour la découverte d'enzymes de dégradation de polymères naturels et synthétiques.

SPÉCIALITÉ : Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries

FILIÈRE : Ingénieries Moléculaires et Enzymatiques

ANNÉE : 2016

LIEU : INSA, Toulouse

DIRECTRICE DE THÈSE : Dr Gabrielle Potocki-Veronese

RÉSUMÉ :

Le microbiote du rumen bovin est un écosystème très diversifié et efficace pour la dégradation de substrats complexes, notamment issus de la biomasse végétale. Composé majoritairement de microorganismes non cultivés, il constitue un réservoir très riche de nouvelles enzymes d'intérêt potentiel pour les biotechnologies industrielles, en particulier les bioraffineries et la bioremédiation. Dans le cadre de cette thèse, nous avons mis en œuvre une approche de criblage fonctionnel du métagenome ruminal pour accélérer la découverte d'enzymes de dégradation des lignocelluloses, mais aussi de divers polluants synthétiques. En particulier, de nouvelles estérases capables de dégrader un insecticide de la famille des carbamates, le fenobucarb, ainsi qu'un polyuréthane commercial, l'Impranil DLN, ont pu être identifiées. De plus, le développement d'une nouvelle stratégie de criblage d'oxydoréductases nous a permis d'isoler trois enzymes bactériennes originales, très polyspécifiques, ne requérant ni cuivre ni manganèse pour dégrader différents substrats polycycliques tels que des polluants majeurs de l'industrie textile, mais aussi des dérivés de lignine. Enfin, le criblage de deux banques issues d'enrichissements *in vivo* et *in vitro* du microbiome du rumen sur paille de blé a permis d'isoler des cocktails d'enzymes lignocellulolytiques au profil fonctionnel et d'origine taxonomique différents, constitués de glycoside-hydrolases, estérases et oxydoréductases. Quinze nouveaux modules CAZy, correspondant à des familles enzymatiques jamais caractérisées, ont été identifiés. L'ensemble de ces résultats met en lumière l'immense potentiel d'innovation biotechnologique contenu dans les écosystèmes microbiens, en particulier dans le microbiote du rumen bovin.

MOTS-CLÉS : métagénomique fonctionnelle, rumen bovin, enzymes, criblage haut débit

ÉCOLE DOCTORALE:

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

LABORATOIRE :

Laboratoire d'Ingénieries des Systèmes Biologiques et des Procédés (LISBP)

INSA/CNRS 5504 - UMR INSA/INRA 792

NAME: UFARTÉ

FIRST NAME: Lisa

TITLE: Functional metagenomics of the bovine rumen microbiota to boost enzyme discovery for complex polymer breakdown.

SPECIALTY: Ecological, Veterinary, Agronomic Sciences and Bioengineering

FIELD: Enzymatic and Molecular engineering

YEAR: 2016

PLACE: INSA, Toulouse

THESIS DIRECTOR: Dr Gabrielle Potocki-Veronese

SUMMARY:

Bovine rumen microbiota is a highly diverse and efficient ecosystem for the degradation of complex substrates, especially those issued from plant biomass. Predominantly composed of uncultivated microorganisms, it constitutes a rich reservoir of new enzymes of potential interest for industrial biotechnologies, especially biorefineries and bioremediation. As part of this thesis, we used the functional screening of the ruminal metagenome to increase the discovery of enzymes able to degrade lignocelluloses, as well as different synthetic pollutants. In particular, new esterases able to degrade a carbamate insecticide, fenoucarb, and a commercial polyurethane, Impranil DLN, have been identified. Moreover, the development of a new screening strategy for oxidoreductases allowed the isolation of three original bacterial enzymes that are very polyspecific, and do not need copper nor manganese to degrade different polycyclic substrates, like major pollutants of the textile industry, as well as lignin derivatives. Finally, the screening of two libraries from *in vivo* and *in vitro* enrichments of the ruminal microbiome on wheat straw allowed the isolation of lignocellulolytic enzymatic cocktails, with different functional profiles and taxonomical origins, comprising glycoside-hydrolases, esterases and oxidoreductases. Fifteen novel CAZy modules, related to enzymatic families never characterized, were identified. All these results highlight the vast potential of microbial ecosystems, in particular the bovine rumen microbiota, for biotechnological innovation.

KEYWORDS: functional metagenomics, bovine rumen, enzymes, high-throughput screening

DOCTORAL SCHOOL:

Ecological, Veterinary, Agronomic Sciences and Bioengineering (SEVAB)

LABORATORY:

Laboratory of Biosystems and Chemical Engineering (LISBP)

INSA/CNRS 5504 - UMR INSA/INRA 792

PUBLICATIONS

Peer reviewed journal articles:

Ufarté L., Potocki-Veronese G., Laville E. (2015).

Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology. *Front. Microbiol.* 6:563. doi: [10.3389/fmicb.2015.00563](https://doi.org/10.3389/fmicb.2015.00563)

Ufarté L., Duquesne S., Laville E., Potocki-Veronese G. (2015)

Metagenomics for the discovery of pollutant degrading enzymes. *Biotechnology Advances.* doi: [10.1016/j.biotechadv.2015.10.009](https://doi.org/10.1016/j.biotechadv.2015.10.009)

Passerini D.*, Vuillemin M.*, Ufarté L., Morel S., Loux V., Fontagné-Faucher C., Monsan P., Remaud-Siméon M., Moulis C. (2015)

Inventory of the GH70 enzymes encoded by *Leuconostoc citreum* NRRL B-1299: identification of three novel α -transglucosylases. *FEBS Journal* 282, 2115–2130. doi:[10.1111/febs.13261](https://doi.org/10.1111/febs.13261) * Contributed equally

Ufarté L., Potocki-Veronese G., Cecchini D., Rizzo A., Morgavi D., Cathala B., Moreau C., Cleret M., Robe P., Klopp C., Bozonnet S., Laville E.

Highly promiscuous oxidases discovered by functional exploration of the bovine rumen microbiome. *To be submitted 2016.*

Ufarté L., Laville E., Cecchini D., Rizzo A., Amblard E., Drula E., Henrissat B., Cleret M., Lazuka A., Hernandez G., Morgavi D., Dumon C., Robe P., Klopp C., Bozonnet S., Potocki-Veronese G.

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown. *To be submitted 2016.*

Ufarté L., Laville E., Duquesne S., Morgavi D., Robe P., Klopp C., Rizzo A., Bozonnet S. and Potocki-Veronese G.

Discovery of carbamate degrading enzymes by functional metagenomics. *To be submitted 2016*

Book chapter:

Laville E., Ufarté L., Potocki-Veronese G. (2015)

Découverte de nouvelles fonctions et familles protéiques: nouveaux défis pour les biotechnologies et l'écologie microbienne. In *La métagénomique: développements et futures applications.* Quae.

Ufarté L., Bozonnet S., Laville E., Cecchini D. A., Pizzut-Serin S., Jacquiod S., Demanèche S., Simonet P., Franqueville L., Potocki-Veronese G.

Functional metagenomics: construction and high-throughput screening of fosmid libraries for discovery of novel carbohydrate-active enzymes. In Martin, F., Uroz, S. (Eds.), *Microbial*

Environmental Genomics (MEG). Springer New York, New York, NY, pp. 257–271. doi : [10.1007/978-1-4939-3369-3_15](https://doi.org/10.1007/978-1-4939-3369-3_15)

ORAL COMMUNICATIONS

Ufarté L., Laville E., Morgavi D., Ladeveze S., Bozonnet S., Robe P., Henrissat B., Remaud-Simeon M., Monsan P., Potocki-Veronese G.

Functional metagenomics boosts enzyme discovery for complex polymer breakdown.

SEVAB Doctoral school day, Toulouse (France), 5 march 2015.

Ufarté L., Laville E., Morgavi D., Ladeveze S., Bozonnet S., Robe P., Henrissat B., Remaud-Simeon M., Monsan P., Potocki-Veronese G

Functional metagenomics boosts enzyme discovery for complex polymer breakdown

2nd International Symposium on Green Chemistry (ISGC), La Rochelle (France), 3-7 may 2015.

Ufarté L., Tauzin A., Gherbovet O., Lebaz N., Lazuka A., Abot A.

Transverse approaches for the optimization of lignocellulose degradation.

1st PhD and Post-Doctoral Students Day, LISBP, Toulouse (France), 8 july 2015.

POSTER COMMUNICATIONS

Ufarté L., Morgavi D., Laville E., Potocki-Veronese G.

Functional metagenomics for biocatalyst discovery: applications for environmental biotechnologies and green chemistry

SEVAB doctoral school day, Toulouse (France), 12 November 2013.

Third place for best poster.

Ufarté L., Morgavi D., Laville E., Potocki-Veronese G.

Functional metagenomics for biocatalyst discovery: applications for environmental biotechnologies and green chemistry

Club Bioconversions en Synthèse Organique (CBSO) 25th symposium, Carry-le-Rouet (France), 3-6 June 2014.

Ufarté L., Laville E., Morgavi D., Hernandez-Raquet G., Bozonnet S., Dumon C., Robe P., Henrissat B., Potocki-Veronese G.

Functional metagenomics boosts enzyme discovery for plant cell wall polymer breakdown.

11th Carbohydrate Bioengineering Meeting (CBM), Espoo (Finland), 10-13 May 2015.

Ufarté L., Laville E., Morgavi D., Hernandez-Raquet G., Bozonnet S., Dumon C., Robe P., Henrissat B., Potocki-Veronese G.

Functional metagenomics boosts enzyme discovery for plant cell wall polymer breakdown.

1st PhD and Post-Doctoral Students Day, LISBP, Toulouse (France), 8 July 2015.

Après rédaction du manuscrit qui suit, je pense être maintenant à la partie la plus difficile de cette rédaction, les remerciements. Non pas parce que je n'ai personne à remercier, bien au contraire, mais surtout parce que rien de ce qui va suivre ne pourra vraiment exprimer toute la reconnaissance que j'ai envers ceux que j'ai côtoyé, que ce soit quelques mois ou quelques années, pendant cette expérience de maintenant trois ans et demie.

Je voudrais commencer par remercier l'ensemble du jury d'avoir accepté de juger mon travail de thèse. Merci à Trevor Charles et Thierry Giardina, d'avoir accepté d'évaluer ce manuscrit et pour l'intérêt que vous avez porté à mon travail de thèse. Un très grand merci aussi à Daniel Zalko et Diego Morgavi, qui en plus d'avoir accepté de faire parti de mon jury, avez aussi été d'excellents conseils en tant que membres de mon comité de thèse. Merci donc pour ces discussions et un merci aussi particulier à Diego pour les échantillons de rumen à la base de ce projet de thèse. Merci à Magali d'avoir accepté l'invitation à ce jury, pour les idées apportées lors de réunions d'équipe, mais surtout pour m'avoir accueillie dans cette équipe depuis mon stage de fin d'étude. Un très grand merci à Elisabeth, pour tes conseils, ton soutien, ta patience... Et sur ces points, je ne peux bien sûr pas t'oublier Gaby ! Que ce soit toi aussi par ta patience, tes conseils, ton optimisme, ton dynamisme... J'ai conscience que cela n'a pas toujours dû être facile (et je pense qu'Elisabeth le confirmera), mais je me rends compte aussi de la chance que j'ai eu, et je voulais vous remercier pour tout ce que vous m'avez apporté, tant au niveau professionnel que personnel. En bref, MERCI !

Il y a aussi deux personnes que j'aimerais tout particulièrement remercier : Claire et Marlène. Claire, tu m'as acceptée comme stagiaire dans cette équipe, bien avant ma thèse, et bien que je travaillais majoritairement avec Marlène, tu as toujours pris le temps de discuter, de t'intéresser aussi à ce que je voulais pour mon futur, et c'est toi qui m'a en premier parlée de Gaby, de son sujet de thèse, et pour tout cela, je ne peux là aussi que te remercier. Et Marlène, tu es celle qui m'a vraiment fait découvrir le travail au labo, ce que c'était vraiment qu'une enzyme, expérimentalement parlant, et qui m'a fait aimer ça. Je me souviens aussi de ton soutien, à la fin de mon stage, pour la préparation du concours SEVAB et ta présence avec moi ce jour là. Pour tout ça et tant d'autres choses, je ne peux que te dire merci. Merci à toutes les deux pour m'avoir fait découvrir les enzymes et cette équipe, sans quoi je n'en serais peut-être pas là aujourd'hui.

Et aujourd'hui, j'en suis à la fin d'un projet de thèse de trois ans et demi, dont les résultats obtenus sont le fruit de diverses collaborations qui ont été pour moi extrêmement enrichissantes.

C'est pour cela que j'aimerais remercier Adèle Lazuka et Guillermina Hernandez-Raquet pour l'enrichissement *in vitro*, réalisé pendant la thèse d'Adèle, et qui m'a permis d'avoir ma deuxième banque.

Merci aussi à Patrick Robe pour la fabrication de ces banques.

Je voudrais remercier par la même Christophe Klopp pour les séquençages et contigages de mes clones.

Merci à Bernard Henrissat et son équipe, notamment Elodie Drula, pour les annotations des CAZy que vous avez réalisées.

Merci à Bernard Cathala et son équipe, notamment Céline Moreau et Nadège Beury, pour les films que vous m'avez envoyés.

De l'équipe, je souhaiterais aussi dire un grand merci à Sophie et Sandra, pour votre aide sur la plateforme ICEO, et à Mégane et Emilie pour votre aide sur certaines des expériences. Sophie D., pour nos discussions sur les estérases et Sandrine et David pour celles sur les oxydoréductases. Angeline, bien sûr, pour ton aide précieuse pendant ces quelques mois, mais aussi pour ton optimisme et ta bonne humeur pendant les années que l'on a passé dans le même labo (et après !). A Nelly aussi, pour ton aide sur les HPLC et la MS, et Isabelle Dufau (TWB), pour les analyses HPSEC que l'on a réalisé.

Merci aussi à mes collègues de l'enseignement, vous qui m'avez fait découvrir le monde d'en face (littéralement ou non) : Stéphane, Mourad, Angelica, les Marion, Marie, Ines, Sandrine, Flo, Aurore, Nathalie... et beaucoup d'autres !

Mais aussi et surtout, merci à toute l'équipe de l'EAD1, que j'ai côtoyé pendant presque quatre ans maintenant, et pour beaucoup, vous avez été là dans les meilleurs moments comme pour les pires, et je ne peux que vous dire merci pour votre soutien, la bonne ambiance, et tout ce que vous m'avez apporté : Yvonne, Haiyang, Nathalie C., Guillaume, Bowl, Hélène, Marie-Laure, Nada, Julien R., Marion C., Marion S., Benjamin, Eleni, Marlène C., Dana, Marie H., Marie G., Pablo (1 et 2 !), Zhongpeng, Laurence, Etienne, Régis, Cédric, Claire D., Gianluca, Alain, Flo, Florent, Alvaro, Maeva, Pauline, Betty, Yoann, Marc, Christopher, Delphine... Et tous ceux présents et passés, que je n'ai pas cités.

Un merci particulier à Marina, Louise et Jean-Christophe. On se connaît depuis plus ou moins longtemps, mais je voulais vous remercier pour les discussions que nous avons eues, les moments de rigolade et de détente bien nécessaires (surtout ces derniers mois !) et d'avoir été là quand il le fallait, avec les mots qu'il fallait. Et à Barbie aussi pour, entre autres, nos après-midi TSS (la prochaine fois, je t'aurai au « Eve lève toi » ! ;-)

Un autre merci particulier (ça en fait beaucoup !), mais non des moindres, à mes compagnons de bureau, qui ont dû me supporter pendant si longtemps, mais qui avez, vous aussi, toujours été là, que ce soit Romain, Vinciane, Cyrielle et Susana au début de mon stage, et Yannick, Virginie, Julien D. et Louison à la fin de ma thèse, en passant par beaucoup d'autres, tout aussi présents.

Enfin, un grand merci à ma famille, mes parents surtout, mes frères, ma belle-sœur, mon neveu et ma nièce, pour leur soutien. Sans vous, je ne serais pas arrivée où j'en suis aujourd'hui.

MERCI

A papa et maman,

« Never doubt that a small group
of committed people can change the world;
indeed, it's the only thing that ever has. »

Margaret Mead

« Take a method and try it.
If it fails, admit it frankly, and try another.
But by all means, try something »

Franklin D. Roosevelt

Table of content

Abbreviations	19
General introduction	23
References.....	28
Chapter I: Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology	29
Abstract:.....	31
1. Introduction.....	31
2. Sampling Strategies	32
3. Functional Screening: New Challenges for the Discovery of Functions	33
3.1. The Sequence, Marker of Originality.....	33
3.2. Activity Screening: Speeding up the Discovery of Biotechnology Tools	37
3.3. The Immense Challenges of Ultra-fast Screening	39
4. Conclusion	42
Author Contributions	42
Acknowledgments	43
Artwork.....	43
Table content	43
References.....	44
Chapter II: Metagenomics for the discovery of pollutant degrading enzymes	55
Abstract	57
1. Introduction.....	57
2. Targeted metagenome sampling	59
3. Screening for pollutant degrading enzymes.....	59
3.1. Sequence-based approaches to understand pollutant degradation mechanisms by microbial communities.....	60
3.2. Activity based metagenomics for the discovery of new tools for bioremediation	64
3.2.1. Oxidoreductases	65
3.2.1.1. Oxygenases.....	65
3.2.1.2. Laccases.	67
3.2.1.3. Alkane degrading enzymes.....	72
3.2.2. Hydrolases	72
3.2.2.1. Esterases.....	72
3.2.2.2. Amidohydrolases.....	74
3.2.2.3. Nitrilases.....	75

4. Conclusion and future goals	75
Acknowledgements	76
Artwork.....	76
Table content	76
References.....	77
Chapter III: The bovine rumen ecosystem	83
1. Physiological properties of the rumen	85
2. Bovine nutrition.....	87
3. Microbial diversity in the bovine gastrointestinal tract	90
3.1. Bacteria.....	93
3.2. Bacteriophages	94
3.3. Methanogens	94
3.4. Protozoa	95
3.5. Anaerobic fungi	96
3.6. Structural and functional modulation of the ruminal microbiota.....	96
4. Enzymes from the rumen	97
4.1. Carbohydrate active enzymes	97
4.1.1. Cellulases	107
4.1.2. Hemicellulases.....	108
4.1.3. Pectinases	109
4.1.4. Auxiliary activities.....	110
4.1.5. Amylases.....	110
4.2. Oxidoreductases (not referenced in the CAZy database)	110
4.3. Phytases.....	111
4.4. Tannases.....	111
4.5. <i>p</i> -coumaric acid esterases	112
4.6. Lipases	112
4.7. Enzymes involved in polyunsaturated fatty acids biohydrogenation	113
5. Application of rumen enzymes.....	114
5.1. Feed additives:	114
5.2. Biorefineries:	114
5.3. Bio-based organic acids production:	115
6. Meta-omic analyses of the rumen microbiota:.....	115
6.1. Metagenomics:.....	115

6.1.1. Sequence-based metagenomics:.....	115
6.1.2. Activity-based metagenomics:	116
6.2. Metatranscriptomics:	117
6.3. Metaproteomics:.....	117
6.4. Metabolomics and meta-metabolomics:	117
7. Conclusion:	118
Artwork.....	122
Table content	123
References.....	124
Thesis goals	135
Chapter IV: Discovery of carbamate degrading enzymes by functional metagenomics.....	141
Abstract:	144
1. Introduction:.....	144
2. Results and discussion:.....	146
2.1. Screening of the bovine rumen metagenomic library.....	146
2.2. Polyurethane degradation.....	150
2.2.1. HPSEC analyses	151
2.2.2. MALDI-TOF analyses.....	152
2.3. Insecticide/herbicide carbamate degradation	153
2.4. Sequence analysis.....	155
3. Material and methods:.....	164
3.1. Chemicals.....	164
3.2. Metagenomic DNA sampling and library construction	164
3.3. Primary high-throughput screening of the metagenomic library	165
3.4. Discrimination screening	165
3.5. Sequencing and data analysis.....	166
3.6. Functional characterisation of enzymatic activities	167
3.6.1. Carbamate and polyurethane degradation reactions	167
3.6.2. High pressure size exclusion chromatography (HP-SEC): polyurethane characterisation	167
3.6.3. Matrix assisted laser desorption/ionization – time-of-flight (MALDI-TOF) mass spectrometry	167
3.6.4. High performance liquid chromatography – mass spectrometry	168
4. Conclusion	168
Acknowledgements	169

Artwork.....	169
Table content	169
Supplementary information	170
References :	173
Chapter V: Highly promiscuous oxidases discovered by functional exploration of the bovine rumen microbiome	179
Abstract	182
1. Introduction.....	182
2. Material and methods:.....	185
2.1. Chemicals.....	185
2.2. Metagenomic DNA sampling and library construction	187
2.3. Metagenomic library screening.....	187
2.4. Functional characterization of the hit clones.....	188
2.4.1. Determination of optimal reaction conditions.....	188
2.4.2. Enzymatic activity on mediators	188
2.4.3. Degradation of semi-relective layers of sulfonated lignin	189
2.4.4. Dye discolouration assays	189
2.4.4.1. Selective growth on solid media with dyes as sole carbon source	189
2.4.4.2. Discolouration halos on solid media	189
2.4.4.3. Discolouration of liquid reaction media	189
2.5. Metagenomic sequence Analysis	190
3. Results & discussion	190
3.1. Metagenomic library screening.....	190
3.2. Red-ox activity characterisation	191
3.2.1. Determination of optimal reaction conditions.....	191
3.2.2. Enzymatic activity on model substrates.....	194
3.2.3. Lignin-derivative depolymerization.....	195
3.2.4. Application for dye elimination.....	197
3.3. Sequence analysis.....	200
3.3.1. Functional annotation	200
3.3.2. Taxonomic assignation and sequence prevalence in the bovine ruminal microbiome	204
4. Conclusion	204
Acknowledgments	205
Artwork.....	205

Table content	206
Supplementary data	206
References.....	210
Chapter VI: Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown	217
Abstract	220
1. Introduction.....	220
2. Results and discussion.....	222
2.1. Primary high-throughput screening of the metagenomic libraries.....	222
2.2. Discrimination screening	226
2.3. Synergic activities of PCW degradation.....	230
2.4. Sequencing and data analysis.....	234
2.4.1. Taxonomic assignation	235
2.4.2. Functional annotation	237
3. Materials and methods	239
3.1. Chemicals.....	239
3.2. Biomass preparation	239
3.3. Metagenomic DNA sampling and library construction	239
3.4. Primary high-throughput screening of the metagenomic libraries.....	240
3.4.1. Esterase/lipase activity.....	240
3.4.2. Carbohydrate degrading activity	240
3.4.3. Oxidoreductase activity.....	241
3.5. Discrimination screening	241
3.5.1. Esterase/lipase activity.....	241
3.5.2. Polysaccharide degrading activity	242
3.5.3. Oxidoreductase activity.....	243
3.6. Sequencing and data analysis.....	244
4. Conclusion:	244
Acknowledgments	245
Artwork.....	245
Table content	245
Supplementary data	245
References:.....	297

Conclusion and perspectives	301
1. Discovery of new bacterial oxidoreductases.....	304
1.1. Set up of novel screening assays	304
1.2. Properties of the new ruminal oxidoreductases	306
2. Ruminal esterases as bioremediation tools	308
3. Effect of enrichment on the functional profile of the rumen microbiome.....	310
Artwork and Table contents	315
1. Artwork content	317
2. Table content.....	318
Résumé en français	321
1. Découverte de nouvelles oxydoréductases bactériennes.....	326
1.1. Mise en place d'un nouveau test de criblage.....	326
1.2. Propriétés des nouvelles oxydoréductases du rumen	327
2. Des estérases issues du rumen comme outils de bioremédiation.....	330
3. Effet de l'enrichissement sur le profil fonctionnel du microbiome du rumen.....	332
Annex I: Functional metagenomics: construction and high-throughput screening of fosmid libraries for discovery of novel carbohydrate-active enzymes	337
Annex II: Inventory of the GH70 enzymes encoded by <i>Leuconostoc citreum</i> NRRL B-1299 – identification of three novel α-transglucosylases	355

Abbreviations

A: amaranth
ABTS: 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)
AF: Acid fuschin
AMD: acid mine drainage
AZCL-HEC : AZCL-hydroxy-ethyl-cellulose
Azo-CMC: AZO-CarboxyMethyl-Cellulose
CAT : Carbohydrate-active-enzymes Analysis Toolkit
CAZy: Carbohydrate-Active enZYmes
CBR3BA: cibracon brillant red 3BA
CDD: Conserved Domain Database
COG/KOG: Clusters of Orthologous Group/ euKaryotic Orthologous Groups
DBP: dibutyl phthalate
DGGE: denaturing gradient gel electrophoresis
eggNOG: evolutionary genealogy of genes: Non-supervised Orthologous Groups
EMBL-EBI: European Molecular Biology Laboratory - European Bioinformatics Institute
GC-MS: gas chromatography-mass spectrometry
GOLD: Genomes Online Database
HC: hydrocarbon
HMM: Hidden Markov Models
HPLC: high performance liquid chromatography
HPSEC: High Performance Size Exclusion Chromatography
IC: indigo carmine
IMG: Integrated Microbial Genomes
IVTE: in vitro enriched
IVVE: in vivo enriched
KEGG: Kyoto Encyclopedia of Genes and Genomes
LA: lignin alkali
LccED : Laccase Engineering Database
LED : Lipase Engineering Database
MALDI-TOF: Matrix-Assisted Laser Desorption/Ionisation-time-of-flight
MetaBioME : Metagenomic BioMining Engine
MG: malachite green
MG-RAST : Metagenomic Rapid Annotations using Subsystems Technology
PAH: polycyclic aromatic hydrocarbons
PBS: poly-butylene succinate
PBSA: poly-butylene succinate-co-adipate
PCB: polychlorinated biphenyl
PCL: polycaprolactone
PCR : Polymerase Chain Reaction
PCR: polymerase chain reaction
PCW: plant cell wall

PES: poly-ethylene succinate

PET: polyethylene terephthalate

Pfam: Protein families

PHB: polyhydroxybutyrate

PLA: polylactic acid

POP: persistent organic pollutant

PU: polyurethane

PUL: Polysaccharide Utilization Loci

RB5: reactive black 5

RBBR: remazol brilliant blue R

RO16: reactive orange 16

RPS BLAST : Reversed Position Specific BLAST

SIP: stable isotope probing

TCP: 3,5,6-trichloro-2-pyridinol

ThYme : Thioester-active enZYme

TIGRFAM: The Institute for Genomic Research's database of protein families

TO: tropaeolin O

General introduction

The large majority of Earth's ecosystems is composed of uncultured microorganisms. In many of them, up to 99% of the microorganisms indeed cannot be cultured using standard laboratory techniques (Kato et al., 2015). Because of this, culture-independent methods, like metagenomics, have become essential to assess the genomic information of an entire microbial community (also called the 'metagenome'), and to broaden our knowledge on the population structure, its functional diversity and ecological roles. This allowed the discovery of numerous useful molecules, among which are the enzymes, which are key tools for many biotechnological applications.

Enzymes are indeed the living world's major catalysts, and are essential for most biological processes. In industry, the exquisite properties of enzymes (catalytic efficiency, regio- and stereo selectivity) have been recognized and exploited in several manufacturing processes for many years (Li et al., 2012). However, the greening of conventional chemical processes, replacing unsustainable manufacturing processes by more sustainable enzyme-driven methods is still challenging, together with industrial pollutant elimination, and continuously require more efficient, specific and stable enzymes.

This is in particular the case for biorefineries that have been developed for many years in order to break down plant biomass and to convert it into biofuels and biosourced synthons and materials (Menon and Rao, 2012). But enzymatic degradation of lignocelluloses, which constitute the main renewable carbon source without being in competition with the food industry, is still a key issue. Disrupting the plant cell wall network to expose cellulose and hemicelluloses for enhanced saccharification and fermentation indeed requires a battery of synergistic carbohydrate active enzymes (glycoside-hydrolases and esterases) and efficient oxidoreductases to break lignin down. There is actually an urgent need to design efficient enzymatic cocktails, and to accelerate the discovery of oxidoreductases, of which the natural diversity is largely underexploited.

Oxidoreductases, with other enzymes like esterases, are also essential biocatalysts for the degradation of many industrial pollutants. Indeed, tons of synthetic polyaromatic compounds, polyesters and phytosanitary products, are leaked in the environment every year (Rieger et al., 2002). Most of their biodegradation routes are still unknown, and thus cannot be exploited at the industrial scale. Both for biorefineries and bioremediation, it is thus essential to expand the catalog of industrial enzymes. Microbial ecosystems are inexhaustible sources of novel biocatalysts that are still hidden in the uncultivated microbial fractions. In order to accelerate their discovery, the challenge is to develop fancy high-throughput screening strategies to explore the extraordinary large functional diversity of complex microbiomes.

The bovine rumen ecosystem is a goldmine to find such biocatalysts. It is indeed one of the most diverse gut microbiota, naturally geared towards the degradation of dietary lignocelluloses. This ecosystem has of course already been mined for cellulases and

hemicellulases in the past few years (Morgavi et al., 2013). However, no novel carbohydrate active enzyme family was discovered until now by using metagenomics, and the discovered enzymes were never tested on real plant cell wall matrixes. In addition, ruminal enzymes have rarely been exploited for pollutant degradation, while enzyme promiscuity often allows them to act on their natural physiologic substrates, but also on synthetic structural analogues.

As such, in this thesis work, we used activity-based functional metagenomics to mine the ruminal microbiome to discover novel glycoside-hydrolases, esterases and oxidoreductases of biotechnological interest. The targeted microbiome was preliminary artificially enriched in such functions, by feeding the ecosystem *in vitro*, and *in vivo*, with wheat straw, an agricultural waste particularly rich in lignin.

The resulting bacterial consortia were then screened in order to:

- exploit their functional potential to find new carbohydrate active enzymes and oxidoreductases for biorefineries. To this aim, we had to develop a multi-steps screening strategy to isolate new bacterial oxidases, and to guide the design of synergistic lignocellulase cocktails active on crude plant biomass.
- compare the effect of various enrichment methods on microbial diversity.
- evaluate the potential of the new oxidases for the degradation of polyaromatic pollutants found in industrial wastewater, especially textile dyes.
- isolate new esterases able to degrade structurally different carbamates, which are major material pollutant and phytosanitary products.

This manuscript is organized in six chapters, the first three constituting a literature review:

The first chapter is a review published in 2015 in *Frontiers in Microbiology*, which gives an overview of the meta-omic strategies that have been developed in this last decade to discover new enzymes.

The second chapter is a review paper published in 2015 in *Biotechnology Advances*, which gives an overview of the enzyme families of interest for pollutant degradation, and their members discovered thanks to functional metagenomics. A list of the different studies, together with their methodological specificities, is presented in two synthetic tables.

The third chapter is an overview of the bovine rumen ecosystem, the microbiome targeted in this thesis. It deals in particular with the diet-microbiota relationships, with a specific focus on the different enzyme classes known to play a major role in the functional potential of this ecosystem.

The next three chapters relate to the results obtained during this thesis. They are presented as research articles that will be submitted in the next weeks:

The fourth chapter deals with the discovery of novel esterases from the bovine rumen microbiome, and with the characterization of their catalytic specificities towards two carbamate pollutants, namely Impranil DLN, a commercial polyurethane, and fenobucarb, a carbamate insecticide.

The fifth chapter describes the development of a functional metagenomic methodology to screen microbiomes for novel oxidases acting on natural and synthetic polyaromatic compounds. This approach, applied to the bovine rumen microbiome, allowed us to discover three metagenomic clones presenting large substrate flexibility. The bacterial enzymes produced, which present low homologies with already characterized enzymes, are indeed efficient for the discoloration of structurally different dyes, but also to degrade lignin derivatives obtained in the paper industries, like kraft lignin.

The sixth chapter compares two different strategies of microbiome enrichment in lignocellulotic activities. In particular, it deals with the functional potential of two bacterial consortia, and with the efficiency of their mining for the discovery of novel lignocellulases, including some belonging to novel enzymatic families, acting synergistically for crude plant biomass breakdown.

A general conclusion, focusing on the major progress obtained from this thesis work, along with the perspectives that can be inferred, concludes this manuscript.

References

- Kato, H., Mori, H., Maruyama, F., Toyoda, A., Oshima, K., Endo, R., Fuchu, G., Miyakoshi, M., Dozono, A., Ohtsubo, Y., Nagata, Y., Hattori, M., Fujiyama, A., Kurokawa, K., Tsuda, M., 2015. Time-series metagenomic analysis reveals robustness of soil microbiome against chemical disturbance. *DNA Res.* 22, 413–424. doi:10.1093/dnares/dsv023
- Li, S., Yang, X., Yang, S., Zhu, M., Wang, X., 2012. TECHNOLOGY PROSPECTING ON ENZYMES: APPLICATION, MARKETING AND ENGINEERING. *Comput. Struct. Biotechnol. J.* 2, 1–11. doi:10.5936/csbj.201209017
- Menon, V., Rao, M., 2012. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. *Prog. Energy Combust. Sci.* 38, 522–550. doi:10.1016/j.pecs.2012.02.002
- Morgavi, D.P., Kelly, W.J., Janssen, P.H., Attwood, G.T., 2013. Rumen microbial (meta)genomics and its application to ruminant production. *animal* 7, 184–201. doi:10.1017/S1751731112000419
- Rieger, P.G., Meier, H.M., Gerle, M., Vogt, U., Groth, T., Knackmuss, H.J., 2002. Xenobiotics in the environment: present and future strategies to obviate the problem of biological persistence. *J. Biotechnol.* 94, 101–123.

Chapter I:

Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology

Lisa Ufarté^{1,2,3}, Gabrielle Potocki-Veronese^{1,2,3} and Élisabeth Laville^{1,2,3*}

¹ Université de Toulouse, Institut National des Sciences Appliquées (INSA), Université Paul Sabatier (UPS), Institut National Polytechnique (INP), Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP), Toulouse, France,

² INRA - UMR792 Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France,

³ CNRS, UMR5504, Toulouse, France

Frontiers in Microbiology (2015)

Abstract:

The rapid expansion of new sequencing technologies has enabled large-scale functional exploration of numerous microbial ecosystems, by establishing catalogs of functional genes and by comparing their prevalence in various microbiota. However, sequence similarity does not necessarily reflect functional conservation, since just a few modifications in a gene sequence can have a strong impact on the activity and the specificity of the corresponding enzyme or the recognition for a sensor. Similarly, some microorganisms harbor certain identified functions yet do not have the expected related genes in their genome. Finally, there are simply too many protein families whose function is not yet known, even though they are highly abundant in certain ecosystems. In this context, the discovery of new protein functions, using either sequence-based or activity-based approaches, is of crucial importance for the discovery of new enzymes and for improving the quality of annotation in public databases. This paper lists and explores the latest advances in this field, along with the challenges to be addressed, particularly where microfluidic technologies are concerned.

Keywords: metagenomics, discovery of new functions, proteins, high throughput screening, microbial ecosystems, microbial ecology, biotechnologies

1. Introduction

The implications of the discovery of new protein functions are numerous, from both cognitive and applicative points of view. Firstly, it improves understanding of how microbial ecosystems function, in order to identify biomarkers and levers that will help optimize the services rendered, regardless of the field of application. Next, the discovery of new enzymes and transporters enables expansion of the catalog of functions available for metabolic pathway engineering and synthetic biology. Finally, the identification and characterization of new protein families, whose functions, three-dimensional structure and catalytic mechanism have never been described, furthers understanding of the protein structure/function relationship. This is an essential prerequisite if we are to draw full benefit from these proteins, both for medical applications (for example, designing specific inhibitors) and for relevant integration into biotechnological processes.

Many reviews have been published on functional metagenomics these last 10 years. Many of them focus on the strategies of library creation and on bio-informatic developments (Di Bella et al., 2013; Ladoukakis et al., 2014), while others describe the various approaches set up to discover novel targets [like therapeutic molecules (Culligan et al., 2014)] for a specific application. In particular several review papers have been written on the numerous activity-based metagenomics studies carried out to find new enzymes for biotechnological

applications, without necessarily finding new functions or new protein families (Ferrer et al., 2009; Steele et al., 2009). The present review focuses on all the functional metagenomics approaches, sequence- or activity-based, allowing the discovery of new functions and families from the uncultured fraction of microbial ecosystems, and makes a recent overview on the advances of microfluidics for ultra-fast microbial screening of metagenomes.

2. Sampling Strategies

The literature describes a wide variety of microbial environments sampled in the search for new enzymes. A large number of studies look at ecosystems with high taxonomic and functional diversity, such as soils or natural aquatic environments that are either undisturbed or exposed to various pollutants (Gilbert et al., 2008; Brennerova et al., 2009; Zanaroli et al., 2010). Extreme environments enable the discovery of enzymes that are naturally adapted to the constraints of certain industrial processes, such as glycoside hydrolases and halotolerant esterases (Ferrer et al., 2005; LeCleir et al., 2007), thermostable lipases (Tirawongsaroj et al., 2008), or even psychrophilic DNA-polymerases (Simon et al., 2009). Other microbial ecosystems, such as anaerobic digesters including both human and/or animal intestinal microbiota and industrial remediation reactors, are naturally specialized in metabolizing certain substrates. These are ideal targets for research into particular functions, such as the degrading activity of lignocellulosic plant biomass (Warnecke et al., 2007; Tasse et al., 2010; Hess et al., 2011; Bastien et al., 2013) or dioxygenases for the degradation of aromatic compounds (Suenaga et al., 2007).

Some studies refer to enrichment steps that occur before sampling, with the aim of increasing the relative abundance of micro-organisms that have the target function. This enrichment can be done by modifying the physical and chemical conditions of the natural environment (van Elsas et al., 2008) or by incorporating the substrate to be metabolized *in vivo* (Hess et al., 2011) or *in vitro*, in reactors (DeAngelis et al., 2010) or mesocosms (Jacquiod et al., 2013). Through stable isotopic probing and cloning of the DNA of micro-organisms able to metabolize a specifically labeled substrate for the creation of metagenome libraries, it is possible to increase the frequency of positive clones by several orders of magnitude (Chen and Murrell, 2010). These approaches require functional and taxonomic controls at the different stages of enrichment, which are often sequential, to prevent the proliferation of populations dependent on the activity of the populations preferred at the outset. These kinds of checks are difficult to do *in vivo*, where there would actually be an increased risk of selecting populations able to metabolize only the degradation products of the initial substrate, to the detriment of those able to attack the more resistant original substrate with its more complex structure.

3. Functional Screening: New Challenges for the Discovery of Functions

Two complementary approaches can be used to discover new functions and protein families within microbial communities. The first involves the analysis of nucleotide, ribonucleotide or protein sequences, and the other the direct screening of functions before sequencing (Figure 1).

3.1. The Sequence, Marker of Originality

There have been a number of large-scale random metagenome sequencing projects (Yooseph et al., 2007; Vogel et al., 2009; Gilbert et al., 2010; Bork et al., with Qin et al., MetaHIT Consortium, 2010; Hess et al., 2011) over the past few years, resulting in catalogs listing millions of genes from different ecosystems, the majority of which are recorded in the GOLD¹ (RRID: nif-0000-02918), MG-RAST² (RRID: OMICS_01456) and EMBL-EBI³ (RRID: nlx_72386) metagenomics databases. At the same time, the obstacles inherent to metatranscriptomic sampling (fragility of mRNA, difficulty with extraction from natural environments, separation of other types of RNA) have been removed, opening a window into the functional dynamics of ecosystems according to biotic or abiotic constraints (Saleh-Lakha et al., 2005; Warnecke and Hess, 2009; Schmieder et al., 2012). Metatranscriptomes sequencing has thus enabled the identification of new gene families, such as those found in microbial communities (prokaryotes and/or eukaryotes) expressed specifically in response to variations in the environment (Bailly et al., 2007; Frias-Lopez et al., 2008; Gilbert et al., 2008) and new enzyme sequences belonging to known carbohydrate active enzymes families (Poretzky et al., 2005; Tartar et al., 2009; Damon et al., 2012).

Regardless of the origin of the sequences (DNA or cDNA, with or without prior cloning in an expression host), the advances made with automatic annotation, most notably thanks to the IMG-M (RRID: nif-0000-03010) and MG-RAST (RRID: OMICS_01456) servers (Markowitz et al., 2007; Meyer et al., 2008), now make it possible to quantify and compare the abundance of the main functional families in the target ecosystems (Thomas et al., 2012), identified through comparison of sequences with the general functional databases: KEGG (RRID: nif-0000-21234) (Kanehisa and Goto, 2000), eggNOG (RRID: nif-0000-02789) (Muller et al., 2010), and COG/KOG (RRID: nif-0000-21313) (Tatusov et al., 2003). They also enable research into specific protein families, thanks to motif detection using Pfam (RRID: nlx_72111) (Finn et al., 2010), TIGRFAM (RRID: nif-0000-03560) (Selengut et al., 2007), CDD (RRID: nif-0000-02647) (Marchler-Bauer et al., 2009), Prosite (RRID: nif-0000-03351) (Sigris

¹ <http://www.genomesonline.org/cgi-bin/GOLD/index>

² <http://metagenomics.anl.gov/>

³ <http://www.ebi.ac.uk/metagenomics>

et al., 2010), and HMM model construction (Hidden Markov Models; Soding, 2005).

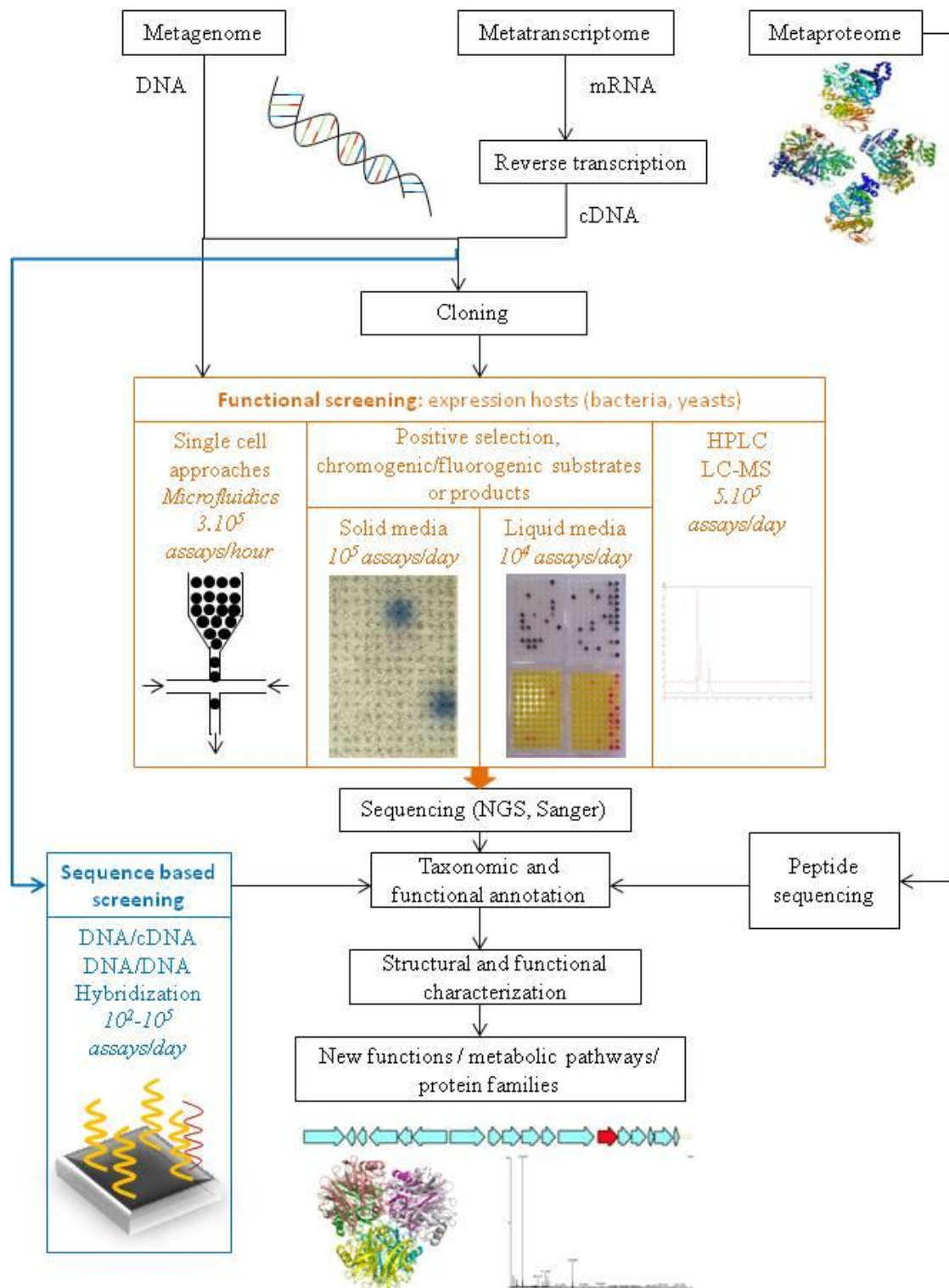


Figure 1: Strategies for the functional exploration of metagenomes, metatranscriptomes and metaproteomics to discover new functions and protein families.

Other servers can be used to interrogate databases specialized in specific enzymatic families (Table 1).

Table 1: Examples of databases specialized in enzymatic functions of biotechnological interest.

Databases	Enzymes	References
MetaBioME	Enzymes of industrial interest	Sharma et al., 2010
CAZy (RRID:OMICS_01677)	carbohydrate active enzymes	Cantarel et al., 2012
	Auxiliary redox enzymes for lignocellulose degradation	Levasseur et al., 2013
CAT (RRID:OMICS_01676)	carbohydrate active enzymes	Park et al., 2010
LccED	Laccases	Sirim et al., 2011
LED (RRID:nif-0000-03084)	Lipases	Pleiss et al., 2000
MEROPS (RRID:nif-0000-03112)	Proteases	Rawlings et al., 2012
ThYme	Thioesterases	Cantu et al., 2011

Finally, the performance of methods used to assemble next generation sequencing reads is set to open up access to a plethora of complete genes to feed expert databases, which currently only contain a tiny percentage of genes from uncultivated organisms—less than 1% for the CAZy database (RRID: OMICS_01677), for example—while the majority of metagenomic studies published target ecosystems with a high number of plant polysaccharide degradation activities by carbohydrate active enzymes (André et al., 2014). Even based on a large majority of truncated genes, metagenomes and metatranscriptomes functional annotation enables *in silico* estimations of the functional diversity of the ecosystem and identification of the most original sequences within a known protein family. It is then possible to use PCR (Polymerase Chain Reaction) to capture those sequences specifically, and test their function experimentally to assess their applicative value. In this way, the sequencing of the rumen metagenome (268 Gb) enabled identification of 27,755 coding genes for carbohydrate active enzymes, and isolation of 51 active enzymes belonging to known families specifically involved in lignocelluloses degradation (Hess et al., 2011).

PCR, and more generally DNA/DNA or DNA/cDNA hybridization, also make it possible to directly capture coding genes for protein families that are abundant and/or expressed in the target ecosystem, but with no need for *a priori* large-scale sequencing. This strategy requires the conception of nucleic acid probes or PCR primers using consensus sequences specific to known protein families. There are plenty of examples of the discovery of enzymes in metagenomes using these approaches, for instance bacterial laccases (Ausec et al., 2011), dioxygenases (Zaprasis et al., 2009), nitrites reductases (Bartossek et al., 2010), hydrogenases (Schmidt et al., 2010), hydrazine oxidoreductases (Li et al., 2010), or chitinases

(Hjort et al., 2010) from various ecosystems. The Gene-Targeted-metagenomics approach (Iwai et al., 2009) combines PCR screening and amplicon pyrosequencing to generate primers in an iterative manner and increase the structural diversity of the target protein families, for example the dioxygenases from the microbiota of contaminated soil. Elsewhere, the use of high-density functional microarrays considerably multiplies the number of probes and is therefore a low-cost way of obtaining a snapshot of the abundance and diversity of sequences within specific protein families and even, where the DNA or cDNA has been cloned (He et al., 2010; Weckx et al., 2010), directly capturing targets of interest while rationalizing sequencing. Using a similar strategy, the solution hybrid selection method enables the selection of fragments of coding DNA for specific enzymatic families using 31-mers capture probes. Applied to the capture of cDNA, this method provides access to entire genes which can be then cloned and their activity tested (Bragalini et al., 2014). Solution hybrid selection can therefore be used to explore the taxonomic and functional diversity of all protein families. More especially, this approach opens the way for the selection and characterization of families that are highly represented in a microbiome but whose function remains unknown, in order to further the understanding of ecosystemic functions and discover novel biocatalysts.

Metaproteomics has recently proved its worth in identifying new protein families and/or functions. Paired with genomic, metagenomic and metatranscriptomic data (Erickson et al., 2012), it provides access to excellent biomarkers of the functional state of the ecosystem. Recent developments, such as high-throughput electrospray ionization paired with mass spectrometry, enable full metaproteome analysis after separation of proteins by liquid chromatography. It is thus possible to highlight hundreds of proteins with no associated function and new enzyme families playing a key functional role in the ecosystem (Ram et al., 2005).

This latter example illustrates the need for research and/or experimental proof of function for proteins where the function remains unknown (products of orphan genes or, on the contrary, genes highly prevalent in the microbial realm but that have never been characterized) or poorly annotated. In fact, annotation errors, which are especially common for multi-modular proteins such as carbohydrate active enzymes, are spread at an increasing rate as a result of the explosion in the number of functional genomics and meta-genomic, -transcriptomic and -proteomic projects. New annotation strategies, most notably based on the prediction of the three-dimensional structure of proteins, are also worth exploring (Uchiyama and Miyazaki, 2009). However, at the present time, it is very difficult to predict the specificity of substrate and the mechanism of action (and therefore the function of the protein) on the basis of sequence or even structure, especially where there is no homologue characterized from a structural and functional point of view. Functional screening can address this challenge.

3.2. Activity Screening: Speeding up the Discovery of Biotechnology Tools

There are three prerequisites for this approach: (i) the cloning of DNA or cDNA in an expression vector for the creation of, respectively, metagenomic or metatranscriptomic libraries, (ii) heterologous expression of cloned genes in a microbial host, (iii) the conception of efficient phenotypic screens to isolate the clones of interest that produce the target activity, also referred to as “hits.”

Using this approach, the functions of a protein can be accessed without any prior information on its sequence. It is therefore the only way of identifying novel protein families that have known functions or previously unseen functions (as long as an adequate screen can be developed). Finally, it helps to rationalize sequencing efforts and focus them only on the hits: for example, those that are of biotechnological interest. The expression potential of the selected heterologous host, the size of the DNA inserts and the type of vectors all determine the success of functional screening. Short fragments of metagenomic DNA (smaller than 15 kb, and most often between 2 and 5 kb), or cDNA for the metatranscriptomic libraries, cloned in plasmids under the influence of a strong expression promoter, enable the overexpression of a single protein, and the easy recovery and sequencing of the hits' DNA (Uchiyama and Miyazaki, 2009). On the other hand, fragments of bacterial DNA measuring between 15 and 40 kb, 25 and 45 kb or even 100 and 200 kb, cloned respectively in cosmids, fosmids or bacterial artificial chromosomes, can be used to explore a functional diversity of several Gb per library and, above all, provide access to operon-type multigene clusters, coding for complete catabolic or anabolic pathways. This is of major interest for the discovery of cocktails of synergistic activities that degrade complex substrates such as plant cell walls for biorefineries. This strategy also ensures high reliability for the taxonomic annotation of inserts, and can even be used to identify the mobile elements responsible for the plasticity of the bacterial metagenome, mediated by horizontal gene transfers (Tasse et al., 2010). However, it requires sensitive activity screens, since the target genes are only weakly expressed, controlled by their own native promoters.

Escherichia coli, whose transformation efficiency is exceptionally high, even for fosmids or bacterial artificial chromosomes, remains the host of choice in the immense majority of studies published. The first exhaustive functional screening study of a fosmid library revealed that *E. coli* can be used to express genes from bacteria that are very different from a taxonomical point of view, including a large number of Bacteroidetes and Gram-positive bacteria (Tasse et al., 2010), contrary to what had been predicted by *in silico* detection of expression signals compatible with *E. coli* (Gabor et al., 2004). However, the value of developing shuttle vectors to screen metagenomic libraries in hosts with different expression and secretion potentials, for example *Bacillus*, *Sphingomonas*, *Streptomyces*, *Thermus*, or the α -, β - and γ -proteobacteria (Taupp et al., 2011; Ekkers et al., 2012) must not

be underestimated, if we are to unlock the functional potential of varied taxons and increase the sensitivity of screens. Finally, it is still very difficult to get access to the uncultivated fraction of eukaryotic microorganisms, due to the lack of screening hosts with sufficient transformation efficiency for the creation of large clone libraries (and thus the exploration of a vast array of sequences) and compatible with the post-translational modifications required to obtain functional recombinant proteins from eukaryotes. Thus, at the present time, only a few studies have been published on the enzyme activity-based screening of metatranscriptomic libraries (making it possible to do away with introns) of eukaryotes from soil, rumen and the gut of the termite (Bailly et al., 2007; Findley et al., 2011; Sethi et al., 2013).

Regardless of the type of library screened, the functional exploration of hundreds of thousands of clones is required, whereas the hit rate rarely exceeds 6‰ (Duan et al., 2009; Bastien et al., 2013). This requires very high throughput primary screens, in a solid medium before or after the automated organization of libraries in 96- or 384-well micro-plate format, in a liquid medium after enzymatic cell lysis and/or thawing and freezing (Bao et al., 2011), or using UV-inducible auto-lytic vectors (Li et al., 2007). This stage is very often followed by medium or low throughput characterization of the properties of the hits obtained, particularly to assess their biotechnological interest (Tasse et al., 2010).

Two generic strategies, used at throughputs exceeding 400,000 tests per week, have been and continue to be applied widely. Positive selection on a medium containing, for example, substrates to be metabolized as the sole source of carbon, can be used to isolate enzymes (Henne et al., 1999), complete catabolic pathways (Cecchini et al., 2013), or membrane transporters (Majernik et al., 2001). This approach also helps easily identify antibiotic resistant genes (Diaz-Torres et al., 2006). The use of chromogenic (Beloqui et al., 2010; Bastien et al., 2013; Nyssönen et al., 2013), fluorescent (LeClerc et al., 2007), or opalescent substrates or reagents, such as insoluble polymers or proteins (Mayumi et al., 2008; Waschowitz et al., 2009), or simply the observation of an original clone phenotype, has already enabled the isolation of several 100 catabolic enzymes, like the numerous hydrolases of very varied taxonomic origin (Simon and Daniel, 2009), some of which were coded by genes that are very abundant in the target ecosystem (Jones et al., 2008; Gloux et al., 2011), but also, although much less frequently, new oxidoreductases (Knietsch et al., 2003). Novel enzymes (laccases, esterases and oxygenases in particular) from microbial communities of very diverse origins (soil, water, activated sludge, digestive tracts) have been highlighted for their capacity to degrade pollutants such as nitriles (Robertson and Steer, 2004), lindane (Boubakri et al., 2006), styrene (van Hellemond et al., 2007), naphthalene (Ono et al., 2007), aliphatic and aromatic carbohydrates (Uchiyama et al., 2004; Brennerova et al., 2009; Lu et al., 2012), organophosphorus (Kambiranda et al., 2009; Math et al., 2010), or plastic materials (Mayumi et al., 2008).

The discovery of proteins involved in prokaryote-eukaryote interactions (Lakhdari et al., 2010) or anabolic pathways is rarer, since it often requires the development of complex screens and lower throughputs. Nonetheless, a few examples of simple screens, based on the aptitude of metagenomic clones to inhibit the growth of a strain by producing antibacterial activity or to complement an auxotrophic strain for a specific compound, have enabled the identification of new pathways for the synthesis of antimicrobials (Brady and Clardy, 2004) or biotin (Entcheva et al., 2001). Nano-technologies, and in particular the latest developments focused on the medium-throughput screening of libraries obtained by combinatorial protein engineering, enable the design of custom microarrays and covered with one to several 100 specific enzymatic substrates, the processing of which may be followed by fluorescence, chemiluminescence, immunodetection, surface plasmon resonance or mass spectrometry (André et al., 2014). Nanostructure-initiator mass spectrometry technology, combining fluorescence and mass spectrometry, is the first example of a functional metagenomic application for the discovery of anabolic enzymes, namely sialyltransferases (Northen et al., 2008).

3.3. The Immense Challenges of Ultra-fast Screening (Figure 2)

Microfluidic technologies are of undeniable interest when it comes to reaching screening rates of a million clones per day. The substrate induced gene-expression screening method has been developed to use fluorescence-activated cell sorting to isolate plasmidic clones containing genes (or fragments of genes) that induce the expression of a fluorescent marker in response to a specific substrate. However, this technique is only suited to small substrates that are non-lethal and internalizable for the host strain (Uchiyama and Watanabe, 2008). Finally, the advances made over the past few years in cellular compartmentalization (Nawy, 2013), selective sorting, based on sequence detection (Pivetal et al., 2014; Lim et al., 2015) or specific metabolites (Kürsten et al., 2014) and the control of reaction kinetics (Mazutis et al., 2009) in microfluidic circuits should allow for a huge acceleration in the discovery of new proteins and metabolic pathways expressed in prokaryotes and eukaryotes in an intercellular, membrane or extracellular manner.

The very first examples of metagenome functional exploration applications have already been used to establish the proof of concept regarding the effectiveness of microfluidics in the discovery of new bioactive molecules and new enzymes. For example, droplet-based microfluidics technology was recently used by the teams of A. Griffiths and A. Dreville to isolate new strains producing cellobiohydrolase and cellulase activities at a rate of 300,000 cells sorted per hour, using just a few microliters of reagent, i.e., 250,000 times less than with the conventional technologies mentioned above (Najah et al., 2014). Here, soil bacteria and a fluorescent substrate were co-encapsulated in micro-droplets in order to sort cells on the basis of the extracellular activity only. In fact, the strategy used, which requires the seeding of cells on a defined medium after sorting, is not compatible with the detection

of intracellular enzymes, which require a lethal lysis step to convert the substrate. Applying a similar principle, the ultra-rapid sorting of eukaryote cells encapsulated with their substrate now also makes it possible to select yeast clones presenting extracellular enzymatic activities (Sjostrom et al., 2014). This technology should, in the short term, make it possible to explore the functional diversity of uncultivated eukaryotes at a very high throughput, by directly sorting fungal populations or libraries of metatranscriptomic clones. In the latter case, access to the sequence involved in the target activity will be easy, since the libraries are built using hosts whose culture is well managed, with insertion of the metatranscriptomic cDNA fragment into a specific region of the genome. Where sorting is done without cloning of the metagenome or metatranscriptome, only microorganisms capable of growth on a defined medium can be recovered, which hugely limits access to functional diversity.

To increase the proportion of cultivable organisms, Kim Lewis' team recently used the iChip to simultaneously isolate and cultivate soil bacteria thanks to the delivery of nutrients from the original medium, into which the iChip is introduced, via semi-permeable membranes. This method enables an increase in cultivable organisms ranging from 1 to 50%. Using colonies cultivated in the chip, the clones isolated in a Petri dish were screened for the production of antimicrobial compounds (Ling et al., 2015). A novel antibiotic was thus identified, together with its biosynthesis pathway, after sequencing and functional annotation of the complete genome.

It is quite another matter when it comes to selecting, on the basis of intracellular activity, completely uncultivable organisms or metagenomic clones containing DNA inserts of several dozen kbp, which are difficult to amplify using PCR. In this case, to liberate the enzymes in question, we are required to include a cellular lysis step, preventing seeding after sorting. On the other hand, this approach is compatible with the sorting of plasmid clone libraries, where the metagenomic or metatranscriptomic inserts can easily be amplified using PCR, on the basis of just a few dozen lysed cells. For libraries with large DNA inserts, the barriers are now being broken down, most notably thanks to the development of the SlipChips microfluidic approach (Ma et al., 2014), which uses two culture microcompartments, where the content of one can be lysed for the detection of enzymatic activities, for example, and the other is used as a backup replicate for the culture and recovery of subsequent DNA for sequencing. In spite of these recent, highly encouraging developments, the proof of concept has not yet been established for the identification of new functions and intracellular metabolic pathways.

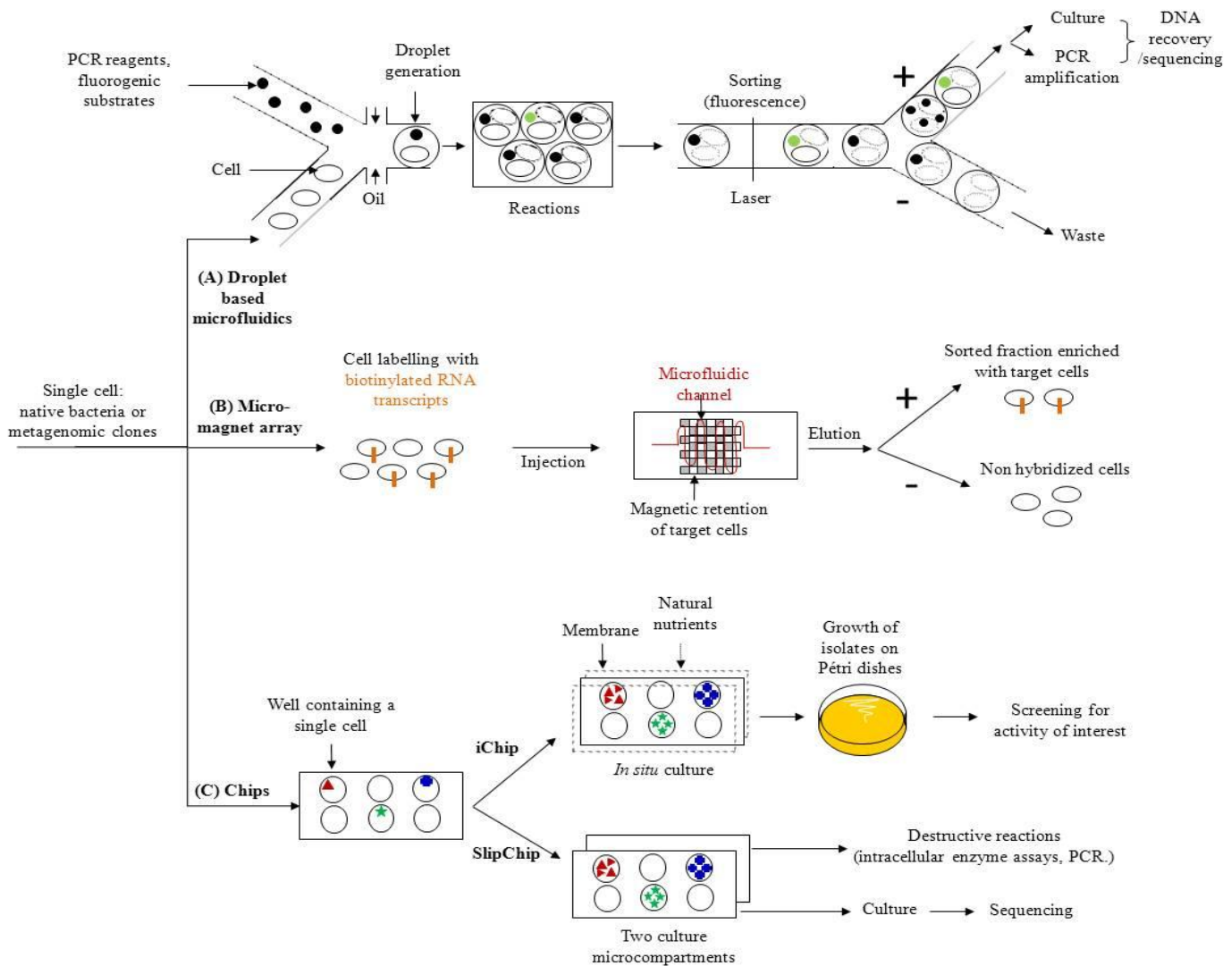


Figure 2 : Microfluidic strategies for new enzyme screening. (A) Droplet based microfluidics: single cells are encapsulated with probes or fluorogenic substrates to create microdroplets, where reactions happen (substrate degradation, PCR). The hits are sorted using fluorescence detection. Non-lysed cells are cultured and DNA fragments from lysed cells are amplified. Both methods allow the recovery and sequencing of DNA. (B) Micro-magnet array: target cells are labeled with biotinylated RNA transcripts probes and injected inside the microchannel. Target cells are captured in the channel thanks to magnetic forces while non-targets cells pass through the device. (C) Chips: the chip wells are filled with a single cell. The *iChip* is covered by membranes, and reintroduced into original environment, where natural nutrients flow through membranes. Colonies are further isolated on Petri dishes to be screened for the activity of interest. The *SlipChip* is composed of two culture microcompartments which are further separated for destructive and non-destructive assays.

4. Conclusion

The rapid expansion of meta-omic technologies over the past decade has shed light on the functions of the uncultivated fraction of microbial ecosystems. A huge number of enzymes have been discovered, in particular through experimental approaches to functional metagenomes exploration. Where their performance can be rapidly assessed within the framework of a known process, or where they catalyze new, previously undescribed reactions, many of them have provided new tools for industrial biotechnologies. However, several challenges still need to be addressed to speed up the rate at which new functions are discovered and to make optimal use of the functional diversity that so far remains unexplored. Firstly, while the uncultivated prokaryote fraction of microbial communities is still extensively studied, the functions of the eukaryote fraction are relatively unexplored from an experimental angle, even though they play a fundamental role for numerous ecosystems. Secondly, in the majority of cases, the functions discovered using meta-omic approaches play a catabolic role, mainly involved in the deconstruction of plant biomass or in bioremediation. It is thus necessary to develop functional screens to access anabolic functions and enrich the catalog of reactions available for synthetic biology. Finally, there are very few studies aimed at identifying the role of protein families that are highly prevalent in the target ecosystem but that have not yet been characterized, even though some of them could be considered as biomarkers of the functional state of the microbial community. Indeed, sequence-based functional metagenomic projects continuously highlight many sequences annotated as domains of unknown function in the Pfam database (RRID: nlx_72111) (Bateman et al., 2010; Finn et al., 2014), some with 3D structures solved thanks to structural genomics initiatives, and available in the Protein Data Bank (RRID: nif-0000-00135). With the goal of characterizing these new protein families and identifying previously unseen functions from the selection the most prevalent protein families (those containing the highest number of homologous sequences without any associated function) in the target ecosystem, the integration of structural, biochemical, genomic and meta-omic data is now also possible (Ladevèze et al., 2013). It allows to benefit from the huge amount of long scaffolds now available in sequence databases, and to access the genomic context of the targeted genes in order to facilitate functional assignment. In the next few years, these strategies should enhance our understanding of how microbial ecosystems function and, at the same time, enable greater control over them.

Author Contributions

LU, GPV, EL contributed equally to this work.

Acknowledgments

This research was funded by the Ministry of Education and Research (Ministère de l'Enseignement supérieur et de la Recherche, MESR), the Agence Nationale de la Recherche (Grant Number ANR2011-Nano00703) and the INRA metaprogramme M2E (project Metascreen).

Artwork

Figure 1: Strategies for the functional exploration of metagenomes, metatranscriptomes and metaproteomics to discover new functions and protein families. 34

Figure 2 : Microfluidic strategies for new enzyme screening. 41

Table content

Table 1: Examples of databases specialized in enzymatic functions of biotechnological interest..... 35

References

- André, I., Potocki-Véronèse, G., Barbe, S., Moulis, C., Remaud-Siméon, M., 2014. CAZyme discovery and design for sweet dreams. *Curr. Opin. Chem. Biol.* 19, 17–24. doi:10.1016/j.cbpa.2013.11.014
- Ausec, L., van Elsas, J.D., Mandic-Mulec, I., 2011. Two- and three-domain bacterial laccase-like genes are present in drained peat soils. *Soil Biol. Biochem.* 43, 975–983. doi:10.1016/j.soilbio.2011.01.013
- Bailly, J., Fraissinet-Tachet, L., Verner, M.-C., Debaud, J.-C., Lemaire, M., Wésolowski-Louvel, M., Marmeisse, R., 2007. Soil eukaryotic functional diversity, a metatranscriptomic approach. *ISME J.* 1, 632–642. doi:10.1038/ismej.2007.68
- Bao, L., Huang, Q., Chang, L., Zhou, J., Lu, H., 2011. Screening and characterization of a cellulase with endocellulase and exocellulase activity from yak rumen metagenome. *J. Mol. Catal. B Enzym.* 73, 104–110. doi:10.1016/j.molcatb.2011.08.006
- Bartossek, R., Nicol, G.W., Lanzen, A., Klenk, H.-P., Schleper, C., 2010. Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ. Microbiol.* 12, 1075–1088. doi:10.1111/j.1462-2920.2010.02153.x
- Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., Noirot, C., Dumon, C., O'Donohue, M., 2013. Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnol. Biofuels* 6, 78. doi:10.1186/1754-6834-6-78
- Bateman, A., Coghill, P., Finn, R.D., 2010. DUFs: families in search of function. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* 66, 1148–1152. doi:10.1107/S1744309110001685
- Beloqui, A., Polaina, J., Vieites, J.M., Reyes-Duarte, D., Torres, R., Golyshina, O.V., Chernikova, T.N., Waliczek, A., Aharoni, A., Yakimov, M.M., Timmis, K.N., Golyshin, P.N., Ferrer, M., 2010. Novel Hybrid Esterase-Haloacid Dehalogenase Enzyme. *ChemBioChem* 11, 1975–1978. doi:10.1002/cbic.201000258
- Boubakri, H., Beuf, M., Simonet, P., Vogel, T.M., 2006. Development of metagenomic DNA shuffling for the construction of a xenobiotic gene. *Gene* 375, 87–94. doi:10.1016/j.gene.2006.02.027
- Brady, S.F., Clardy, J., 2004. Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water. *J Nat Prod* 67, 1283 – 1286.
- Bragalini, C., Ribiere, C., Parisot, N., Vallon, L., Prudent, E., Peyretailade, E., Girlanda, M., Peyret, P., Marmeisse, R., Luis, P., 2014. Solution Hybrid Selection Capture for the Recovery of Functional Full-Length Eukaryotic cDNAs From Complex Environmental Samples. *DNA Res.* 21, 685–694. doi:10.1093/dnares/dsu030
- Brennerova, M.V., Josefiova, J., Brenner, V., Pieper, D.H., Junca, H., 2009. Metagenomics reveals diversity and abundance of meta -cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging

- bioremediation. *Environ. Microbiol.* 11, 2216–2227. doi:10.1111/j.1462-2920.2009.01943.x
- Cantarel, B.L., Lombard, V., Henrissat, B., 2012. Complex Carbohydrate Utilization by the Healthy Human Microbiome. *PLoS ONE* 7, e28742. doi:10.1371/journal.pone.0028742
- Cantu, D.C., Chen, Y., Lemons, M.L., Reilly, P.J., 2011. ThYme: a database for thioester-active enzymes. *Nucleic Acids Res.* 39, D342–D346. doi:10.1093/nar/gkq1072
- Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré, J., Henrissat, B., Remaud-Siméon, M., Monsan, P., Potocki-Véronèse, G., 2013. Functional Metagenomics Reveals Novel Pathways of Prebiotic Breakdown by Human Gut Bacteria. *PLoS ONE* 8, e72766. doi:10.1371/journal.pone.0072766
- Chen, Y., Murrell, J.C., 2010. When metagenomics meets stable-isotope probing: progress and perspectives. *Trends Microbiol.* 18, 157–163. doi:10.1016/j.tim.2010.02.002
- Culligan, E.P., Sleator, R.D., Marchesi, J.R., Hill, C., 2014. Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. *Virulence* 5, 399–412. doi:10.4161/viru.27208
- Damon, C., Lehembre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L., Marmeisse, R., 2012. Metatranscriptomics Reveals the Diversity of Genes Expressed by Eukaryotes in Forest Soils. *PLoS ONE* 7, e28967. doi:10.1371/journal.pone.0028967
- DeAngelis, K.M., Gladden, J.M., Allgaier, M., D’haeseleer, P., Fortney, J.L., Reddy, A., Hugenholtz, P., Singer, S.W., Vander Gheynst, J.S., Silver, W.L., Simmons, B.A., Hazen, T.C., 2010. Strategies for Enhancing the Effectiveness of Metagenomic-based Enzyme Discovery in Lignocellulolytic Microbial Communities. *BioEnergy Res.* 3, 146–158. doi:10.1007/s12155-010-9089-z
- Diaz-Torres, M.L., Villedieu, A., Hunt, N., McNab, R., Spratt, D.A., Allan, E., Mullany, P., Wilson, M., 2006. Determining the antibiotic resistance potential of the indigenous oral microbiota of humans using a metagenomic approach. *FEMS Microbiol. Lett.* 258, 257–262. doi:10.1111/j.1574-6968.2006.00221.x
- Di Bella, J.M., Bao, Y., Gloor, G.B., Burton, J.P., Reid, G., 2013. High throughput sequencing methods and analysis for microbiome research. *J. Microbiol. Methods* 95, 401–414. doi:10.1016/j.mimet.2013.08.011
- Duan, C.-J., Xian, L., Zhao, G.-C., Feng, Y., Pang, H., Bai, X.-L., Tang, J.-L., Ma, Q.-S., Feng, J.-X., 2009. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J. Appl. Microbiol.* 107, 245–256. doi:10.1111/j.1365-2672.2009.04202.x
- Ekkers, D.M., Cretoiu, M.S., Kielak, A.M., Elsas, J.D. van, 2012. The great screen anomaly—a new frontier in product discovery through functional metagenomics. *Appl. Microbiol. Biotechnol.* 93, 1005–1020. doi:10.1007/s00253-011-3804-3
- Entcheva, P., Liebl, W., Johann, A., Hartsch, T., Streit, W.R., 2001. Direct Cloning from Enrichment Cultures, a Reliable Strategy for Isolation of Complete Operons and

- Genes from Microbial Consortia. *Appl. Environ. Microbiol.* 67, 89–99.
doi:10.1128/AEM.67.1.89-99.2001
- Erickson, A.R., Cantarel, B.L., Lamendella, R., Darzi, Y., Mongodin, E.F., Pan, C., Shah, M., Halfvarson, J., Tysk, C., Henrissat, B., Raes, J., Verberkmoes, N.C., Fraser, C.M., Hettich, R.L., Jansson, J.K., 2012. Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. *PLoS ONE* 7, e49138.
doi:10.1371/journal.pone.0049138
- Ferrer, M., Beloqui, A., Timmis, K.N., Golyshin, P.N., 2009. Metagenomics for Mining New Genetic Resources of Microbial Communities. *J. Mol. Microbiol. Biotechnol.* 16, 109–123. doi:10.1159/000142898
- Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Martins Dos Santos, V.A.P., Yakimov, M.M., Timmis, K.N., Golyshin, P.N., 2005. Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem. Biol.* 12, 895–904.
doi:10.1016/j.chembiol.2005.05.020
- Findley, S.D., Mormile, M.R., Sommer-Hurley, A., Zhang, X.-C., Tipton, P., Arnett, K., Porter, J.H., Kerley, M., Stacey, G., 2011. Activity-Based Metagenomic Screening and Biochemical Characterization of Bovine Ruminant Protozoan Glycoside Hydrolases. *Appl. Environ. Microbiol.* 77, 8106–8113. doi:10.1128/AEM.05925-11
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230.
doi:10.1093/nar/gkt1223
- Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., 2010. The Pfam protein families database. *Nucleic Acids Res.* 38, D211–D222.
doi:10.1093/nar/gkp985
- Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., DeLong, E.F., 2008. Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci.* 105, 3805–3810. doi:10.1073/pnas.0708897105
- Gabor, E.M., Alkema, W.B.L., Janssen, D.B., 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* 6, 879 – 886. doi:10.1111/j.1462-2920.2004.00640.x
- Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., Joint, I., 2008. Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *PLoS ONE* 3, e3042. doi:10.1371/journal.pone.0003042
- Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., Weynberg, K., Huse, S., Hughes, M., Joint, I., Somerfield, P.J., Mühlhling, M., 2010. The Taxonomic and Functional Diversity of Microbes at a Temperate Coastal Site: A “Multi-Omic” Study of Seasonal and Diel Temporal Variation. *PLoS ONE* 5, e15545.
doi:10.1371/journal.pone.0015545

- Gloux, K., Berteau, O., El oumami, H., Beguet, F., Leclerc, M., Dore, J., 2011. A metagenomic - glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proc. Natl. Acad. Sci.* 108, 4539–4546. doi:10.1073/pnas.1000066107
- Henne, A., Daniel, R., Schmitz, R.A., Gottschalk, G., 1999. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.* 65, 3901 – 3907.
- He, S., Kunin, V., Haynes, M., Martin, H.G., Ivanova, N., Rohwer, F., Hugenholtz, P., McMahon, K.D., 2010. Metatranscriptomic array analysis of “*Candidatus Accumulibacter phosphatis*”-enriched enhanced biological phosphorus removal sludge: Metatranscriptomic array analysis of EBPR sludge. *Environ. Microbiol.* 12, 1205–1217. doi:10.1111/j.1462-2920.2010.02163.x
- Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z., Rubin, E.M., 2011. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* 331, 463–467. doi:10.1126/science.1200387
- Hjort, K., Bergström, M., Adesina, M.F., Jansson, J.K., Smalla, K., Sjöling, S., 2010. Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil. *FEMS Microbiol. Ecol.* 71, 197–207. doi:10.1111/j.1574-6941.2009.00801.x
- Iwai, S., Chai, B., Sul, W.J., Cole, J.R., Hashsham, S.A., Tiedje, J.M., 2009. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *ISME J.* 4, 279–285. doi:10.1038/ismej.2009.104
- Jacquioid, S., Franqueville, L., Cécillon, S., M. Vogel, T., Simonet, P., 2013. Soil Bacterial Community Shifts after Chitin Enrichment: An Integrative Metagenomic Approach. *PLoS ONE* 8, e79699. doi:10.1371/journal.pone.0079699
- Jones, B.V., Begley, M., Hill, C., Gahan, C.G.M., Marchesi, J.R., 2008. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc. Natl. Acad. Sci.* 105, 13580–13585. doi:10.1073/pnas.0804437105
- Kambiranda, D.M., Asraful-Islam, S.M., Cho, K.M., Math, R.K., Lee, Y.H., Kim, H., Yun, H.D., 2009. Expression of esterase gene in yeast for organophosphates biodegradation. *Pestic. Biochem. Physiol.* 94, 15–20. doi:10.1016/j.pestbp.2009.02.006
- Kanehisa, M., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27
- Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., Daniel, R., 2003. Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl. Environ. Microbiol.* 69, 1408 – 1416. doi:10.1128/AEM.69.3.1408-1416.2003
- Kürsten, D., Kothe, E., Wetzels, K., Bergmann, K., Köhler, J.M., 2014. Micro-segmented flow and multisensor-technology for microbial activity profiling. *Env. Sci. Process. Impacts* 16, 2362–2370. doi:10.1039/C4EM00255E

- Ladevèze, S., Tarquis, L., Cecchini, D.A., Bercovici, J., André, I., Topham, C.M., Morel, S., Laville, E., Monsan, P., Lombard, V., Henrissat, B., Potocki-Véronèse, G., 2013. Role of glycoside phosphorylases in mannose foraging by human gut bacteria. *J. Biol. Chem.* 288, 32370–32383. doi:10.1074/jbc.M113.483628
- Ladoukakis, E., Kolisis, F.N., Chatziioannou, A.A., 2014. Integrative workflows for metagenomic analysis. *Front. Cell Dev. Biol.* 2. doi:10.3389/fcell.2014.00070
- Lakhdari, O., Cultrone, A., Tap, J., Gloux, K., Bernard, F., Ehrlich, S.D., Lefèvre, F., Doré, J., Blottière, H.M., 2010. Functional Metagenomics: A High Throughput Screening Method to Decipher Microbiota-Driven NF- κ B Modulation in the Human Gut. *PLoS ONE* 5, e13092. doi:10.1371/journal.pone.0013092
- LeCleur, G.R., Buchan, A., Maurer, J., Moran, M.A., Hollibaugh, J.T., 2007. Comparison of chitinolytic enzymes from an alkaline, hypersaline lake and an estuary. *Environ. Microbiol.* 9, 197–205. doi:10.1111/j.1462-2920.2006.01128.x
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., Henrissat, B., 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6, 41. doi:10.1186/1754-6834-6-41
- Li, M., Hong, Y., Klotz, M.G., Gu, J.-D., 2010. A comparison of primer sets for detecting 16S rRNA and hydrazine oxidoreductase genes of anaerobic ammonium-oxidizing bacteria in marine sediments. *Appl. Microbiol. Biotechnol.* 86, 781–790. doi:10.1007/s00253-009-2361-5
- Lim, S.W., Tran, T.M., Abate, A.R., 2015. PCR-Activated Cell Sorting for Cultivation-Free Enrichment and Sequencing of Rare Microbes. *PLOS ONE* 10, e0113549. doi:10.1371/journal.pone.0113549
- Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., Mueller, A., Schäberle, T.F., Hughes, D.E., Epstein, S., Jones, M., Lazarides, L., Steadman, V.A., Cohen, D.R., Felix, C.R., Fetterman, K.A., Millett, W.P., Nitti, A.G., Zullo, A.M., Chen, C., Lewis, K., 2015. A new antibiotic kills pathogens without detectable resistance. *Nature* 517, 455–459. doi:10.1038/nature14098
- Li, S., Xu, L., Hua, H., Ren, C., Lin, Z., 2007. A set of UV-inducible autolytic vectors for high throughput screening. *J. Biotechnol.* 127, 647–652. doi:10.1016/j.jbiotec.2006.07.030
- Lu, Z., Deng, Y., Van Nostrand, J.D., He, Z., Voordeckers, J., Zhou, A., Lee, Y.-J., Mason, O.U., Dubinsky, E.A., Chavarria, K.L., Tom, L.M., Fortney, J.L., Lamendella, R., Jansson, J.K., D'haeseleer, P., Hazen, T.C., Zhou, J., 2012. Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME J.* 6, 451–460. doi:10.1038/ismej.2011.91
- Majernik, A., Gottschalk, G., Daniel, R., 2001. Screening of Environmental DNA Libraries for the Presence of Genes Conferring Na⁺(Li⁺)/H⁺ Antiporter Activity on *Escherichia coli*: Characterization of the Recovered Genes and the Corresponding Gene Products. *J. Bacteriol.* 183, 6645–6653. doi:10.1128/JB.183.22.6645-6653.2001
- Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson,

- J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Bryant, S.H., 2009. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* 37, D205–D210. doi:10.1093/nar/gkn845
- Markowitz, V.M., Ivanova, N.N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., Chen, I.-M.A., Grechkin, Y., Dubchak, I., Anderson, I., Lykidis, A., Mavromatis, K., Hugenholtz, P., Kyrpides, N.C., 2007. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* 36, D534–D538. doi:10.1093/nar/gkm869
- Math, R.K., Asraful Islam, S.M., Cho, K.M., Hong, S.J., Kim, J.M., Yun, M.G., Cho, J.J., Heo, J.Y., Lee, Y.H., Kim, H., Yun, H.D., 2010. Isolation of a novel gene encoding a 3,5,6-trichloro-2-pyridinol degrading enzyme from a cow rumen metagenomic library. *Biodegradation* 21, 565–573. doi:10.1007/s10532-009-9324-5
- Mayumi, D., Akutsu-Shigeno, Y., Uchiyama, H., Nomura, N., Nakajima-Kambe, T., 2008. Identification and characterization of novel poly(DL-lactic acid) depolymerases from metagenome. *Appl. Microbiol. Biotechnol.* 79, 743–750. doi:10.1007/s00253-008-1477-3
- Mazutis, L., Baret, J.-C., Griffiths, A.D., 2009. A fast and efficient microfluidic system for highly selective one-to-one droplet fusion. *Lab. Chip* 9, 2665–2672. doi:10.1039/b903608c
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386. doi:10.1186/1471-2105-9-386
- Muller, J., Szklarczyk, D., Julien, P., Letunic, I., Roth, A., Kuhn, M., Powell, S., von Mering, C., Doerks, T., Jensen, L.J., Bork, P., 2010. eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.* 38, D190–D195. doi:10.1093/nar/gkp951
- Najah, M., Calbrix, R., Mahendra-Wijaya, I.P., Beneyton, T., Griffiths, A.D., Drevelle, A., 2014. Droplet-Based Microfluidics Platform for Ultra-High-Throughput Bioprospecting of Cellulolytic Microorganisms. *Chem. Biol.* 21, 1722–1732. doi:10.1016/j.chembiol.2014.10.020
- Nawy, T., 2013. Lab-On-A-Chip: Receptive cells feel the squeeze. *Nat. Methods* 10, 198–198. doi:10.1038/nmeth.2395
- Northen, T.R., Lee, J.-C., Hoang, L., Raymond, J., Hwang, D.-R., Yannone, S.M., Wong, C.-H., Siuzdak, G., 2008. A nanostructure-initiator mass spectrometry-based enzyme activity assay. *Proc. Natl. Acad. Sci.* 105, 3678–3683. doi:10.1073/pnas.0712332105
- Nyysönen, M., Tran, H.M., Karaoz, U., Weihe, C., Hadi, M.Z., Martiny, J.B.H., Martiny, A.C., Brodie, E.L., 2013. Coupled high-throughput functional screening and next generation sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries. *Front. Microbiol.* 4. doi:10.3389/fmicb.2013.00282

- Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y., Tsuda, M., 2007. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl. Microbiol. Biotechnol.* 74, 501–510. doi:10.1007/s00253-006-0671-4
- Park, B.H., Karpinets, T.V., Syed, M.H., Leuze, M.R., Uberbacher, E.C., 2010. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* 20, 1574–1584. doi:10.1093/glycob/cwq106
- Pivetal, J., Toru, S., Frenea-Robin, M., Haddour, N., Cecillon, S., Dempsey, N.M., Dumas-Bouchiat, F., Simonet, P., 2014. Selective isolation of bacterial cells within a microfluidic device using magnetic probe-based cell fishing. *Sens. Actuators B Chem.* 195, 581–589. doi:10.1016/j.snb.2014.01.004
- Pleiss, J., Fischer, M., Peiker, M., Thiele, C., Schmid, R.D., 2000. Lipase engineering database. *J. Mol. Catal. B Enzym.* 10, 491–508. doi:10.1016/S1381-1177(00)00092-8
- Poretzky, R.S., Bano, N., Buchan, A., LeClerc, G., Kleikemper, J., Pickering, M., Pate, W.M., Moran, M.A., Hollibaugh, J.T., 2005. Analysis of Microbial Gene Transcripts in Environmental Samples. *Appl. Environ. Microbiol.* 71, 4121–4126. doi:10.1128/AEM.71.7.4121-4126.2005
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., MetaHIT Consortium, Bork, P., Ehrlich, S.D., Wang, J., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi:10.1038/nature08821
- Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., Shah, M., Hettich, R.L., Banfield, J.F., 2005. Community proteomics of a natural microbial biofilm. *Science* 308, 1915–1920. doi:10.1126/science. 1109070
- Rawlings, N.D., Barrett, A.J., Bateman, A., 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 40, D343–D350. doi:10.1093/nar/gkr987
- Robertson, D.E., Steer, B.A., 2004. Recent progress in biocatalyst discovery and optimization. *Curr. Opin. Chem. Biol.* 8, 141–149. doi:10.1016/j.cbpa.2004.02.010
- Saleh-Lakha, S., Miller, M., Campbell, R.G., Schneider, K., Elahimanesh, P., Hart, M.M., Trevors, J.T., 2005. Microbial gene expression in soil: methods, applications and challenges. *J. Microbiol. Methods* 63, 1–19. doi:10.1016/j.mimet.2005.03.007
- Schmidt, O., Drake, H.L., Horn, M.A., 2010. Hitherto Unknown [Fe-Fe]-Hydrogenase Gene Diversity in Anaerobes and Anoxic Enrichments from a Moderately Acidic Fen. *Appl. Environ. Microbiol.* 76, 2027–2031. doi:10.1128/AEM.02895-09

- Schmieder, R., Lim, Y.W., Edwards, R., 2012. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* 28, 433–435. doi:10.1093/bioinformatics/btr669
- Selengut, J.D., Haft, D.H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W.C., Richter, A.R., White, O., 2007. TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res.* 35, D260–D264. doi:10.1093/nar/gkl1043
- Sethi, A., Slack, J.M., Kovaleva, E.S., Buchman, G.W., Scharf, M.E., 2013. Lignin-associated metagene expression in a lignocellulose-digesting termite. *Insect Biochem. Mol. Biol.* 43, 91–101. doi:10.1016/j.ibmb.2012.10.001
- Sharma, V.K., Kumar, N., Prakash, T., Taylor, T.D., 2010. MetaBioME: a database to explore commercially useful enzymes in metagenomic datasets. *Nucleic Acids Res.* 38, D468–D472. doi:10.1093/nar/gkp1001
- Sigrist, C.J.A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A., Hulo, N., 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* 38, D161–D166. doi:10.1093/nar/gkp885
- Simon, C., Daniel, R., 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Appl. Microbiol. Biotechnol.* 85, 265–276. doi:10.1007/s00253-009-2233-z
- Simon, C., Herath, J., Rockstroh, S., Daniel, R., 2009. Rapid Identification of Genes Encoding DNA Polymerases by Function-Based Screening of Metagenomic Libraries Derived from Glacial Ice. *Appl. Environ. Microbiol.* 75, 2964–2968. doi:10.1128/AEM.02644-08
- Sirim, D., Wagner, F., Wang, L., Schmid, R.D., Pleiss, J., 2011. The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases. *Database* 2011, bar006–bar006. doi:10.1093/database/bar006
- Sjostrom, S.L., Bai, Y., Huang, M., Liu, Z., Nielsen, J., Joensson, H.N., Andersson Svahn, H., 2014. High-throughput screening for industrial enzyme production hosts by droplet microfluidics. *Lab Chip* 14, 806–813. doi:10.1039/C3LC51202A
- Soding, J., 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951–960. doi:10.1093/bioinformatics/bti125
- Steele, H.L., Jaeger, K.-E., Daniel, R., Streit, W.R., 2009. Advances in Recovery of Novel Biocatalysts from Metagenomes. *J. Mol. Microbiol. Biotechnol.* 16, 25–37. doi:10.1159/000142892
- Suenaga, H., Ohnuki, T., Miyazaki, K., 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ. Microbiol.* 9, 2289–2297. doi:10.1111/j.1462-2920.2007.01342.x
- Tartar, A., Wheeler, M.M., Zhou, X., Coy, M.R., Boucias, D.G., Scharf, M.E., 2009. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol. Biofuels* 2, 25. doi:10.1186/1754-6834-2-25

- Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., Dore, J., Monsan, P., Remaud-Simeon, M., Potocki-Veronese, G., 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* 20, 1605–1612. doi:10.1101/gr.108332.110
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A., 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41. doi:10.1186/1471-2105-4-41
- Taupp, M., Mewis, K., Hallam, S.J., 2011. The art and design of functional metagenomic screens. *Curr. Opin. Biotechnol.* 22, 465–472. doi:10.1016/j.copbio.2011.02.010
- Thomas, T., Gilbert, J., Meyer, F., 2012. Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* 2, 3. doi:10.1186/2042-5783-2-3
- Tirawongsaroj, P., Sriprang, R., Harnpicharnchai, P., Thongaram, T., Champreda, V., Tanapongpipat, S., Pootanakit, K., Eurwilaichitr, L., 2008. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *J. Biotechnol.* 133, 42–49. doi:10.1016/j.jbiotec.2007.08.046
- Uchiyama, T., Abe, T., Ikemura, T., Watanabe, K., 2004. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat. Biotechnol.* 23, 88–93. doi:10.1038/nbt1048
- Uchiyama, T., Miyazaki, K., 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr. Opin. Biotechnol.* 20, 616–622. doi:10.1016/j.copbio.2009.09.010
- Uchiyama, T., Watanabe, K., 2008. Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nat. Protoc.* 3, 1202–1212. doi:10.1038/nprot.2008.96
- van Elsas, J.D., Costa, R., Jansson, J., Sjöling, S., Bailey, M., Nalin, R., Vogel, T.M., van Overbeek, L., 2008. The metagenomics of disease-suppressive soils – experiences from the METACONTROL project. *Trends Biotechnol.* 26, 591–601. doi:10.1016/j.tibtech.2008.07.004
- van Hellemond, E.W., Janssen, D.B., Fraaije, M.W., 2007. Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl. Environ. Microbiol.* 73, 5832–5839. doi:10.1128/AEM.02708-06
- Vogel, T.M., Simonet, P., Jansson, J.K., Hirsch, P.R., Tiedje, J.M., van Elsas, J.D., Bailey, M.J., Nalin, R., Philippot, L., 2009. TerraGenome: a consortium for the sequencing of a soil metagenome. *Nat. Rev. Microbiol.* 7, 252. doi:Article
- Warnecke, F., Hess, M., 2009. A perspective: Metatranscriptomics as a tool for the discovery of novel biocatalysts. *J. Biotechnol.* 142, 91–95. doi:10.1016/j.jbiotec.2009.03.022
- Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S.G., Podar, M., Martin, H.G., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., Szeto,

- E., Salamov, A., Barry, K., Mikhailova, N., Kyrpides, N.C., Matson, E.G., Ottesen, E.A., Zhang, X., Hernández, M., Murillo, C., Acosta, L.G., Rigoutsos, I., Tamayo, G., Green, B.D., Chang, C., Rubin, E.M., Mathur, E.J., Robertson, D.E., Hugenholtz, P., Leadbetter, J.R., 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565. doi:10.1038/nature06269
- Waschkowitz, T., Rockstroh, S., Daniel, R., 2009. Isolation and Characterization of Metalloproteases with a Novel Domain Structure by Construction and Screening of Metagenomic Libraries. *Appl. Environ. Microbiol.* 75, 2506–2516. doi:10.1128/AEM.02136-08
- Weckx, S., Van der Meulen, R., Allemeersch, J., Huys, G., Vandamme, P., Van Hummelen, P., De Vuyst, L., 2010. Community Dynamics of Bacteria in Sourdough Fermentations as Revealed by Their Metatranscriptome. *Appl. Environ. Microbiol.* 76, 5402–5408. doi:10.1128/AEM.00570-10
- Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C.S., Li, H., Mashiyama, S.T., Joachimiak, M.P., van Belle, C., Chandonia, J.-M., Soergel, D.A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B.J., Bafna, V., Friedman, R., Brenner, S.E., Godzik, A., Eisenberg, D., Dixon, J.E., Taylor, S.S., Strausberg, R.L., Frazier, M., Venter, J.C., 2007. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* 5, e16. doi:10.1371/journal.pbio.0050016
- Zanaroli, G., Balloi, A., Negroni, A., Daffonchio, D., Young, L.Y., Fava, F., 2010. Characterization of the microbial community from the marine sediment of the Venice lagoon capable of reductive dechlorination of coplanar polychlorinated biphenyls (PCBs). *J. Hazard. Mater.* 178, 417–426. doi:10.1016/j.jhazmat.2010.01.097
- Zaprasis, A., Liu, Y.-J., Liu, S.-J., Drake, H.L., Horn, M.A., 2009. Abundance of Novel and Diverse tfdA-Like Genes, Encoding Putative Phenoxalkanoic Acid Herbicide-Degrading Dioxygenases, in Soil. *Appl. Environ. Microbiol.* 76, 119–128. doi:10.1128/AEM.01727-09

Chapter II:

Metagenomics for the discovery of pollutant degrading enzymes

Lisa Ufarté^{1,2,3}, Élisabeth Laville^{1,2,3}, Sophie Duquesne^{1,2,3}, Gabrielle Potocki-Veronese^{1,2,3*}

¹Université de Toulouse, INSA, UPS, INP, LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

²INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France

³CNRS, UMR5504, F-31400 Toulouse, France

Biotechnology Advances (2015)

Abstract

Organic pollutants, including xenobiotics, are often persistent and toxic organic compounds resulting from human activities and released in large amounts into terrestrial, fluvial and marine environments. However, some microbial species which are naturally exposed to these compounds in their own habitat are capable of degrading a large range of pollutants, especially poly-aromatic, halogenated and polyester molecules. These microbes constitute a huge reservoir of enzymes for the diagnosis of pollution and for bioremediation. Most are found in highly complex ecosystems like soils, activated sludge, compost or polluted water, and more than 99% have never been cultured. Meta-omic approaches are thus well suited to retrieve biocatalysts from these environmental samples. In this review, we report the latest advances in functional metagenomics aimed at the discovery of enzymes capable of acting on different kinds of polluting molecules.

Keywords: Pollutants, Metagenomics, Enzyme discovery, Screening

1. Introduction

Large amounts of pollutants are released into the air, water and soil as a result of industrial, agricultural and domestic activities. Organic pollutants are synthetic compounds in the form of herbicides, dyes, pesticides, plastics and drugs (Rieger et al., 2002). Most are aromatic molecules, polymers of ring shaped molecules or planar molecules, and are therefore among the most stable and persistent molecules. Those that most frequently accumulate in natural environments are polycyclic aromatic hydrocarbons (PAH) especially chlorinated hydrocarbons, steroids (phenols, phthalates) and dyes, but also organocyanides including nitriles, long chain aliphatics found in numerous plastics and insulating materials like polyurethanes, or organophosphates and pyrethroid herbicides and pesticides.

The toxicity of these compounds for the environment and for mammalian organisms results from their resistance to natural degradation, which is linked to their stable structures. They are grouped under the name of persistent organic pollutants (POP), some of them being potent carcinogens or mutagens, and/or having endocrine disrupting properties (Ballschmiter et al., 2002).

Cleaning up contaminated soils, wastewater, groundwater and marine environments is a challenging task. Physical and chemical technologies (electrochemical treatments, oxidising agents, activation by ultraviolet rays, adsorption of pollutants, membrane filtration, ion exchange, electrokinetic coagulation, etc.) can be combined to reduce contamination to a safe and acceptable level. Many of these processes have drawbacks, such as formation of by-products, high sludge production, and high processing costs (Riser-Roberts, 1998; Robinson et al., 2001).

In this context, recent decades have seen the development of biological processes based on the breakdown of pollutant organic compounds into stable, innocuous end-products by bacteria (Watanabe, 2001; Yam et al., 2010), fungi (Cerniglia, 1997; Bertrand et al., 2015), algae or their enzymes (Sutherland et al., 2004). Since bioremediation may result in the complete metabolisation of pollutants, it is considered to be a highly effective and environmentally friendly strategy (Colleran, 1997). Microorganisms have indeed developed a wide range of aerobic and anaerobic catabolic strategies to degrade the huge range of organic compounds present in the ecosystems they colonise. Because pollutant molecules are often structurally similar to natural molecules, one can assume that there are always some organisms in contaminated ecosystems that are able to metabolise pollutants, which serve as their main carbon source.

During microbial degradation, all changes in the chemical structure of pollutants are due to the action of enzymes, whose specificity is often broad enough to accommodate several molecules of similar structures. Once identified and isolated, these enzymes can therefore be engineered by directed evolution to improve their stability or efficiency with respect to a particular compound (Festa et al., 2008; Theerachat et al., 2012). However, the greatest advances in recent decades have been made thanks to the development of next generation sequencing and other 'omics' technologies. These new technologies have enabled numerous insights into the genes and metabolic pathways involved in pollutant catabolism by bacteria and fungi, thereby improving our understanding of the genetic and molecular bases of the reactions involved in the degradation processes. Most knowledge has been obtained by functional genomics of pollutant-degrading microorganisms or model microbial communities. This allowed one, after genome sequencing and/or activity-based screening of genomic libraries, to identify genes encoding the enzymes that are directly involved in the breakdown of the targeted compound (Pieper et al., 2004). However, since more than 99% of the microbial species of the terrestrial (Pham and Kim, 2012) and aqueous (Zengler et al., 2002) ecosystems are uncultured, one can consider that most of the microorganisms that are capable of breaking down pollutants remain uncharacterised.

Functional metagenomics, which consists of assigning functions to proteins encoded by all genomes of a microbial community with no isolation and cultivation step, is a highly efficient way to boost the discovery of novel biocatalysts from the huge diversity of uncultured microbes. In this review, we present the various functional metagenomic approaches that require, or not, cloning of metagenomic fragments in an expression host in order to screen and experimentally validate gene functions, which have been used in recent years to identify enzymes and metabolic pathways involved in pollutant degradation.

2. Targeted metagenome sampling

Numerous microbial communities have the potential to degrade and metabolise most pollutants. This is even the case of those in natural, non-polluted ecosystems, as long as (i) their taxonomical and functional diversity is sufficient, and (ii) the structure of the targeted pollutant is similar to that of one of the natural substrates used as carbon source to support microbial growth. Therefore, several such natural microbiota that had never or only rarely been exposed to pollutants, were successfully mined for xenobiotic degrading enzymes. For example, as detailed later in subsections 3.2.2.1 and 3.2.2.2, organophosphate and 3,5,6-trichloro-2-pyridinol (TCP) degrading enzymes (Math et al., 2010), chlorpyrifos insecticides degrading esterases (Kambiranda et al., 2009) were discovered in the cow rumen microbiome. Likewise, in the human gut metagenomic gene catalogue, 15% of genes have been annotated as being involved in the degradation and metabolism of xenobiotics, like halogenated aromatic compounds (Qin et al., 2010).

Nevertheless, since in polluted environments the microbial communities are enriched in microorganisms able to thrive by degrading pollutant compounds, most metagenomic studies have been conducted on polluted soils, activated sludge, sediment, or wastewater environments. To reduce the complexity of the sample, stable isotope probing (SIP) can be used to specifically label the genomes of pollutant metabolising species. For instance, Wang et al. (2012) amended microcosms from contaminated groundwater with ¹³C-labelled naphthalene. The ¹³C-labelled DNA was then easily separated from the unlabelled DNA, to be specifically sequenced and concurrently cloned before activity-based screening. This approach is suitable for focusing on active degraders, especially those which act on relatively simple pollutant structures. Indeed, when more complex compound structures are targeted, polyaromatics for instance, there is a risk of missing the first steps of degradation undertaken by microbes acting on the most recalcitrant molecules, with the simpler reaction products being metabolised and incorporated in the DNA of the end-user microbes.

3. Screening for pollutant degrading enzymes

Two complementary approaches can be used to identify pollutant degrading biocatalysts in microbial communities. One is guided by gene sequence analysis and functional annotation, based on the content of available sequence databases (Cardenas and Tiedje, 2008). The other is guided by the observation of pollutant degrading phenotypes, harboured by recombinant metagenomic clones (Fig. 3). In both cases, only the cloning of the targeted metagenomic sequences allows the gene function to be validated experimentally.

3.1. Sequence-based approaches to understand pollutant degradation mechanisms by microbial communities

Sequence-based metagenomics relies on whole-genome DNA extraction from microbial communities, shotgun sequencing, and read assembly (Fig. 3). The assembled sequences are subjected to bioinformatic analysis to assess the composition and functions of the microbial populations. Over the past 10 years many sequencing projects have targeted heavily polluted ecosystems to study the structure of microbial populations (Chouari et al., 2003; Guerhazi et al., 2008), species interactions, metabolic pathways, and the genes involved in species survival in such environments (Yamada et al., 2012). Among these projects, we cite the culture-independent recovery of nearly complete genomes of five dominant members of an acid mine drainage (AMD), by in-depth metagenome sequencing (Tyson et al., 2004). Functional annotation has provided new insights into the ecological roles of individual members of this community, with ferrous iron oxidation playing a key role in the microbial and geochemical processes in AMD ecosystems. More recently, the metagenome of the activated biomass from a common effluent treatment plant was sequenced to retrieve novel oxygenase sequences (Jadeja et al., 2014). The sequences were compared to enzymes identified in previous metagenomes from different wastewater compositions and at different locations, including a sewage effluent treatment plant, an enhanced biological phosphorus treatment plant, and a tannery waste treatment plant (Albertsen et al., 2012; Yu and Zhang, 2012; Wang et al., 2013). In addition, samples of the activated biomass were assessed for the degradation of aromatic compounds like anthracene, biphenyl, naphthalene, phenol and *o*-toluidine. Reaction products were further analysed by gas chromatography–mass spectrometry (GC–MS). However, this study did not ascribe the degradation potential to specific bacteria and their enzymes. Another study succeeded in discovering a key functional operon for naphthalene degradation by using the SIP method coupled with shotgun sequencing targeted to the ¹³C-labelled DNA of metabolically active naphthalene degraders (Wang et al., 2012). Specific primers were therefore designed from the operon [¹³C]DNA sequence to clone it into a biosensor-based genetic transducer system activated by salicylate, an intermediate metabolite of naphthalene catabolism. The involvement of the labelled operon in the degradation of naphthalene was thus experimentally confirmed through the activation of the biosensor by salicylate. To our knowledge, this is the only study of massive metagenome sequencing that has resulted in the characterisation of enzymes for the degradation of pollutant molecules. The other sequencing projects resulted in the creation of catalogues of annotated genes, representing a considerable reservoir of potentially useful catalysts for bioremediation. This is the case in particular for the 1,200 bacterial laccase-like genes identified in several genomic and metagenomic datasets (Ausec et al., 2011), whose functions remain to be tested for the breakdown of polyaromatic pollutants.

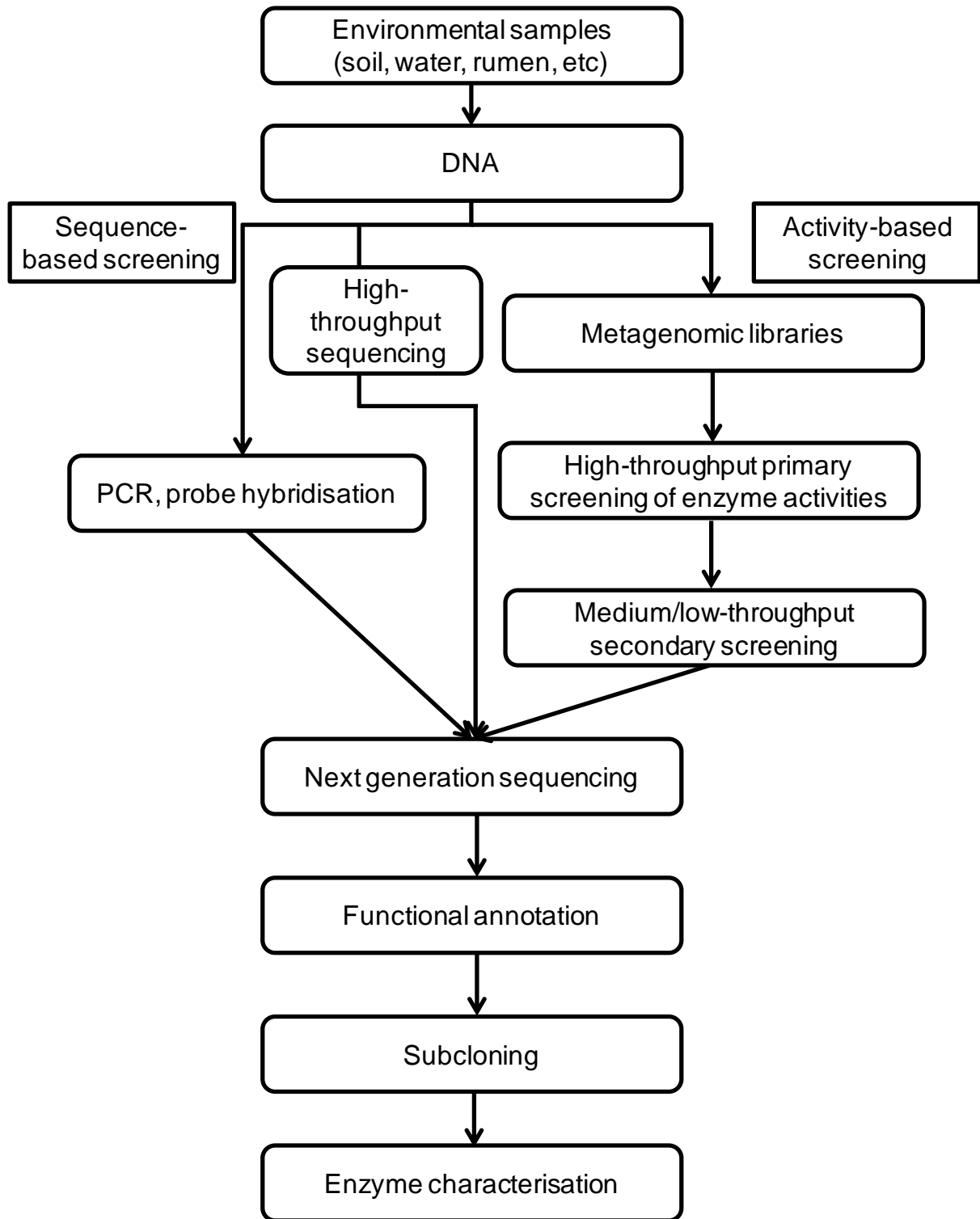


Fig. 3. Schematic representation of sequence-based and activity-based functional metagenomics approaches.

Metagenomes can also be specifically screened for known enzyme families via amplification by polymerase chain reaction (PCR) or DNA/ DNA or DNA/cDNA hybridisation, using conserved sequences as primers or probes. These strategies bypass massive and/or random sequencing, but as sequencing, they limit capture to genes with sufficient homology with previously characterised enzymes. For this reason, these approaches are not suitable for the discovery of new protein families whose sequences are not available, or not properly annotated in databases. The literature is rich in examples of such sequence-based metagenomic studies focused on xenobiotic degrading enzymes. Quantitative PCR assays were, for instance, developed to analyse the diversity and abundance of genes encoding 2,4-D/ α -ketoglutarate and Fe^{2+} -dependent dioxygenases that initiate the biotic degradation of phenoxyalkanoic acid herbicides (Zaprasis et al., 2009). In the same way, ammonia monooxygenases, the key enzymes of denitrification, have been amplified from different environmental samples (soils, the sea and microcosms) contaminated with nitrates (Nogales et al., 2002; Treusch et al., 2005; Bartossek et al., 2010).

The use of high density arrays designed to detect specific functions makes it possible to dramatically multiply the number of probes, thereby giving a glimpse of the abundance and diversity of sequences in specific protein families (Eyers et al., 2004). One example is the GeoChip 4.0, containing 83,992 50-mer oligonucleotide probes targeting 152,414 genes in 410 gene categories for different chemical processes used to detect functional genes specific to hydrocarbon (HC) degradation processes in a deep sea water sample contaminated by an oil spill, in comparison with a non-contaminated sample (Lu et al., 2012). This approach was particularly suited to investigate modifications in the functional structure of microbial communities and to quantify the number of genes involved in aerobic HC degradation. As is often the case with sequence-based approaches, in the last two examples, the main aim of the study was to investigate the diversity of the targeted genes and species involved in bioremediation, but was not dedicated to the capture of new catalysts.

Indeed, only rare studies using the PCR-based approach have resulted in the identification of a full length gene sequence and the characterisation of its enzyme product. Among them, we cite the discovery of bacterial laccases by PCR-screening of a sea metagenomic library, of which the activity towards syringaldazine, azo-dyes, 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS), 2,6-dimethoxyphenol and/or catechol was validated experimentally (Fang et al., 2010, 2012). Several other studies were targeted to ecosystems polluted by polychlorinated biphenyl (PCB). Indeed, PCBs were manufactured widely from 1929 to the 1970s and an estimated 1.5 million tons of them have been produced worldwide (Aken et al., 2010), which explains the importance of identifying new degradation mechanisms. Using the SIP approach, microcosms from polychlorinated biphenyl PCB-contaminated river sediments were amended with labelled ^{13}C biphenyl (Sul et al., 2009). The labelled DNA was used to construct a cosmid library containing 30 to 40 kb DNA inserts, which was PCR screened using primers specific for aromatic dioxygenases. One

cosmid clone was retrieved containing a biphenyl dioxygenase encoding gene, which was subsequently subcloned and expressed to confirm its biphenyl degrading function.

In the last decade, metagenomic sequencing has revealed that enzyme sequence diversity is much broader than previously assumed on the basis of functional genomic data. Thus, many enzyme encoding genes cannot be captured with conventional primer sets. In addition, the metagenomic DNA fragments retrieved using PCR-based approaches are often only fragments of the targeted genes. The PCR based approach is thus coupled to other methods enabling the recovery of entire genes. Indeed, the gene targeted metagenomic approach (Iwai et al., 2009) combines PCR screening with pyrosequencing of amplicons to iteratively generate new primers which can be designed from other conserved regions, thus increasing the structural diversity of the targeted protein families. This approach has enabled the discovery of thousands of full length dioxygenase encoding genes, assigned to dozens of novel clusters containing still unknown sequences, from the metagenomes of soils contaminated by polychlorinated biphenyl (Iwai et al., 2009) and 3-chlorobenzoate (Morimoto and Fujii, 2009). In these studies, the authors used two approaches: first the PCR denaturing gradient gel electrophoresis (DGGE) to demonstrate DNA bands becoming dominant after addition of the targeted pollutant to soil. Second, the dominant bands were used as target fragments for the PCR–metagenome walking approach to retrieve complete functional genes. To our knowledge, their functional characterisation has not yet been reported.

Another way to exploit the diversity of metagenomic sequences is metagenomic DNA shuffling. This method has been developed for the creation of novel genes by recombination of the DNA sequence of a gene of interest with metagenomic DNA fragments issued from highly complex microbial communities, for instance those of contaminated soils. Boubakri et al. used this approach to create a library of more than one thousand clones harbouring hybrid sequences that was screened on a solid medium containing lindane, an organochlorine pesticide. No less than 23 new genes encoding lindane degrading enzymes were obtained by using this smart technology (Boubakri et al., 2006).

In most of the studies cited above, proof of enzymatic function has not been established, even if these investigations enabled the capture of the sequences of interest. Their sub-cloning, expression and biochemical characterisation of the protein they encode are indeed needed for confirmation of the targeted activities. As illustration of such in-depth characterisation of enzyme properties, one could cite the work of Vergne-Vaxelaire et al. (2013), who selected thirty four sequences from experimentally confirmed nitrilases for use as reference set to choose 290 candidate nitrilase sequences from both genomes of their strain collection and metagenome of a wastewater treatment plant. These nitrilases, assumed to be representative of nitrilase diversity, were further cloned through a high-throughput cloning platform for overexpression in *Escherichia coli* and further

characterisation, though not as pollutant degrading tools. Until now, there have been few published reports of the discovery and experimental proof of the function of pollutant degrading enzymes by sequence-based metagenomics (Table 2).

3.2. Activity based metagenomics for the discovery of new tools for bioremediation

Contrary to sequence-based approaches, activity-based metagenomics is highly efficient to simultaneously screen a microbiome for novel enzymes and acquire biochemical data regarding substrate specificity. This technology relies on the observation of a phenotype, linked to the reaction(s) involved in the breakdown of the targeted pollutant. It has three prerequisites: (i) cloning of DNA or cDNA fragments between 2 and 200 kbp in length into an expression vector (plasmids, cosmids, fosmids or even bacterial artificial chromosomes) for the creation of metagenomic or metatranscriptomic libraries, respectively; (ii) heterologous expression of cloned genes into a microbial host; (iii) design of sensitive phenotypic screens to isolate clones of interest with the targeted activity, also called screening 'hits'. Sequencing efforts are then concentrated on the hits and provide entire gene sequences coding for functional proteins. Because hit genes are not retrieved on the basis of sequence homology with already known genes, activity-based metagenomics is the only known way to identify new protein families or to attribute new functions to already known protein families.

Two generic high-throughput strategies are widely used for primary screening. One is based on direct detection of colouration or discolouration, when coloured substrates or chromogenic substrates are used, or of a reduction in opacity of the reaction medium when insoluble substrates are used (Shah et al., 2008). The other is based on heterologous complementation of an auxotrophic host by foreign genes, which allows microbial host growth on selective culture media (Xing et al., 2012) (Fig. 3). However, the design of screens with pollutant compounds as substrates can be challenging. Such compounds may indeed be toxic for the screening host, and even for the operator, depending on the amount of substrate used for screening. In this case, non-toxic structural analogues of pollutants are used for high-throughput screening, and the real toxic substrates are used only in confined media in secondary screening assays with lower throughput, often involving HPLC, GC and/or MS analysis of reaction substrates and products.

Many more activity-based metagenomic studies have provided evidence of function for pollutant degrading enzymes than sequence based metagenomic studies. Two main families of such enzymes have been unearthed from activity based screening: oxidoreductases and hydrolases. *E. coli* was the most frequently used expression host (Table 3), even if other bacterial hosts from the *Bacillus*, *Sphingomonas*, *Streptomyces*, *Thermus*, and *Pseudomonas* genera have been streamlined for constructing metagenomic libraries, that were exploited in other contexts (Taupp et al., 2011; Ekkers et al., 2012). The

transformation efficiency of *E. coli* is indeed particularly high, even for vectors sizing several dozen kbp, and this host was showed to properly express genes from various taxonomically distant bacterial species (Tasse et al., 2010).

3.2.1. Oxidoreductases

3.2.1.1. Oxygenases.

Oxygenases are mostly sought after for the elimination of aromatic compounds. To break them down, bacteria use a typical aerobic degradation pathway, which can be broken down into two critical steps: ring hydroxylation of adjacent carbon atoms involving phenol hydroxylases (phenol 2-monooxygenases) and ring cleavage of the resulting catecholic intermediates involving catechol 1,2- or 2,3-dioxygenases (Silva et al., 2013).

In the works presented in this review, the libraries used to search for oxygenases were constructed from samples of environmental bacterial DNA ranging from 3 to 50 kb in size, expressed in *E. coli* strains, except in the studies by Ono et al. (2007) and Nagayama et al. (2015), in which the expression hosts was the naphthalene-degrading *Pseudomonas putida* strain specifically designed for heterologous complementation, in order to detect naphthalene dioxygenases and enzymes involved in phenol and catechol degradation (monooxygenases, multicomponent hydroxylase, and a reductase).

These new oxygenases were all retrieved from highly polluted environments, like activated sludge collected from wastewater treatment facilities belonging to the petroleum, coke and pharmaceutical industries (Suenaga et al., 2007; Sharma et al., 2012; Silva et al., 2013), or soils contaminated with PAH, PCBs, and oil (Ono et al., 2007; Brennerova et al., 2009; Lu et al., 2011). Artificial enrichment strategies in aromatic compound (phenol, biphenyl, phenanthrene, carbazole, and 3-chlorobenzoate) degrading bacteria were also used in several cases (Silva et al., 2013; Nagayama et al., 2015).

In these studies, two kinds of primary screening assays on solid medium were used to search for mono- or dioxygenases. In one case only, positive selection was used by spotting the metagenomic library on agar plates containing naphthalene as the sole carbon source (Ono et al., 2007). But this strategy is limited to compounds that are non-toxic for the cloning host, and that can be metabolised. Growing metagenomic clones on rich medium was thus used much more often, before detection of coloured halos around clones producing indigo or indirubin pigments by phenolic substrate cleavage. A substrate like indole can either be added to the medium or formed from tryptophan by the endogenous *E. coli* enzyme tryptophanase (van Hellemond et al., 2007; Nagayama et al., 2015). But in most cases, after cell growth, agar plates are further sprayed with substrate solutions, like catechol (Brennerova et al., 2009), phenol (Sharma et al., 2012), 2,3-dihydroxybiphenyl

(Wang et al., 2015), or 3,5-dichlorocatechol (a precursor of a widely used herbicide 2,4-dichlorophenoxyacetic acid) (Lu et al., 2011).

Positive selection and screening can also be performed in liquid media, as done by Silva et al. (2013), who grew the clones in liquid medium containing phenol as sole carbon source. Cell growth was monitored using a chromogenic reagent [3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2 H-tetrazolium bromide], to evaluate microbial respiration and subsequent phenol consumption.

To screen for extradiol dioxygenases, Suenaga et al. (2007) incubated metagenomic clone cellular extracts on catechol, ring cleavage being identified from the absorption coefficients of the ring cleavage products formed from catechol.

Using either solid or liquid assays, phenolic degrading hits were retrieved at a yield varying between 9.10^{-4} and 1.6 per Mbp of screened metagenomic DNA, depending on the library tested and on the screening strategy (Table 3). In most studies, the activity of the hit clones was further estimated or quantified in liquid medium, using the same substrates as in the primary screens, or by testing structural analogues to evaluate enzyme promiscuity towards these complex compounds. Suenaga et al. thus tested all the hit clones on catechol derivatives (3-methylcatechol, 4-methylcatechol, 4-chlorocatechol and 2,3-dihydroxybiphenyl), using the same strategy as in primary screening to monitor the appearance of ring cleavage products. Discrimination assays were also often based on substrate quantification by GC-MS, after incubation with the hit clones (Ono et al., 2007; van Hellemond et al., 2007; Kimura et al., 2010; Nagayama et al., 2015).

The sequences of the enzymes responsible for the screened activities were then retrieved by direct sequencing, sometimes after transposon mutagenesis (Kimura et al., 2010) in order to unambiguously identify the gene target when long metagenomic DNA inserts were used to create the libraries. PCR based screening can also be used to target specific enzyme families encoded by the hit from activity based screening to clone DNA inserts. Using a set of primers specific for eight groups of extradiol dioxygenases of different taxonomic origin, Brennerova et al. isolated and subcloned a particularly interesting enzyme presenting a broad specificity towards different catecholic substrates (3-methylcatechol, catechol, 2,3-dihydroxybiphenyl and 1,2-dihydroxynaphthalene) (Brennerova et al., 2009).

Subcloning of the best gene targets and in depth enzyme characterisation is the last step required to evaluate the applicative potential of the newly discovered enzymes. For instance, Lu et al. (2011) identified a very promising enzyme for bioremediation of chlorophenol contaminated environments. The enzyme indeed exhibited monooxygenase activity, checked by monitoring the substrate dependent oxidation of NADPH towards all fifteen of the tested substrates which derived from the herbicide 2,4-dichlorophenoxyacetic

acid. In very rare cases, real bioremediation trials have been performed using the best enzyme candidates extracted from metagenomes. We cite the pioneering study of Wang et al. (2015), who created transgenic lines of alfalfa plants harbouring a 2,3-dihydroxybiphenyl-1,2-dioxygenase acting on 2,3-dihydroxybiphenyl, 4-methylcatechol, 3-methylcatechol, 4-chlorocatechol and catechol, used to remove PCBs from their culture medium.

3.2.1.2. Laccases.

Laccases are multi-copper oxidoreductases which oxidise a wide variety of phenolic and non-phenolic compounds, polycyclic aromatic hydrocarbons like industrial dyes, pesticide alkenes and recalcitrant biopolymers such as lignin (Beloqui et al., 2006). Due to their broad substrate specificity, they are good candidates for bioremediation (Ausec et al., 2011). Substrate oxidation by laccases (and subsequent reduction of molecular oxygen) creates reactive radicals which may be involved in (i) polymerisation (oxidative coupling of monomers), (ii) degradation of the covalent linkages between the monomers or (iii) phenol degradation by cleavage of aromatic rings. Substrate specificity is broadened by laccase mediators, which, once oxidised into radicals, can then oxidise a variety of other substrates (Ausec et al., 2011). So far, the majority of characterised laccases have been retrieved from fungi. However, laccase encoding genes are widely represented in bacterial genomes and metagenomes (Ausec et al., 2011), which has encouraged the functional exploration of prokaryotic microbial ecosystems for laccases, especially those acting on dyes. Indeed, more than 10,000 commercially available ones and about 7×10^5 metric tons of dyestuffs are produced annually by the textile, plastic, and tannery industries. Inefficient dyeing results in the direct loss of large amounts of dyestuff in wastewater, which ultimately finds its way into the environment, thereby disrupting aquatic ecosystems. These concerns have led to strict regulations for the discharge of coloured wastewater, obliging dye manufacturers and users to adopt “cleaner technology” approaches, based, for instance, on the use of oxidase type enzymes (Vaghela et al., 2005; Satar and Husain, 2011).

Ferrer's team was a pioneer in this field (Ferrer et al., 2005; Beloqui et al., 2006). As soon as 2005, they screened a phagemid *E. coli* library created from bovine rumen metagenome on agar plates containing syringaldazine as substrate. One phage conferring laccase activity was sequenced, and the laccase encoding gene was subcloned for further enzyme characterisation on syringaldazine, 2,6-dimethoxyphenol, veratryl alcohol, guaiacol, 4-methoxybenzyl alcohol and ABTS. Phenol and 1-hydroxybenzotriazole were also oxidised despite being unusual substrates for laccase. This new enzyme was classified as a polyphenol oxidase with laccase activity, and it exhibited much higher catalytic efficiency than previously described for laccases. This enzyme is the first functionally characterised member of this new laccase family. Based on sequence homology, two other members of this family, from *E. coli* and *Bacteroides thetaiotaomicron* were subsequently functionally characterised (Beloqui et al., 2006).

Table 2: Discovery of pollutant degrading enzymes using sequence based metagenomics approaches. All the cited studies provided biochemical proof of function for the hits obtained.

Ecosystem	Pollutant	Category	Method	Sequenced DNA size	Hits	Reference
Groundwater	Naphthalene/PAH	Mothball, fumigant	PCR amplification and functional annotation	~ 431 Mb	3 multigenic systems encoding 1,2-dihydroxynaphthalene degrading enzymes	(Wang et al., 2012)
Wastewater	Azo-dyes	Dyes	PCR amplification	1.4 Gb	1 laccase	(Fang et al., 2010)
Wastewater	Catechol; Azo-dyes	Pesticide precursor; Dyes	PCR amplification	1.4 Gb	1 laccase	(Fang et al., 2012)
Wastewater	Naphthalene, Anthracene; Biphenyl; Phenol; <i>o</i> -toluidine	PAH; Precursor to PCBs; Toxic effect on neurosystem; Dye precursor	Functional annotation and comparative metagenomics	~ 30 Mb	191 oxygenases annotated, around 50% involved in metabolism of aromatic compound degradation	(Jadeja et al., 2014)
River sediments	Polychlorinated biphenyl	Dielectric, coolant fluid; POP	PCR amplification	~ 55 Mb	1 dioxygenase fragment	(Sul et al., 2009)
Soil	Lindane	Insecticide	DNA shuffling	Unspecified size	23 DNA hybrids	(Boubakri et al., 2006)
Soil	3-chlorobenzoate	Pesticide precursor	PCR and metagenome walking	PCR product: 3,649 bp	2 dioxygenases	(Morimoto and Fujii, 2009)
Soil	Phenoxyalkanoic acid	Herbicide	Degenerated PCR primers	PCR products: 260, 360 and 815 bp	11 multigenic clusters encoding dioxygenase	(Zapras et al., 2009)
Soil	Polychlorinated biphenyl	Dielectric, coolant fluid; POP	Pyrosequencing/functional annotation	524 bp region	1 biphenyl dioxygenase	(Iwai et al., 2009)

Table 3: Discovery of pollutant degrading enzymes by activity-based metagenomics. All screening studies were performed by using *E. coli* as cloning host, except for the work of Ono et al. (2007) and Nagayama et al. (2015), performed in *P. putida*.

Ecosystem	Screening Substrate	Pollutant family	Category	Activity screened	Screened library size (sequence coverage)	Number of hit clones	Reference
Activated sludge	Catechol	Aromatic hydrocarbon	Pesticide (precursor)	Extradiol dioxygenase	96,000 fosmid clones (3.2 Gb)	91	(Suenaga et al., 2007)
Activated sludge	4-nitrotoluene	Aromatic compound	Dye precursor	Monooxygenase	40,000 fosmid clones (1.6 Gb)	6	(Kimura et al., 2010)
Industrial effluent	Catechol	Aromatic hydrocarbon	Pesticide (precursor)	Dioxygenase and monooxygenase	40,000 BAC clones (440 Mb) + 32,000 plasmid clones (192 Mb)	4	(Sharma et al., 2012)
Loam Soil	Styrene, <i>o</i> -chlorostyrene, <i>m</i> -chlorostyrene, <i>p</i> -chlorostyrene	Plastics	Suspected human carcinogen	Styrene monooxygenase	65,000 plasmid clones (357.5 Mb)	2 (identical)	(van Hellemond et al., 2007)
Loam Soil	Naphthalene	PAH	Fumigant, precursor	Naphthalene dioxygenase	24,000 cosmid clones (600 Mb)	2 clones	(Ono et al., 2007)
Soil	Catechol	Aromatic hydrocarbon	Pesticide (precursor)	Extradiol dioxygenase	87,000 + 300,000 fosmid clones (2.6 + 10.5 Gb)	254	(Brennerova et al., 2009)
Soil	3,5-dichlorocatechol	Phenoxy acid herbicide	2,4-Dichlorophenoxyacetic acid (herbicide) precursor	Chlorocatechol 1,2-dioxygenase	Plasmid clones, size unspecified	1	(Lu et al., 2011)
Soil	Catechol; α -naphthol	Aromatic hydrocarbon; Carbaryl (carbamate)/naphthalene (PAH)	Pesticide (precursor) Insecticide (precursor)	Laccase	8,000 plasmid clones (44 Mb)	1	(Ye et al., 2010)

Metagenomics for the discovery of pollutant degrading enzymes

Soil	Malathion; <i>cis</i> -permethrin, <i>trans</i> -permethrin, cypermethrin, fenvalerate, deltamethrin	Organophosphate; Pyrethroids	Insecticide	Pyrethroid-hydrolysing esterase	93,000 plasmid clones (390 Mb)	1	(Li et al., 2008)
Soil	Cyhalothrin, cypermethrin, sumicidin, deltamethrin	Pyrethroids	Insecticide	Pyrethroid-hydrolysing esterase	21,000 plasmid clones (100 Mb)	1	(Fan et al., 2012)
Soil	Poly(diethylene glycol adipate)	Polyurethane	Many uses in industry (plastics...)	Esterase	13,000 fosmid clones, unspecified size	10	(Kang et al., 2011)
Artificially polluted soil	Catechol Phenol Related compounds	Aromatic compounds	Pesticide (precursor) Endocrine disruptor	Dioxygenase, phenol hydroxylase	208,000 cosmid clones (5.2 Gb)	29	(Nagayama et al., 2015)
PCBs-contaminated Soil	4-methylcatechol, 3-methylcatechol, 4-chlorocatechol, catechol	Aromatic compounds	PCBs	2,3-dihydroxybiphenyl-1,2-dioxygenase	Plasmid clones, unspecified size	1	(Wang et al., 2015)
Compost	PLA, PHB, PCL, PBSA, PES	Polyester	Plastics	Esterase	40,000 plasmid clones (100 Mb)	7	(Mayumi et al., 2008)
Leaf-branch compost	PCL, PET	Polyester	Plastics	Cutinase	21,000 fosmid clones (735 Mb)	19	(Sulaiman et al., 2012)
Cow rumen	Cadusafos, chlorpyrifos, coumaphos, diazinon, dyfonate, ethoprophos, fenamiphos, methylparathion, parathion	Organophosphorus	Insecticide	Esterase	Fosmid clones, unspecified size	1	(Kambiranda et al., 2009)

Metagenomics for the discovery of pollutant degrading enzymes

Cow rumen	Poly-R 478	Dye	Anthrapyridone chromophore	Polyphenol oxidase with laccase activity	14,000 phagemid clones (1.1 Gb)	1	(Beloqui et al., 2006)
Cow rumen	3,5,6-trichloro-2-pyridinol (TCP)	Organophosphorus	Chlorpyrifos (insecticide) precursor	TCP degrading enzyme	Fosmid clones, unspecified size	1	(Math et al., 2010)
Biofilms, from wastewater treatment plant	Dibutyl phthalate	Phthalate ester	Ectoparasiticide	Dialkyl phthalate ester hydrolase	Fosmid clones (400 Mb)	3	(Jiao et al., 2013)
Wastewater treatment sludge	Phenol	Aromatic organic compound	Herbicide precursor	Phenol degrading enzyme	10,000 fosmid clones + 3,200 fosmid clones (528 Mb)	413	(Silva et al., 2013)
Oil contaminated water and phenanthrene/pyrene amended water	PLA, PHB, polycaprolactones polybutylene-succinate-co-adipate	Polyester	Plastics	Carboxylesterase	100,000 phagemid clones + 100,000 phagemid clones + 60,000 phagemid clones (1.43 Gb)	95 (3 unique)	(Tchigvintsev et al., 2014)
Oil contaminated water	Hexadecane	Alkane hydrocarbon	Diesel fuel	Hexadecane degrading enzyme	31,000 fosmid clones (1.24 Gb)	72	(Vasconcellos et al., 2010)
	Naphthalene, phenanthrene	PAH	Diesel fuel	PAH degrading enzyme	31,000 fosmid clones (1.24 Gb)	3	(Sierra-García et al., 2014)
	n-alkanes; phytane, pristane; Phenanthrene, methylphenanthrene	Alkane; Isoprenoids; PAH	Diesel fuel	Alkane/Isoprenoid/PAH degrading enzymes	31,000 fosmid clones (1.24 Gb)	5	(Dellagnezze et al., 2014)

The discovery of another bacterial laccase by activity-based metagenomics was reported by Ye et al. (2010). These authors constructed a plasmid *E. coli* library from environmental DNA isolated from mangrove soil. Laccase screening was performed in liquid medium, based on the detection of a brownish red colour resulting from the hydrolysis of guaiacol incubated with cellular extracts. The only positive clone obtained displayed laccase activity on α -naphthol and catechol (respectively insecticide and pesticide precursors) and other related substrates. The optimal pH of this polyspecific enzyme is highly alkaline, which is thus compatible with dyestuff processing, lignocellulose biorefinery and biobleaching of paper pulp.

More recently, a pseudolaccase owing to the requirement of exogenous Cu(II) for oxidase activity was discovered from coal bed metagenomes, screened for lignin catabolic activities. This pioneering work, which included development of a very original method of screening for polyaromatic degrading enzymes, highlighted the promiscuity of many biocatalysts towards this kind of substrates, and the metagenome flexibility in loci encoding polyaromatic degrading pathways (Strachan et al., 2014).

3.2.1.3. Alkane degrading enzymes.

Only a few studies have targeted alkane degrading enzymes. Vasconcellos et al. (2010) published a major one in this field. They constructed a metagenomic library from the pooled DNA of aerobic and anaerobic cultures in mineral medium supplemented with *n*-hexadecane or with a non-biodegraded oil sample as carbon source, inoculated with an oil sample derived from a moderately biodegraded petroleum reservoir. The library was screened on minimum medium agar plates supplemented with hexadecane. In all, 72 clones were isolated, and their performance was then quantified using GC–MS. Five clones displayed exceptional hexadecane biodegradation ability (>70%). Further characterisation of these hits was presented in two subsequent publications (Dellagnezze et al., 2014; Sierra-García et al., 2014). GC–MS and GC coupled with flame ionisation detection were used to test their biodegradation ability on *n*-alkanes (C14 to C33), different aromatic compounds (naphthalene, phenanthrene, methylphenanthrene) and isoprenoids (phytane and pristane). Two clones, harbouring inserts of 32 and 40 ORFs, were shown to be exceptionally efficient in degrading both linear and branched *n*-alkanes.

3.2.2. Hydrolases

3.2.2.1. Esterases.

Esterase is a generic term for a hydrolase that catalyses the cleavage of ester bonds. Esterases are of particular interest for detoxification of pesticides and herbicides. Indeed, organophosphorus, pyrethroids and carbamate pesticides and herbicides, which are known to affect the mammalian nervous system (Kuhr and Dorough, 1976; Sogorb and Vilanova,

2002; Singh and Walker, 2006), can be hydrolysed and detoxified by these enzymes. The first organophosphorus insecticide was used in 1937 (Dragun et al., 1984), while the synthesis and sale of carbamate pesticides started in the 1950s. In 2007, organophosphates accounted for 10% and 36% of total insecticides used in Europe and the US, respectively, and carbamates accounted for about 1,500 tons of active ingredients used in the agriculture sector (data obtained from the Food and Agriculture Organization of the United Nations, FAOSTAT and from Grube et al., 2011).

The new esterases presented in this review were isolated from metagenomic libraries constructed in *E. coli* from highly complex prokaryotic ecosystems, most of which were exposed to the targeted pollutant (Table 3). Enzymes that are both stable and active in extreme conditions are of considerable interest for many industrial applications. For instance, in order to access thermostable enzymes, Fan et al. (2012) constructed a library from topsoil samples of 'the turban basin' in China where the surface temperature is more than 82 °C during a certain period of the year. Tchigvintsev et al. (2014) searched for salt-resistant esterases in a cold, salty oil contaminated marine metagenome, and in phenanthrene and pyrene amended water from the Mediterranean Sea.

Screening of enzymes that act on highly toxic compounds like pyrethroids requires assays in confined media. Primary screening is thus often performed using solid plate assays containing classical esterase substrates, like the chromogenic compound 5-bromo-4-chloro-3-indolyl-caprylate (X-caprylate) (Li et al., 2008; Fan et al., 2012), or tributyrin, which degradation results in a clear halo around positive clones (Kambiranda et al., 2009; Kang et al., 2011; Sulaiman et al., 2012). To assess chain length specificity and enzyme stability in various temperatures, pH and salinity conditions, *para*-nitrophenyl esters of chain length ranging from C2 to C18 and alpha-naphthyl substrates of chain length ranging from C2 to C4, are commonly used in liquid medium containing cell lysates.

Enzymatic extracts of the hit clones are then incubated with the toxic compounds and analysed to quantify substrate consumption or product release. Several thermostable esterases expressed in *E. coli* and shown to be extremely efficient in breaking down pesticides such as malathion (an organophosphate analogous to pyrethroids) and other pyrethroids, were discovered using this approach (Li et al., 2008; Fan et al., 2012). In contrast, after primary screening in *E. coli*, Kambiranda et al. (2009) chose to express the discovered esterase in *Pichia pastoris*, to overcome the drawbacks of antibacterial activities of the main transformation product of chlorpyrifos, the main insecticide targeted in the study. The specificity of the purified enzyme towards a wide panel of organophosphorus insecticides was then assessed by HPLC. This versatile enzyme was able to degrade them all, with varying degrees of efficiency, making it a very interesting candidate for the remediation of highly toxic organophosphorus nerve agents.

In addition, numerous synthetic materials which are widely used in the insulation and plastic industries, like polyurethanes and aliphatic or aromatic polyesters, can also be degraded by some esterases. The breakdown of these compounds can easily be assessed using agar plates containing emulsions of the targeted insoluble polymers, a transparent halo appearing around the hit clones. Using this approach, Kang et al. (2011) found a polyspecific enzyme acting both on polyurethane and diethylene glycol adipate. Many other studies have targeted polylactic acid (PLA) and other aliphatic polyester degrading activities. After primary screening on solid plates containing PLA emulsions, three new esterases identified by Mayumi et al. (2008) were tested positive in liquid media on other aliphatic polyesters (PLA of different molecular weights, poly-butylene succinate (PBS), poly-butylene succinate-co-adipate (PBSA), poly-ethylene succinate (PES), polycaprolactone (PCL), polyhydroxybutyrate (PHB)), by measuring changes in the emulsion turbidity. Using the same approach, Tchigvintsev et al. (2014) found five new esterases that act on a large collection of monoester and polyester substrates, mixed as an emulsion in agarose plates. Sulaiman et al. (2012), determined the polyethylene terephthalate (PET) and the poly(ϵ -caprolactone)-degrading activities by measuring the weight loss of PET and PCL films after incubation with the subcloned and purified new cutinase.

In the same field of application, other functional metagenomic studies allowed the discovery of enzymes acting on phthalate esters, components also commonly used in the plastics industry. Jiao et al. (2013) mined an *E. coli* metagenomic library obtained from biofilms of dibutyl phthalate (DBP) (and other phthalate esters) from a wastewater treatment plant, maintained at low temperature (10 °C), by screening the clones on DBP spread in LB agar plates. The enzyme retrieved using this strategy was further characterised in liquid medium by HPLC/MS quantification of residual phthalate esters. The enzyme, which is capable of hydrolysing dipropyl- and dipentyl phthalates at low temperatures, could be used for bioremediation processes in wastewater treatment and for in situ degradation in contaminated cold environments.

3.2.2.2. Amidohydrolases.

The biological breakdown of organophosphorus insecticides requires various enzymatic activities. As described earlier in subsection 3.2.2.1. Esterases., an esterase able to cleave the P–O bound of chlorpyrifos was isolated from cow rumen, leading to 3,5,6-trichloro-2-pyridinol (TCP). This primary degradation product is more polar and more water soluble, but still classified as persistent and mobile by the US Environmental Protection Agency. Math et al. (2010) screened a cow rumen metagenomic library cloned in *E. coli* using TCP as sole carbon source. Only one positive clone was evidenced. The activity of the target enzyme on TCP was checked using both thin layer chromatography and HPLC analysis of *E. coli* cell supernatants.

As sequence annotation revealed that the enzyme responsible for this activity showed significant homology with ureases and amidohydrolases, authors postulated that this enzyme attacked the terminal amide bond of TCP. Nevertheless, further investigation is needed to confirm that this enzyme is an amide-linkage hydrolase.

3.2.2.3. Nitrilases.

Nitriles are components of plastics, polymers and herbicides which can be hydrolysed to their corresponding carboxylic acids and ammonia by nitrilases in a one-step reaction (Pace and Brenner, 2001). In this respect, mention should be made of two publications describing screening of metagenomic libraries to search for nitrilases. The libraries were constructed with bacterial DNA isolated from environmental samples collected from terrestrial and aquatic microenvironments worldwide (Robertson et al., 2004) or from oil-contaminated soil, wastewater treatment from a refinery and forest soils (Bayer et al., 2011). In both studies environmental DNA was cloned in *E.coli* and the libraries were screened using a liquid growth-dependent selection strategy with nitriles as sole nitrogen source (either adiponitrile or (R,S)-4-chloro-3-hydroxyglutaronitrile (Robertson et al., 2004) and 4-tolunitrile, 4-hydroxybenzonitrile, cinnamonitrile, or a mix of benzonitrile, glutaronitrile and α -methylbenzylcyanide (Bayer et al., 2011)). After sequencing and subcloning, around 140 putative nitrilases were further characterised for their specificity, especially their regioselective and enantioselective conversion properties. Indeed, the conversion of (poly)nitriles produces synthons of interest for the production of fine chemicals and pharmaceuticals. That was the goal of both metagenomic studies, although the approach was also mentioned as being useful for the discovery of enzymes involved in the catabolism of xenobiotic derived nitriles.

4. Conclusion and future goals

The rise of meta-omic technologies in the last decade has enabled to unlock the functional potential of uncultivable microbial biodiversity. In particular, many enzymes, exclusively produced by bacteria, have been discovered thanks to the activity-based exploration of metagenomes, especially those originating from highly polluted environments, and sometimes also exposed to extreme physical conditions, all with the aim of finding biocatalysts that are naturally adapted to industrial constraints. However, this approach is still limited i) by the lack of the facilities required to screen hundreds of thousands of clones for the desired functions, ii) by the fact that novel screens have to be designed for each enzymatic reaction, and, iii) when xenobiotic degradation activities are sought, by the need to handle large amounts of toxic compounds. These limits are currently being overcome thanks to the development of the different multi-step activity-based screening strategies cited in this review, which are often based on the use of non-toxic structural analogues for primary screening. Many of the enzymes that have been discovered

represent new tools for environmental biotechnology. Nevertheless, several challenges still have to be faced before we can accelerate the pace of the discovery of new functions by expanding the space of unexplored functional diversity. Firstly, although prokaryotic microbial communities are still the subject of numerous studies, the functions of uncultured eukaryotes are only rarely explored using activity-based metatranscriptomic approaches. However, eukaryotes play a fundamental role in many ecosystems, and some cultivable fungi are highly valuable bioremediation tools which have already been in use for a long time. New screening hosts are thus needed for expression of eukaryotic cDNA, to expand the panel of biocatalysts issued from the uncultivated fungal fraction. Secondly, the literature is very rich in functional metagenomics studies targeting oxidases and esterases with a wide range of specificities. But references to the discovery of proteases, which are likely to be effective for the hydrolysis of amide bonds (like those found for example in polyurethanes), are almost inexistent. In addition, too few studies aim to identify the role of uncharacterised protein families that are highly prevalent in polluted ecosystems, even though some of them could be considered as biomarkers of the functional status of the microbial community. Combining genomic and meta-omic data with data delivered by high-throughput activity-based screening approaches is now possible, and should accelerate the discovery of new functions and allow a better understanding and control of ecosystem functions. Finally, whatever their origin, using enzymes for bioremediation processes at low cost requires breaking the locks of enzyme expression in appropriate hosts, and, for some applications in open environments, taking legal regulations and ethical considerations into account.

Acknowledgements

This research was funded by the French Ministry of Education and Research (Ministère de l'Enseignement supérieur et de la Recherche) and the INRA metaprogramme MEM (project Metascreen, grant number: P10054).

Artwork

Fig. 3. Schematic representation of sequence-based and activity-based functional metagenomics approaches. 61

Table content

Table 2: Discovery of pollutant degrading enzymes using sequence based metagenomics approaches. All the cited studies provided biochemical proof of function for the hits obtained. 68

Table 3: Discovery of pollutant degrading enzymes by activity-based metagenomics. All screening studies were performed by using *E. coli* as cloning host, except for the work of Ono et al. (2007) and Nagayama et al. (2015), performed in *P. putida*. 69

References

- Aken, B.V., Correa, P.A., Schnoor, J.L., 2010. Phytoremediation of Polychlorinated Biphenyls: New Trends and Promises. *Environmental Science & Technology* 44, 2767–2776. doi:10.1021/es902514d
- Albertsen, M., Hansen, L.B.S., Saunders, A.M., Nielsen, P.H., Nielsen, K.L., 2012. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *ISME J* 6, 1094–1106. doi:10.1038/ismej.2011.176
- Ausec, L., Zakrzewski, M., Goesmann, A., Schlüter, A., Mandic-Mulec, I., 2011. Bioinformatic analysis reveals high diversity of bacterial genes for laccase-like enzymes. *PLoS ONE* 6, e25724. doi:10.1371/journal.pone.0025724
- Ballschmiter, K., Hackenberg, R., Jarman, W.M., Looser, R., 2002. Man-made chemicals found in remote areas of the world: the experimental definition for POPs. *Environ Sci Pollut Res Int* 9, 274–288.
- Bartossek, R., Nicol, G.W., Lanzen, A., Klenk, H.-P., Schleper, C., 2010. Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ. Microbiol.* 12, 1075–1088. doi:10.1111/j.1462-2920.2010.02153.x
- Bayer, S., Birkemeyer, C., Ballschmiter, M., 2011. A nitrilase from a metagenomic library acts regioselectively on aliphatic dinitriles. *Applied Microbiology and Biotechnology* 89, 91–98. doi:10.1007/s00253-010-2831-9
- Beloqui, A., Pita, M., Polaina, J., Martinez-Arias, A., Golyshina, O.V., Zumarraga, M., Yakimov, M.M., Garcia-Arellano, H., Alcalde, M., Fernandez, V.M., Elborough, K., Andreu, J.M., Ballesteros, A., Plou, F.J., Timmis, K.N., Ferrer, M., Golyshin, P.N., 2006. Novel Polyphenol Oxidase Mined from a Metagenome Expression Library of Bovine Rumen: BIOCHEMICAL PROPERTIES, STRUCTURAL ANALYSIS, AND PHYLOGENETIC RELATIONSHIPS. *Journal of Biological Chemistry* 281, 22933–22942. doi:10.1074/jbc.M600577200
- Bertrand, J.-C., Doumenq, P., Guyoneaud, R., Marrot, B., Martin-Laurent, F., Matheron, R., Moulin, P., Soulas, G., 2015. Applied Microbial Ecology and Bioremediation, in: Bertrand, J.-C. (Ed.), *Environmental Microbiology: Fundamentals and Applications*. Springer Netherlands, Dordrecht, pp. 659–753.
- Boubakri, H., Beuf, M., Simonet, P., Vogel, T.M., 2006. Development of metagenomic DNA shuffling for the construction of a xenobiotic gene. *Gene* 375, 87–94. doi:10.1016/j.gene.2006.02.027
- Brennerova, M.V., Josefiova, J., Brenner, V., Pieper, D.H., Junca, H., 2009. Metagenomics reveals diversity and abundance of meta-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environmental Microbiology* 11, 2216–2227. doi:10.1111/j.1462-2920.2009.01943.x
- Cardenas, E., Tiedje, J.M., 2008. New tools for discovering and characterizing microbial diversity. *Current Opinion in Biotechnology* 19, 544–549. doi:10.1016/j.copbio.2008.10.010
- Cerniglia, C.E., 1997. Fungal metabolism of polycyclic aromatic hydrocarbons: past, present and future applications in bioremediation. *J. Ind. Microbiol. Biotechnol.* 19, 324–333.
- Chouari, R., Le Paslier, D., Daegelen, P., Ginestet, P., Weissenbach, J., Sghir, A., 2003. Molecular Evidence for Novel Planctomycete Diversity in a Municipal Wastewater Treatment Plant. *Applied and Environmental Microbiology* 69, 7354–7363. doi:10.1128/AEM.69.12.7354-7363.2003
- Colleran, E., 1997. Uses of Bacteria in Bioremediation, in: *Bioremediation Protocols*. Humana Press, New Jersey, pp. 3–22.
- Dellagnezze, B.M., de Sousa, G.V., Martins, L.L., Domingos, D.F., Limache, E.E.G., de Vasconcellos, S.P., da Cruz, G.F., de Oliveira, V.M., 2014. Bioremediation potential of microorganisms derived from petroleum reservoirs. *Marine Pollution Bulletin* 89, 191–200. doi:10.1016/j.marpolbul.2014.10.003

- Dragun, J., Kuffner, A.C., Schneiter, R.W., 1984. A chemical engineer's guide to ground-water contamination — Part 1: Transport and transformations of organic chemicals. *Chemical engineering* 91, 64–70.
- Ekkers, D.M., Cretoiu, M.S., Kielak, A.M., Elsas, J.D. van, 2012. The great screen anomaly--a new frontier in product discovery through functional metagenomics. *Appl. Microbiol. Biotechnol.* 93, 1005–1020. doi:10.1007/s00253-011-3804-3
- Eyers, L., George, I., Schuler, L., Stenuit, B., Agathos, S.N., El Fantroussi, S., 2004. Environmental genomics: exploring the unmined richness of microbes to degrade xenobiotics. *Applied Microbiology and Biotechnology* 66, 123–130. doi:10.1007/s00253-004-1703-6
- Fang, Z., Li, T., Wang, Q., Zhang, X., Peng, H., Fang, W., Hong, Y., Ge, H., Xiao, Y., 2010. A bacterial laccase from marine microbial metagenome exhibiting chloride tolerance and dye decolorization ability. *Applied Microbiology and Biotechnology* 89, 1103–1110. doi:10.1007/s00253-010-2934-3
- Fang, Z.-M., Li, T.-L., Chang, F., Zhou, P., Fang, W., Hong, Y.-Z., Zhang, X.-C., Peng, H., Xiao, Y.-Z., 2012. A new marine bacterial laccase with chloride-enhancing, alkaline-dependent activity and dye decolorization ability. *Bioresour. Technol.* 111, 36–41. doi:10.1016/j.biortech.2012.01.172
- Fan, X., Liu, X., Huang, R., Liu, Y., 2012. Identification and characterization of a novel thermostable pyrethroid-hydrolyzing enzyme isolated through metagenomic approach. *Microbial Cell Factories* 11, 33. doi:10.1186/1475-2859-11-33
- FAOSTAT [WWW Document], n.d. URL <http://faostat3.fao.org/home/E> (accessed 7.5.15).
- Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A.P.M.D., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N., Golyshin, P.N., 2005. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora: Enzymatic diversity from bovine rumen metagenome. *Environmental Microbiology* 7, 1996–2010. doi:10.1111/j.1462-2920.2005.00920.x
- Festa, G., Autore, F., Fraternali, F., Giardina, P., Sannia, G., 2008. Development of new laccases by directed evolution: Functional and computational analyses. *Proteins: Structure, Function, and Bioinformatics* 72, 25–34. doi:10.1002/prot.21889
- Grube, A., Donaldson, D., Kiely, T., Wu, L., 2011. Pesticides Industry Sales and Usage 2006 and 2007 Market Estimates.
- Guermaz, S., Daegelen, P., Dauga, C., Rivire, D., Bouchez, T., Godon, J.J., Gyapay, G., Sghir, A., Pelletier, E., Weissenbach, J., Le Paslier, D., 2008. Discovery and characterization of a new bacterial candidate division by an anaerobic sludge digester metagenomic approach. *Environmental Microbiology* 10, 2111–2123. doi:10.1111/j.1462-2920.2008.01632.x
- Iwai, S., Chai, B., Sul, W.J., Cole, J.R., Hashsham, S.A., Tiedje, J.M., 2009. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *The ISME Journal* 4, 279–285. doi:10.1038/ismej.2009.104
- Jadeja, N.B., More, R.P., Purohit, H.J., Kapley, A., 2014. Metagenomic analysis of oxygenases from activated sludge. *Bioresour. Technol.* doi:10.1016/j.biortech.2014.02.045
- Jiao, Y., Chen, X., Wang, X., Liao, X., Xiao, L., Miao, A., Wu, J., Yang, L., 2013. Identification and Characterization of a Cold-Active Phthalate Esters Hydrolase by Screening a Metagenomic Library Derived from Biofilms of a Wastewater Treatment Plant. *PLoS ONE* 8, e75977. doi:10.1371/journal.pone.0075977
- Kambiranda, D.M., Asraful-Islam, S.M., Cho, K.M., Math, R.K., Lee, Y.H., Kim, H., Yun, H.D., 2009. Expression of esterase gene in yeast for organophosphates biodegradation. *Pesticide Biochemistry and Physiology* 94, 15–20. doi:10.1016/j.pestbp.2009.02.006
- Kang, C.-H., Oh, K.-H., Lee, M.-H., Oh, T.-K., Kim, B., Yoon, J.-, 2011. A novel family VII esterase with industrial potential from compost metagenomic library. *Microbial Cell Factories* 10, 41. doi:10.1186/1475-2859-10-41

- Kimura, N., Sakai, K., Nakamura, K., 2010. Isolation and Characterization of a 4-Nitrotoluene-Oxidizing Enzyme from Activated Sludge by a Metagenomic Approach. *Microbes and Environments* 25, 133–139. doi:10.1264/jsme2.ME10110
- Kuhr, R.J., Dorrough, H.W., 1976. *Carbamate insecticides: chemistry, biochemistry, and toxicology*. CRC Press, Cleveland.
- Li, G., Wang, K., Liu, Y.H., 2008. Molecular cloning and characterization of a novel pyrethroid-hydrolyzing esterase originating from the Metagenome. *Microbial Cell Factories* 7, 38. doi:10.1186/1475-2859-7-38
- Lu, Y., Yu, Y., Zhou, R., Sun, W., Dai, C., Wan, P., Zhang, L., Hao, D., Ren, H., 2011. Cloning and characterisation of a novel 2,4-dichlorophenol hydroxylase from a metagenomic library derived from polychlorinated biphenyl-contaminated soil. *Biotechnology Letters* 33, 1159–1167. doi:10.1007/s10529-011-0549-0
- Lu, Z., Deng, Y., Van Nostrand, J.D., He, Z., Voordeckers, J., Zhou, A., Lee, Y.-J., Mason, O.U., Dubinsky, E.A., Chavarria, K.L., Tom, L.M., Fortney, J.L., Lamendella, R., Jansson, J.K., D'haeseleer, P., Hazen, T.C., Zhou, J., 2012. Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME J* 6, 451–460. doi:10.1038/ismej.2011.91
- Math, R.K., Asraful Islam, S.M., Cho, K.M., Hong, S.J., Kim, J.M., Yun, M.G., Cho, J.J., Heo, J.Y., Lee, Y.H., Kim, H., Yun, H.D., 2010. Isolation of a novel gene encoding a 3,5,6-trichloro-2-pyridinol degrading enzyme from a cow rumen metagenomic library. *Biodegradation* 21, 565–573. doi:10.1007/s10532-009-9324-5
- Mayumi, D., Akutsu-Shigeno, Y., Uchiyama, H., Nomura, N., Nakajima-Kambe, T., 2008. Identification and characterization of novel poly(DL-lactic acid) depolymerases from metagenome. *Appl. Microbiol. Biotechnol.* 79, 743–750. doi:10.1007/s00253-008-1477-3
- Morimoto, S., Fujii, T., 2009. A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Applied Microbiology and Biotechnology* 83, 389–396. doi:10.1007/s00253-009-1992-x
- Nagayama, H., Sugawara, T., Endo, R., Ono, A., Kato, H., Ohtsubo, Y., Nagata, Y., Tsuda, M., 2015. Isolation of oxygenase genes for indigo-forming activity from an artificially polluted soil metagenome by functional screening using *Pseudomonas putida* strains as hosts. *Appl. Microbiol. Biotechnol.* doi:10.1007/s00253-014-6322-2
- Nogales, B., Timmis, K.N., Nedwell, D.B., Osborn, A.M., 2002. Detection and diversity of expressed denitrification genes in estuarine sediments after reverse transcription-PCR amplification from mRNA. *Appl. Environ. Microbiol.* 68, 5017–5025.
- Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y., Tsuda, M., 2007. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl. Microbiol. Biotechnol.* 74, 501–510. doi:10.1007/s00253-006-0671-4
- Pace, H.C., Brenner, C., 2001. The nitrilase superfamily: classification, structure and function. *Genome Biol.* 2, REVIEWS0001.
- Pham, V.H.T., Kim, J., 2012. Cultivation of unculturable soil bacteria. *Trends in Biotechnology* 30, 475–484. doi:10.1016/j.tibtech.2012.05.007
- Pieper, D.H., Martins dos Santos, V.A., Golyshin, P.N., 2004. Genomic and mechanistic insights into the biodegradation of organic pollutants. *Current Opinion in Biotechnology* 15, 215–224. doi:10.1016/j.copbio.2004.03.008
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., MetaHIT Consortium, Bork, P., Ehrlich, S.D., Wang, J., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi:10.1038/nature08821

- Rieger, P.G., Meier, H.M., Gerle, M., Vogt, U., Groth, T., Knackmuss, H.J., 2002. Xenobiotics in the environment: present and future strategies to obviate the problem of biological persistence. *J. Biotechnol.* 94, 101–123.
- Riser-Roberts, E., 1998. Remediation of petroleum contaminated soils: biological, physical, and chemical processes. Lewis Publishers, Boca Raton.
- Robertson, D.E., Chaplin, J.A., DeSantis, G., Podar, M., Madden, M., Chi, E., Richardson, T., Milan, A., Miller, M., Weiner, D.P., Wong, K., McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Snead, M.A., Keller, M., Mathur, E., Kretz, P.L., Burk, M.J., Short, J.M., 2004. Exploring Nitrilase Sequence Space for Enantioselective Catalysis. *Applied and Environmental Microbiology* 70, 2429–2436. doi:10.1128/AEM.70.4.2429-2436.2004
- Robinson, T., McMullan, G., Marchant, R., Nigam, P., 2001. Remediation of dyes in textile effluent: a critical review on current treatment technologies with a proposed alternative. *Bioresour. Technol.* 77, 247–255.
- Satar, R., Husain, Q., 2011. Catalyzed degradation of disperse dyes by calcium alginate-pectin entrapped bitter melon (*Momordica charantia*) peroxidase. *Journal of Environmental Sciences* 23, 1135–1142. doi:10.1016/S1001-0742(10)60525-6
- Shah, A.A., Hasan, F., Hameed, A., Ahmed, S., 2008. Biological degradation of plastics: a comprehensive review. *Biotechnol. Adv.* 26, 246–265. doi:10.1016/j.biotechadv.2007.12.005
- Sharma, N., Tanksale, H., Kapley, A., Purohit, H.J., 2012. Mining the metagenome of activated biomass of an industrial wastewater treatment plant by a novel method. *Indian Journal of Microbiology.* doi:10.1007/s12088-012-0263-1
- Sierra-García, I.N., Correa Alvarez, J., de Vasconcellos, S.P., Pereira de Souza, A., dos Santos Neto, E.V., de Oliveira, V.M., 2014. New hydrocarbon degradation pathways in the microbial metagenome from Brazilian petroleum reservoirs. *PLoS ONE* 9, e90087. doi:10.1371/journal.pone.0090087
- Silva, C.C., Hayden, H., Sawbridge, T., Mele, P., De Paula, S.O., Silva, L.C.F., Vidigal, P.M.P., Vicentini, R., Sousa, M.P., Torres, A.P.R., Santiago, V.M.J., Oliveira, V.M., 2013. Identification of Genes and Pathways Related to Phenol Degradation in Metagenomic Libraries from Petroleum Refinery Wastewater. *PLoS ONE* 8, e61811. doi:10.1371/journal.pone.0061811
- Singh, B.K., Walker, A., 2006. Microbial degradation of organophosphorus compounds. *FEMS Microbiol. Rev.* 30, 428–471. doi:10.1111/j.1574-6976.2006.00018.x
- Sogorb, M.A., Vilanova, E., 2002. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. *Toxicol. Lett.* 128, 215–228.
- Strachan, C.R., Singh, R., VanInsberghe, D., Ievdokymenko, K., Budwill, K., Mohn, W.W., Eltis, L.D., Hallam, S.J., 2014. Metagenomic scaffolds enable combinatorial lignin transformation. *Proceedings of the National Academy of Sciences* 111, 10143–10148. doi:10.1073/pnas.1401631111
- Suenaga, H., Ohnuki, T., Miyazaki, K., 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environmental Microbiology* 9, 2289–2297. doi:10.1111/j.1462-2920.2007.01342.x
- Sulaiman, S., Yamato, S., Kanaya, E., Kim, J.-J., Koga, Y., Takano, K., Kanaya, S., 2012. Isolation of a Novel Cutinase Homolog with Polyethylene Terephthalate-Degrading Activity from Leaf-Branched Compost by Using a Metagenomic Approach. *Applied and Environmental Microbiology* 78, 1556–1562. doi:10.1128/AEM.06725-11
- Sul, W.J., Park, J., Quensen, J.F., Rodrigues, J.L.M., Seliger, L., Tsoi, T.V., Zylstra, G.J., Tiedje, J.M., 2009. DNA-Stable Isotope Probing Integrated with Metagenomics for Retrieval of Biphenyl Dioxygenase Genes from Polychlorinated Biphenyl-Contaminated River Sediment. *Applied and Environmental Microbiology* 75, 5501–5506. doi:10.1128/AEM.00121-09
- Sutherland, T., Horne, I., Weir, K., Coppin, C., Williams, M., Selleck, M., Russell, R., Oakeshott, J., 2004. ENZYMATIC BIOREMEDIATION: FROM ENZYME DISCOVERY TO APPLICATIONS. *Clinical and Experimental Pharmacology and Physiology* 31, 817–821. doi:10.1111/j.1440-1681.2004.04088.x

- Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., Dore, J., Monsan, P., Remaud-Simeon, M., Potocki-Veronese, G., 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* 20, 1605–1612. doi:10.1101/gr.108332.110
- Taupp, M., Mewis, K., Hallam, S.J., 2011. The art and design of functional metagenomic screens. *Curr. Opin. Biotechnol.* 22, 465–472. doi:10.1016/j.copbio.2011.02.010
- Tchigvintsev, A., Tran, H., Popovic, A., Kovacic, F., Brown, G., Flick, R., Hajighasemi, M., Egorova, O., Somody, J.C., Tchigvintsev, D., Khusnutdinova, A., Chernikova, T.N., Golyshina, O.V., Yakimov, M.M., Savchenko, A., Golyshin, P.N., Jaeger, K.-E., Yakunin, A.F., 2014. The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. *Applied Microbiology and Biotechnology.* doi:10.1007/s00253-014-6038-3
- Theerachai, M., Emond, S., Cambon, E., Bordes, F., Marty, A., Nicaud, J.-M., Chulalaksananukul, W., Guieysse, D., Remaud-Siméon, M., Morel, S., 2012. Engineering and production of laccase from *Trametes versicolor* in the yeast *Yarrowia lipolytica*. *Bioresource Technology* 125, 267–274. doi:10.1016/j.biortech.2012.07.117
- Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.-P., Schleper, C., 2005. Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ. Microbiol.* 7, 1985–1995. doi:10.1111/j.1462-2920.2005.00906.x
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37 – 43.
- Vaghela, S.S., Jethva, A.D., Mehta, B.B., Dave, S.P., Adimurthy, S., Ramachandraiah, G., 2005. Laboratory Studies of Electrochemical Treatment of Industrial Azo Dye Effluent. *Environmental Science & Technology* 39, 2848–2855. doi:10.1021/es035370c
- van Hellemond, E.W., Janssen, D.B., Fraaije, M.W., 2007. Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl. Environ. Microbiol.* 73, 5832–5839. doi:10.1128/AEM.02708-06
- Vasconcellos, S.P. de, Angolini, C.F.F., García, I.N.S., Martins Dellagnezze, B., Silva, C.C. da, Marsaioli, A.J., Neto, E.V. dos S., de Oliveira, V.M., 2010. Reprint of: Screening for hydrocarbon biodegraders in a metagenomic clone library derived from Brazilian petroleum reservoirs. *Organic Geochemistry* 41, 1067–1073. doi:10.1016/j.orggeochem.2010.08.003
- Vergne-Vaxelaire, C., Bordier, F., Fossey, A., Besnard-Gonnet, M., Debard, A., Mariage, A., Pellouin, V., Perret, A., Petit, J.-L., Stam, M., Salanoubat, M., Weissenbach, J., De Bernardinis, V., Zapparucha, A., 2013. Nitrilase Activity Screening on Structurally Diverse Substrates: Providing Biocatalytic Tools for Organic Synthesis. *Advanced Synthesis & Catalysis* 355, 1763–1779. doi:10.1002/adsc.201201098
- Wang, Y., Chen, Y., Zhou, Q., Huang, S., Ning, K., Xu, J., Kalin, R.M., Rolfe, S., Huang, W.E., 2012. A Culture-Independent Approach to Unravel Uncultured Bacteria and Functional Genes in a Complex Microbial Community. *PLoS ONE* 7, e47530. doi:10.1371/journal.pone.0047530
- Wang, Y., Ren, H., Pan, H., Liu, J., Zhang, L., 2015. Enhanced tolerance and remediation to mixed contaminants of PCBs and 2,4-DCP by transgenic alfalfa plants expressing the 2,3-dihydroxybiphenyl-1,2-dioxygenase. *Journal of Hazardous Materials* 286, 269–275. doi:10.1016/j.jhazmat.2014.12.049
- Wang, Z., Zhang, X.-X., Huang, K., Miao, Y., Shi, P., Liu, B., Long, C., Li, A., 2013. Metagenomic Profiling of Antibiotic Resistance Genes and Mobile Genetic Elements in a Tannery Wastewater Treatment Plant. *PLoS ONE* 8, e76079. doi:10.1371/journal.pone.0076079
- Watanabe, K., 2001. Microorganisms relevant to bioremediation. *Curr. Opin. Biotechnol.* 12, 237–241.
- Xing, M.-N., Zhang, X.-Z., Huang, H., 2012. Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. *Biotechnology Advances* 30, 920–929. doi:10.1016/j.biotechadv.2012.01.021

- Yamada, T., Waller, A.S., Raes, J., Zelezniak, A., Perchat, N., Perret, A., Salanoubat, M., Patil, K.R., Weissenbach, J., Bork, P., 2012. Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours. *Molecular Systems Biology* 8. doi:10.1038/msb.2012.13
- Yam, K.C., van der Geize, R., Eltis, L.D., 2010. Catabolism of Aromatic Compounds and Steroids by *Rhodococcus*, in: Alvarez, H.M. (Ed.), *Biology of Rhodococcus*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 133–169.
- Ye, M., Li, G., Liang, W.Q., Liu, Y.H., 2010. Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. *Applied Microbiology and Biotechnology* 87, 1023–1031. doi:10.1007/s00253-010-2507-5
- Yu, K., Zhang, T., 2012. Metagenomic and Metatranscriptomic Analysis of Microbial Community Structure and Gene Expression of Activated Sludge. *PLoS ONE* 7, e38183. doi:10.1371/journal.pone.0038183
- Zapras, A., Liu, Y.-J., Liu, S.-J., Drake, H.L., Horn, M.A., 2009. Abundance of Novel and Diverse tfdA-Like Genes, Encoding Putative Phenoxalkanoic Acid Herbicide-Degrading Dioxygenases, in Soil. *Applied and Environmental Microbiology* 76, 119–128. doi:10.1128/AEM.01727-09
- Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E.J., Short, J.M., Keller, M., 2002. Cultivating the uncultured. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15681–15686. doi:10.1073/pnas.252630999

Chapter III:

The bovine rumen ecosystem

1. Physiological properties of the rumen

Grazing animals have a four-chambered stomach, rumen being the biggest part with an average volume corresponding to 80% of the total stomach. A highly complex microbial ecosystem develops and survives in this compartment by extracting nutrients coming mostly from dietary plant constituents. Since mammals do not produce the carbohydrate active enzymes (CAZymes (Lombard et al., 2014)) able to hydrolyse most of plant polysaccharides, it is this symbiotic microbial ecosystem, present in the digestive tract that do so, generating energy for themselves and their host: they produce volatile fatty acids which are directly absorbed through the rumen papillae to supply up to 80% of the energy needed by the ruminant (Hall and Silver, 2009).

According to a study of Hess et al. (2011), who sequenced 268 Gbp of bovine rumen metagenomic DNA, several dozen thousands CAZymes (around 100 putative CAZyme encoding genes per Gbp of sequence) are produced by the microbes adherent to plant fibre in cow rumen, since most of the glycoside hydrolysis occurs in the biofilm adhering to the ingested feed particles. These enzymes, which include glycoside hydrolases, carbohydrate esterases, polysaccharide lyases and, probably (as they still have not been searched in the ruminal microbiome to our knowledge), LPMOs (lytic polysaccharide monooxygenases (Levasseur et al., 2013; Agger et al., 2014)), mostly catalyse the degradation of the plant cell wall lignocellulosic components, together with plant energy storage polysaccharides (resistant starch and inulin) and, in a minor extent, microbial glycans that are also present in fermented feeds.

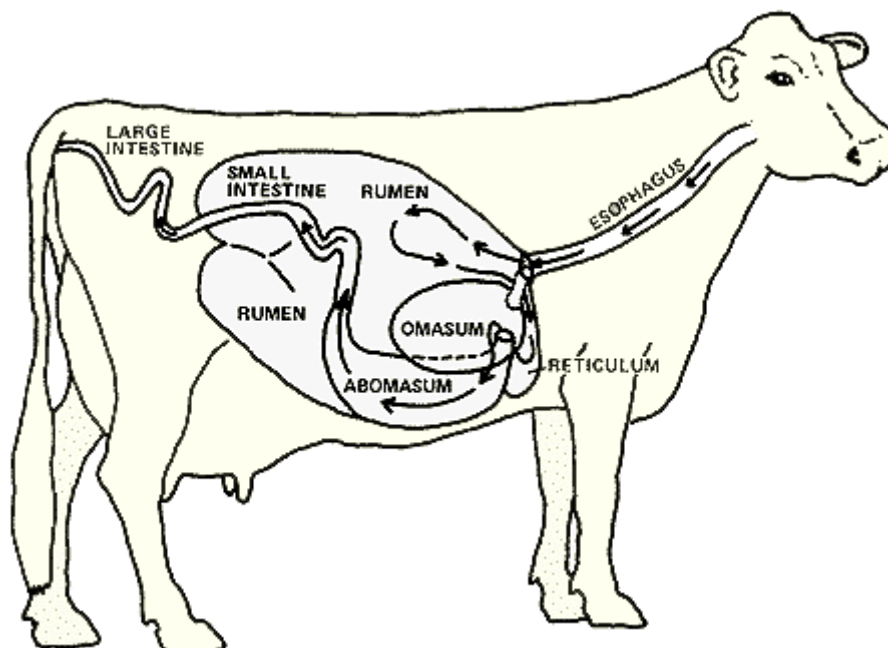


Figure 4: Schematic representation of the bovine digestive tract

As mentioned before, the bovine digestive tract is made of different important parts: the mouth cavity, oesophagus, a complex stomach (made of the rumen, the reticulum, the omasum and the abomasum) and intestines (small intestine, including duodenum, and the large intestine, consisting of the caecum, colon and rectum) (Figure 4). Other glands are also important in digestion: pancreas, spleen, and liver.

In the first part of the digestive tract, the mouth and the teeth configuration allows for the chewing of large amount of fibrous feed. Inside the mouth, the feed particle size is thus reduced and some food components begin to be digested by the enzymes present in the saliva. Bovine lipase indeed hydrolyses short chain fatty acids, which are only a small fraction of food components. Saliva also helps to keep the rumen at the right neutral pH for the growth of bacteria. In the mouth, the digestion is thus mostly mechanical, while in the rumen, it is both mechanical and enzymatic (Agarwal et al., 2015; Krause et al., 2013).

Digestion by ruminants is mostly done in the rumen. Its capacity reaches up to 150 litres for adults, which allows storage of a large quantity of food ingested by the animal throughout the day, during 30 to 50 hours depending on the diet. The bovine rumen also contains billions of symbiotic microorganisms of which the functional diversity allows digestion of different feed (grass, hay, corn, silage...) by the ruminant. Most of the digestion is done by bacteria and protozoa (for more details, see part 3). They produce energy needed for the ruminant, vitamins (B, C and K) along with proteins, which provide essential amino acids and nitrogen. Lipolysis and biohydrogenation of dietary lipids also happen in the rumen, converting them in saturated fatty acids (palmitic, stearic acids). When animals are healthy, the rumen temperature oscillates between 38 and 41°C, while physiological pH, one of the most variable factors, varies from 5.5 to 6.9 (Puniya et al., 2015).

During rumination, cows chew, swallow and regurgitate their food again and again, in order to cut it into small particles to digest the fibres more efficiently. Cows also eructate in order to release the carbon dioxide and methane produced during carbohydrate fermentation.

The next part of the digestive tract, the omasum, allows the filtering of large particles back into the rumino-reticulum, and the smaller particles to go on to the abomasum. The omasum also helps in water absorption.

The abomasum is known as the “true stomach”, producing hydrochloric acid and digestive enzymes to start protein digestion. It is comparable to the stomach of non-ruminants (Agarwal et al., 2015; Hall and Silver, 2009).

The digestion rate changes depending on numerous factors, like the retention time of feeds in the rumino-reticulum, the mastication and the mixing of saliva along with its

buffering action, the flow rate of the digesta to the omasum and the lower part of the digestive tract, along with the continuous removal of some metabolites (volatile fatty acids, methane, carbon dioxide, ammonia) (Agarwal et al., 2015).

2. Bovine nutrition

Polysaccharides are found in high quantity in the diet of bovines, being their first energy source. They mostly come from plant cell wall, of which structure is dynamic and complex. It allows cohesion of cells (Burton et al., 2010), and also has a protective role, by being the physical barrier against pathogens (Underwood, 2012).

Plant cell wall is composed of a majority (up to 90%) of polysaccharides (cellulose, hemicelluloses, pectins) and of lipids, proteins and, lignin. It is made of many layers: the middle lamella, rich in pectins, is the first to be synthesised; the primary cell wall, made of cellulose fibres, hemicelluloses and pectins, linked by different kinds of links that can be covalent or not, are synthesised during cell growth (Figure 5); the secondary cell wall is synthesised after growth and is composed of a compact matrix of hemicellulose, crystalline microfibrils of cellulose, and lignin that gives the rigidity and a great mechanical and chemical resistance.

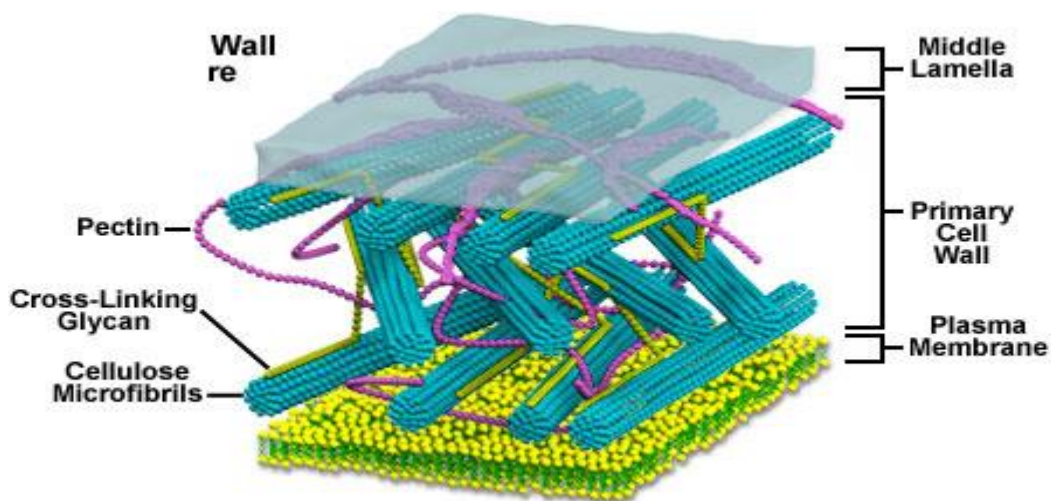


Figure 5: Composition and organisation of primary cell wall (from <http://micro.magnet.fsu.edu/cells/plants/cellwall.html>)

The different components of plant cell walls are described below:

- Cellulose is a linear homopolymer of D-glucose with β -1,4 linkages. It is one of the most abundant polymers on Earth. Polymerisation degree varies between 250 and 15,000, depending on its origins. Inside plant cell wall, linear chains are organised in microfibrils thanks to intra and intermolecular hydrogen linkages, which guide their organisation: the non-organised amorphous zones and the crystalline ones, where

the hydrogen linkages are more numerous, making the crystalline zones more resistant to enzymatic and chemical degradation.

- Hemicelluloses are heteropolymers composed of pentoses (D-xylose and L-arabinose), hexoses (D-glucose, D-galactose and D-mannose) and sugar acids (acetic acid, ferulic acid, 4-O-methylglucuronic acid...). Depending on their primary structure, they are classified in four categories, which proportion depends on their origin:
 - Xylans: in this case, the backbone is made of β -1,4-linked xylosyl residues, most of the time acetylated, and which can be ramified. Presence of ramifications and their structure are the basis of another classification: glucuronoxylans are ramified with glucuronic acids and/or one of their reaction product, 4-O-methylglucuronic acid; arabinoxylans are substituted with L-arabinofuranosyl residues; glucuronoarabinoxylans or arabinoglucuronoxylans contain both previous substitutions, and named depending on the proportion of each one; heteroxylans are ramified with arabinosyl, xylosyl, galactosyl residues, and glucuronic acid in various amounts. The arabinosyl residues can also be substituted by phenolic acids (ferulic acid, *p*-coumaric acid), which allows the connexion of two free hemicellulose chains, which can then be covalently linked, by ether linkage, to lignin.
 - Xyloglucans: the backbone is made of β -1,4-linked glucosyl residues, ramified with xylosyl units linked to the C6 position. Depending on the plant origin, xylosylation varies from 30 to 75%. Xylosyl residues can also be substituted by β -galactosyl or α -L-arabinofuranosyl residues.
 - Mannans: they are classified in 2 sub categories depending on their backbone structure. Galactomannans have a backbone of β -1,4-linked mannosyl residues, with D-galactosyl ramifications on the C6 position; glucomannans have a backbone with both β -1,4-linked D-mannosyle and D-glucosyl residues, with D-galactosyl ramifications on the C6 position. Moreover, these different mannan chains, especially the most linear ones tend to self-associate, leading to insolubility and crystallinity. As the best example, even if it is not a bovine dietary constituent, let's cite the siphonous green algae of the families *Codiaceae* and the *Dasycladaceae*, of which crystalline mannan microfibrils act as a functional replacement for cellulose (Whitney et al., 1998)
 - (1,3)(1,4)- β -D-glucans or mixed-linkage glucans: it is a defining feature of grass cell wall. They are unsubstituted chains of β -glucopyranosyl monomers linked through (1,3) (around 30 % of the total linkages) and (1,4) linkages (around 70 %). In the aleurone and starchy endosperme of grains, including barley, wheat, and oat, mixed-linkage glucans are also found in high concentrations and are thought to act as storage carbohydrates (Kiemle et al., 2014).

- Pectins: these polysaccharides are rich in galacturonic acids linked with β -1,4 linkages. They have an important role for hydration of cell wall and its wholeness, thanks to their gelification ability. They also enable an easier organisation of parietal components along with the diffusion of ions and small molecules. Depending on their composition, they are classified in different subcategories covalently associated to create the pectic matrix:
 - Homogalacturonans: they are composed of a linear chain of galacturonic acids, which can be methylesterified or O-acetylated.
 - Substituted galacturonans: while they have a similar backbone to homogalacturonans, the ramification is different: apiogalacturonans with D-apiofuranose; xylogalacturonan with xylose; rhamnogalacturonan II with up to 12 different sugars and 20 different linkages.
 - Rhamnogalacturonans I: their backbone consists of repeats of the [α -D-Galacturonan-1,2- α -L-rhamnose-1,4-] disaccharide with ramified oligosaccharides that are specific to the kind of cell and their growth stage.
 - Arabinan: they have a backbone of (1,5)-linked α -L-arabinofuranosyl units. The neutral arabinans and arabinogalactans are linked to the acidic pectins. They would promote wall flexibility and would bind to the surface of cellulose (Cosgrove, 2005)
- Lignin: this complex polymer is specific to the secondary cell wall, and confers to it its hydrophobicity, while increasing its rigidity. Because it is difficult to degrade, it helps protecting the cells from pathogen microorganisms. It is an amorphous aromatic polymer predominantly composed of three hydroxycinnamoyl alcohols (conyferylic alcohol, sinapylic alcohol and *p*-coumarylic alcohol), of which the proportion depends on the cell type and plant species. Lignins are only a little depolymerised inside the rumen. The presence of high amounts of lignin in the cell walls of cow dietary compounds limits carbohydrate fermentation, volatile fatty acid and microbial biomass production, along with reducing feed intake and prolonging retention time, which induce low energy value for animal feeding (Calabrò, 2015).
- Extensines: they act during growth and plant development, being involved in differentiation and morphogenesis control, along with cell-cell interactions. They can also, by auto-assembly, create a net able to strengthen and stabilise the cell wall.
- Enzymatic and non-enzymatic proteins: respectively oxidative or hydrolytic and expansins which intervene in metabolism and remodelling of the cell wall.
- Lipidic components: they can be found on the surface of secondary cell wall (suberin, cutin, wax) and create a protective layer against pathogens attacks and limit water loss.

Starch is also one of the main plant cell components, and thus an important constituent of the bovine diet. This energy storage polysaccharide is made of around 20-30% of amylose, a helical polymer made of α -1,4-linked D-glucosyl units, and of 70-80% of amylopectin, a

soluble polysaccharide consisting of α -1,4-linked D-glucosyl units branched with α -1,6 linkages every 24 to 30 glucosyl units.

Inulin is another storage polysaccharide, even much less abundant than starch. It is found in many kinds of plants, mostly commonly extracted from chicory, and is composed of a backbone of β -1,2-linked fructosyl residues, with a sucrose molecule at the reducing end. Its polymerisation degree ranges from 2 to 60.

Finally, phytic acid is also an important component of bovine diet, as it is the primary storage form of phosphorus and inositol in the majority of seeds. It is a complex of calcium or magnesium with myo-inositol. It is an anti-nutritional component of plant-derived feeds and it can form insoluble salts with different minerals (zinc, magnesium...) or complexes with proteins and proteolytic enzymes.

3. Microbial diversity in the bovine gastrointestinal tract

Since only around 11% of the rumen bacterial species seem to be culturable (Wu et al., 2012), monogenic (16S rRNA) metagenomics is the most appropriate method to evaluate the microbial diversity along the bovine gastrointestinal tract.

16S rDNA (or 18S for fungi and protozoa), which encodes part of the ribosome needed for protein synthesis, is indeed composed of highly conserved DNA motifs, interrupted by nine separate variable regions, that are specific to bacterial and archaeal species. The development of next generation sequencing enabled the characterisation of complex communities based on 16S rDNA sequencing (Fouts et al., 2012; Highlander, 2014; Jami and Mizrahi, 2012; Sadet et al., 2007).

Since the rumen is predominantly a strict anaerobic environment, the natural rumen microbiota is composed of anaerobic microorganisms that can be divided in three main superkingdoms: Bacteria, Archaea containing the methanogens, and Eukaryota composed of protozoa and fungi. However, since the rumen is open to the external environment and because of the continual flow of materials going in and out of the rumen, aerobic microorganisms can be found, coming, for example from the food they ingest. The commensal microbial community must be at least facultative anaerobe, though the proportion of facultative/strict anaerobe is, to our knowledge, never explicitly mentioned (Highlander, 2014).

At birth, the calf rumen microflora is considered sterile, rapidly colonized by aerobic and facultative anaerobic microbial strains, close to the birth, with a gradual decrease and replacement by anaerobic strains over time (Figure 6). Thanks to a pyrosequencing and quantitative real time PCR approach, the dominant phyla, no matter the age, were

determined by Jami et al to be Bacteroidetes, Firmicutes and Proteobacteria. However, the difference is also in the composition of said phyla: for example Bacteroidetes, inside the rumen of newborns, are predominantly from the genus *Bacteroides*, while for older animal's rumen, they are almost exclusively from the genus *Prevotella* (Jami et al., 2013).

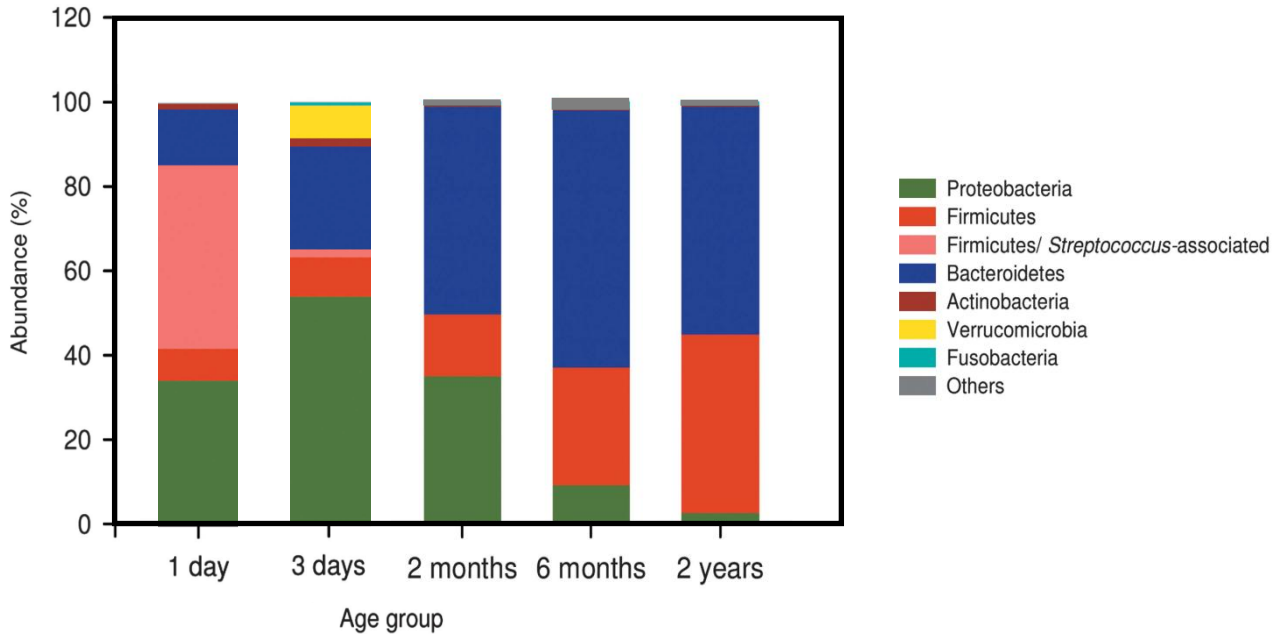


Figure 6: Phylum level composition depending on the age of the animals (adapted from Jami et al., 2013)

In the study of Peng et al., the microbial community of the reticulum and the omasum of 3 adult cows, fed with dried distillers' grains containing soluble and corn silage, has been studied for the first time (Peng et al., 2015), along with the rumen. They were able to identify the seven most abundant phyla of the three chambers of a cow stomach, which correspond to more than 98% of the bacteria identified (Figure 7).

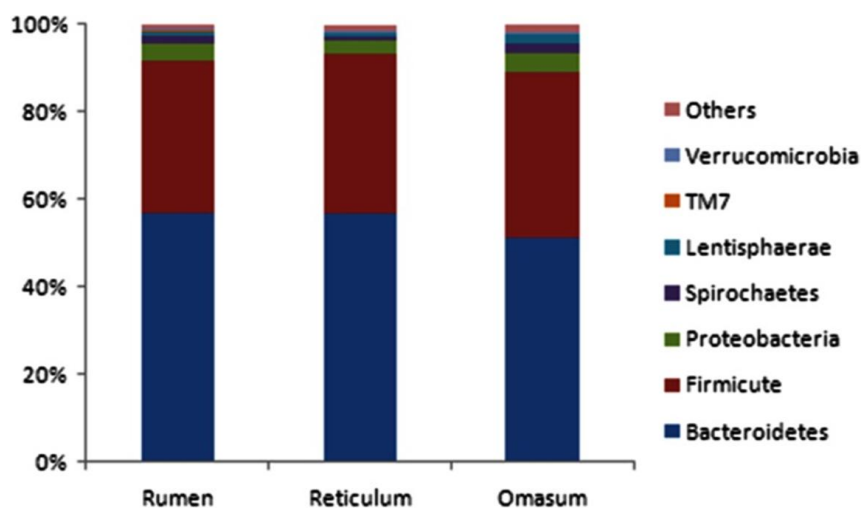


Figure 7: relative distribution of the most abundant phyla in three chambers of cow stomach (adapted from Peng et al (2015))

In the rumen of adult cows, as can also be seen in other studies (Jami et al., 2013), the most abundant phylum is Bacteroidetes, which include Gram-negative bacteria known to be one of the main actors in the polysaccharide catabolism. The predominant genus is also, as previously seen, *Prevotella*, known for its importance in digestion, one of the most prevalent species, *Prevotella ruminicola*, being able to digest hemicelluloses, cellulose, pectin, and proteins. The second most abundant phylum is Firmicutes, with *Succinivibrio* as the predominant genus. The latter is particularly specific, as it is unable to ferment amino acids, carbohydrates, mono-, di-, and tricarboxylic acids other than succinate (which is converted to propionate). Moreover, bacteria belonging to the *Succinivibrio* genus are not proteolytic, do not produce a catalase or urease, nor reduced nitrate. *Butyrivibrio* is less prevalent in the rumen than in the other chambers, but it has been known to be very important for the degradation of proteins, urea, cellulose and hemicelluloses. *Treponema*, a genus from the Spirochaetes phylum, was found in the three chambers of the stomach, as part of the ten most abundant in all three. In the rumen, they are known to efficiently degrade plant polysaccharides (Peng et al., 2015).

The abomasal microbiota was studied in the study of Li and al, (Li et al., 2011). As can be seen in Figure 8, the predominant phyla are the same as the ones found in the rumen: Bacteroidetes, Firmicutes and Proteobacteria. However, the Actinobacteria, Verrucomicrobia and Fusobacteria cannot be found in the abomasum. On the other hand, *Fibrobacter* and Spirochaetes were found in the abomasum and not notably in the rumen. These differences could be due to the pH differences between the rumen and the abomasum (pH around 6.0 for the rumen, around 2.0 for the abomasum).

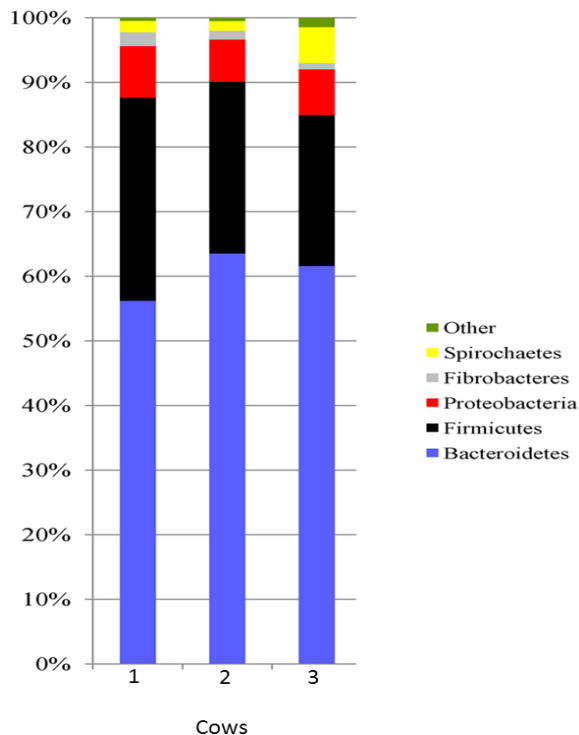


Figure 8: Relative distribution of phyla of the cow abomasum (adapted from Li et al, 2011)

Since most of the plant component degradation also happens in the rumen, the most effective enzymes for polysaccharide degradation should be found there, which explains why the ruminal ecosystem is the most studied for mining the gut microbiome for CAZymes (van Vliet et al., 2007).

If diversity of the bovine gut microbiota has been extensively studied these last years, some differences are frequently observed between the studies, at the phyla and genera levels. They are due to differences in the sampling and sequencing technologies, the samples themselves (liquid and solid fractions), or to the diet of the animals (Ramirez et al., 2012). The different microbial constituents of the ruminal ecosystem are presented below:

3.1. Bacteria

They are the most actively involved in plant fibre degradation. Bacteria represent 40 to 50 of the total microbial mass, or 10^{10} - 10^{11} bacteria per gram of rumen content, and as such are the predominant microbes in the rumen. Among the cultivated ruminal species, *Fibrobacter succinogenes*, *Ruminococcus albus*, and *Ruminococcus flavefaciens* are known to degrade cellulose, while hemicelluloses are predominantly degraded by *Butyrivibrio fibrisolvens*, *Prevotella ruminicola*, *Ruminococcus albus*, and *Ruminococcus flavefaciens*, these two last species being important H₂ producers during plant cell wall degradation. *Eubacterium ruminantium* can also ferment extensively glucose, cellobiose and fructose. Major pectin fermenters are *Butyrivibrio fibrisolvens*, *Prevotella ruminicola*, *Lachnospirousa multiparus*. And known starch digesters are *Butyrivibrio fibrisolvens*, *Prevotella ruminicola*, *Streptococcus bovis*, *Ruminobacter amylophilus*, *Succinimonas amylolytica*, *Selenomonas ruminantium*, together with some *Clostridium* species and some strains of *Fibrobacter succinogenes* (Zhou et al., 2015).

All these bacteria are present in the rumen, in different proportions depending, in particular, on diet (see part 3.6). These bacteria can either freely float in the liquid phase, be attached to the feed particles (like many fibrolytic bacteria), or to the rumen epithelium (Choudhury et al., 2015). The ruminal epithelium is indeed covered with papillae that facilitate the absorption of fermentation products, like volatile fatty acids. Using culture based methods and microscopy (Cheng et al., 1979; McCowan et al., 1978), it was found that epimural bacteria (*Micrococcus*, *Staphylococcus*, *Streptococcus*, *Corynebacterium*, *Lactobacillus*, *Fusobacterium*, *Propionibacterium* species, along with other unidentified anaerobic species) are taxonomically different from those present in the rumen fluid and/or solid particles. Using recent molecular based technics (PCR-DGGE, 16S rRNA sequence analysis, qRT-PCR), the difference between microbial community inside the rumen and attached to the ruminal tissue has been confirmed (Li et al., 2012). However, although the bacterial diversity has been defined, it is not the case accurately for the bovine epimural bacterial population.

3.2. Bacteriophages

They are the second most present components, in terms of number, of the rumen microbial ecosystem, since 10^8 - 10^9 bacteriophages are counted per gram of ruminal content. They are considered as obligate pathogens for bacteria. Bacteriophages can help in bacterial mass turnover in the rumen that can happen after a radical change in diet, which makes available bacterial proteins to the host as an amino acid source. They are specific for some bacteria present in the rumen, and can be used to kill or remove unwanted bacterial strains, like *Streptococcus bovis*, and methanogens, to be used, for example, to lower the production of methane (Ross et al., 2013). Berg Miller et al. were able to identify the most abundant lytic phages in the rumen. They belong to the following families: *Siphoviridae*, *Myoviridae*, *Podoviridae*, *Herpesviridae*, *Phycodnaviridae*, along with many unclassified ones (Berg Miller et al., 2012; Gilbert and Klieve, 2015). However, little information is currently available on the rumen virome and its interaction with ruminal microbiome.

3.3. Methanogens

Classified in the separate superkingdom of Archaea, they count for 10^7 - 10^8 cells per grams of rumen content. In the majority of ruminants, *Methanobacteriales* (including *Methanobrevibacter* spp., *Methanobacterium* spp., *Methanosphaera* spp.) are the most prevalent, *Methanomicrobiales* (including *Methanomicrobium* spp.) being the second most important. The least in number are *Methanosarcinales* and *Methanococcales*. In bovine rumen, some methanogens are associated with protozoa, depending on the species. Using 16S rRNA analyses, Tymensen et al were able to find out that *Methanobrevibacter* are predominant in methanogens associated with protozoa, while *Methanomicrobium* can be found as free-living methanogens (Agarwal et al., 2015; Tymensen et al., 2012). Methanogens can produce 100% of the rumen methane emissions. Methanogens thus became interesting targets in the context of global warming. The amount of methane produced by ruminants indeed represents about one-quarter of anthropogenic emission of methane. Many methods have been tried to lower methane emission by cattle (Janssen and Kirs, 2008). For example, plants containing secondary metabolites which act as methane inhibitors can be used as feed additives. Another approach is based on the use of microbial feed additives, though it has not been as successful as the first one. Some of the strains used are: *Saccharomyces cerevisiae*, for which the results of methane emission decrease are quite chaotic (from 0 to 58% inhibition); *Propionibacterium* species producing propionate, since the production of propionate is inversely proportional to the amount of methane, both production pathway consuming hydrogen; acetogens, which uses hydrogen for the reduction of carbon dioxide to acetate and can bring 5 to 10% of energy intake by the animal back to the energy cycle of said animal instead of methanogenesis. The problem of the latter is that acetogen affinity to hydrogen is lower than the one for methanogens in the rumen. Defaunation, meaning the removal of protozoa in the rumen, is another method; however data compilation showed that more study is required to help fully explain the relationship between defaunation and methanogenesis (Morgavi et al., 2010). As previously mentioned, increase of propionate concentration could be a way to lower methane production. Organic acids have thus been tested as propionate precursors. However, they are expensive, and the

results obtained variable. Use of halogenated methane analogues is another approach, to block the terminal enzyme of methanogenesis, the methyl CoM reductase. The problem would be the possible short term effect of such a technique, the rumen treated being able to rapidly adapt to the inhibitory effect. The fact that they can be toxic to the host as well as Humans depending on the concentration is another drawback. Another method is the use of commercially available ionophores that would inhibit hydrogen producing bacteria. However, like halogenated methane analogues, the risk of the host's microbes adaptation to these compounds is too high, along with the possibility to recover residues in animal products. The last method would be to add inorganic salt to the diet, since nitrate and sulphate reducing bacteria already present in the rumen use hydrogen to reduce substrates to nitrite and H₂S that become toxic at high concentration for animals. This means that this method would not be the optimal one either (Kamra et al., 2015). However, to find the optimal method to decrease methane emission, a more in depth understanding of the functional pattern of rumen methanogens is essential.

3.4. Protozoa

Rumen protozoa represent only a small part of the microbial population in terms of number (10^4 - 10^6 per gram of rumen content). In terms of total mass, however, it can account for up to 40-50% of the microbial mass. The two major groups of rumen *Ciliates* are: *Entodiniomorpha* (*Entodinium simplex*, *Entodinium caudatum*, *Eudiplodinium maggii*, *Metadinium medium*, *Diploplastron affine*, *Polyplastron multivesiculatum* and *Epidinium ecaudatum*) and *Vestibuliferida*, previously called *Holotricha* (*Isotricha prostoma*, *Isotricha intestinalis* and *Dasytricha ruminantium*) (Agarwal et al., 2015; Choudhury et al., 2015; Ricard et al., 2006).

They play a very important role in rumen fermentation. They help ferment lignocellulosic parts of the food intake into volatile fatty acids. In their study, Ricard et al were able to identify many carbohydrate catabolic enzymes (like xylanases, cellulases, β -hexosaminidases, pectate lyases, polygalacturonases, cellobiose phosphorylases, carbohydrate esterases, several glycosidases, fructokinases and aldose epimerase precursors) , of which the encoding genes were acquired by horizontal gene transfer between *Bacteria* and *Ciliates* (Ricard et al., 2006).

Most importantly, protozoa protect easily fermentable carbohydrates, including starch, from the bacteria that could use them, by being incorporated in protozoa biomass. Protozoa can indeed also engulf bacteria and feed particles, along with being able to degrade carbohydrates, lipids and proteins. As such, carbohydrates are released slowly over time to give the host a constant energy supply in the form of short chain volatile fatty acids, and avoid an overproduction of organic acids directly after feeding that could be detrimental to the host (Williams and Coleman, 1992).

3.5. Anaerobic fungi

Anaerobic fungi are the smallest components (in number, 10^3 - 10^5 per gram of rumen content) of the rumen microbial ecosystem, yet they are thought to be the most effective for the degradation of fibres. Indeed, because they are able to produce many different glycoside-hydrolases (cellulases, hemicellulases, amylases, glycosidases, glycoside esterases, pectinases), they are key degraders of plant components. Because they are anaerobic fungi, the energy they use comes from the fermentation of carbohydrates, by a mechanism where the energy source is both the electron acceptor and donor. They do not have mitochondria but hydrogenosomes which seems to be specialised or derived mitochondria that contain unusual enzymes like hydrogenases and pyruvate/ferredoxin oxidoreductases (not the usual mitochondrial pyruvate dehydrogenase). Because of their low number, they were not extensively studied for as long as other components of the ruminal ecosystem, though this has been changing for some years (Robert J. Gruninger et al., 2014). Up until now, only six genera have been described (*Caecomyces*, *Piromyces*, *Neocallimastix*, *Anaeromyces*, *Orpinomyces*, *Cyllamyces*) and some uncharacterised isolates have been mentioned. Even though fungi are located throughout the gastrointestinal tract, 90% are found in the reticulo-rumen, with 10% along the rest of the gastrointestinal tract (Agarwal et al., 2015; Choudhury et al., 2015; Davies et al., 1993).

3.6. Structural and functional modulation of the ruminal microbiota

Ruminal microbiota structure varies depending on many factors, linked to the animal (diet, feed intake, frequency of feeding, age of the animal, antibiotic use, health of the host, stress level, ...), or not (geographic location, season, photoperiod, environment...).

The most known way to affect ruminal microbiota is diet, which directly impacts the balance among species. For example, a high fibre diet increases the population of *Ruminococcus albus*, *Ruminococcus flavefaciens* and *Fibrobacter succinogenes*, even though if a cow is on an all straw diet, the fibre digesters still never account for more than 25% of the rumen bacterial population (Choudhury et al., 2015). Works on rumen manipulation suggest that bacterial diversity is mostly dependant on diet or location rather than species of host (Choudhury et al., 2015; Zhou et al., 2015). Not many studies can be found where animals are fed predominantly on wheat straw, like in this PhD project. One of them, by Chaudhary et al, showed the difference in quantification of the major microbial species of the rumen of Indian cattle and buffalo, fed on wheat straw diet (concentrate: roughage ratio = 40:60), using real time PCR method. While no significant change in total bacterial and methanogens population could be seen, it is not the same for the other microbes. Indeed, the amount of fungi and protozoa seems to be higher in cattle than in buffalo. A significant difference could also be seen when focusing on cellulolytic bacteria: the copy number of *Fibrobacter succinogenes* and *Ruminococcus albus* was higher in buffalo than in cattle, while for *Ruminococcus flavefaciens*, the highest number could be found in cattle. While no

explanation can be given for the moment, this shows that while most of the prevalent microbial species are found in all ruminants, their quantity differs depending on the species (Chaudhary et al., 2012).

In vitro enrichments are also informative to identify functional specificities of ruminal microbes. Lazuka et al indeed analysed diversity and function changes of the cow rumen microbiota, enriched in a bioreactor, with wheat straw as sole carbon source. The authors observed that the microbial diversity of the samples evolved towards a structure able to better degrade wheat straw and produce volatile fatty acids. After the enrichment, the dominant phyla were Bacteroidetes, Firmicutes and Proteobacteria, in this order, which corresponded to 85% of the total community. During the anaerobic enrichment, for the Bacteroidetes, the *Bacteroidia* class decreased in favor of an unclassified class. For the Firmicutes, the *Clostridia* class amount decreased by more than half after the first cycle of enrichment, though they were still well represented. This analysis highlighted the interest of non-sterile microbial enrichment to develop high performance biomass degrading consortia (Lazuka et al., 2015).

4. Enzymes from the rumen

Thanks to the huge diversity of ruminal microbial species, many different enzymes are produced in this ecosystem, allowing them to thrive by metabolising the different dietary components.

The rumen is considered as a specialised fermentation place, where the digestion of plant components is essential for the animals. This specificity makes it an ecosystem of choice to mine for novel enzymes, especially fibrolytic ones that can be used for the production of biofuels, and, more generally, for valorisation of lignocellulosic wastes (Goel et al., 2015). This topic will be further developed in part 5. Extensive functional analysis of the bovine ruminal microbiota would not be possible at such a high-throughput without genomics, transcriptomics and sequence-based metagenomics. All together, these approaches allowed, for the past years, a step increase in the identification of the most important classes of ruminal enzymes.

4.1. Carbohydrate active enzymes

While many kinds of enzymes are present in the rumen, the most studied are the CAZymes (Brulc et al., 2009; Hess et al., 2011; Wang et al., 2013) because of their role in the plant biomass degradation, of interest for white biotechnologies, but also of course for maintaining host health and nutritional efficiency.

CAZymes are of high importance in all living organisms, in particular catabolic ones, because of their ability to transform complex glycans into simple carbohydrates that are required to support cell growth. The sequence-based CAZy classification was created in 1991 to address some of the problems present in the more general Enzyme Classification (EC) system, namely the incompleteness of numbering in EC codes, the non-adapted classification for enzymes having promiscuous activities, or multiple domains with various activities or for enzymes that contain elements with no catalytic function.

CAZymes contain one or several catalytic domain, and most of the time one or several elements along the polypeptide chain, involved in i) addressing the protein to the correct cell compartment or outside the cell (signal peptides); ii) specific binding to carbohydrates (carbohydrate binding modules); iii) protein-protein interaction (dockerins). Catalytic domains and CBMs are defined as “modules” and are the base of the CAZy classification (Lombard et al., 2014). It allows the description of enzymes, even ones with multiple modules, which is highly relevant to interpret biochemical data, and, of course, genomic and metagenomic data. Each CAZy module has been classified in “classes” based on its catalytic or non-catalytic function:

- Glycoside hydrolases (GHs) were the first described in this classification. They contain catalytic modules able to i) hydrolyse glycosidic bonds, ii) perform transglycosylation reactions, or iii) break down glycosidic linkages by phosphorolysis, this particular reaction being performed by glycoside-phosphorylases, which can be classified in GH families because of their structural and mechanistic similarities with real glycoside hydrolases. These three reactions happen between two carbohydrate monomers or between a carbohydrate motif and another molecule (protein or lipid) (Henrissat and Davies, 1997).
- Glycosyl Transferases (GTs) are involved in the synthesis of glycosidic linkages. This needs a phosphor-activated sugar donor and an acceptor that can be a carbohydrate or another molecule containing a hydroxyl group (Coutinho et al., 2003). Some GT families also contain glycoside-phosphorylases for the same reasons as for GH families.
- Polysaccharide lyases (PLs) cleave glycosidic bonds in polysaccharides containing uronic acid. Instead of hydrolysis, the mechanism used is β -elimination, which does not need a water molecule, creating an unsaturated hexenuronic acid residue, along with a new reducing end (Lombard et al., 2010).
- Carbohydrate esterases (CEs) can O- or N-deacetylate esterified sugars. There are two classes of CEs, depending on the role of the sugar moiety in the ester formation: acid (uronic acid) or alcohol (neutral sugars esterified with short chain fatty acids) (Jayani et al., 2005).
- Auxiliary activities (AAs), which perform redox reactions. It is the newest class in the CAZy classification, containing two main classes of enzymes: the lignolytic enzymes, which, while they are not active on polysaccharides, help other CAZymes to access

the carbohydrates hidden in the plant cell wall by oxidising the surrounding lignin; the lytic polysaccharide mono-oxygenases (LPMO), degrading polysaccharides using an oxidative mechanism (Levasseur et al., 2013). AA modules can also be associated on the same polypeptidic chain with other CAZy modules, in order to degrade recalcitrant substrates, like chitin (Vaaje-Kolstad et al., 2010). Their presence and abundance in the ruminal microbiome has still not been studied, to our knowledge.

- Carbohydrate binding modules (CBMs): they are non-catalytic modules, often associated with catalytic modules on a single polypeptidic chain. CBMs facilitate association of catalytic domains with the substrate through specific binding (Boraston et al., 2004).

Depending on their catalytic mechanism, GHs and GTs can be divided in two different groups. During reaction, the asymmetric carbon C1 can undergo a conformational change, leading to a carbohydrate with the same anomery as the substrate, or a reversed one: the enzymes are classified as “retaining” enzymes or “inverting” enzymes, respectively (Coutinho et al., 2003).

As previously mentioned, the CAZy classification is a sequence-based one, describing structural similarities in the fold and catalytic machinery between enzymes with related sequences (Henrissat et al., 1995). As such, the probability of a catalytic mechanism being identical between two enzymes is higher the more their sequences are similar. Another advantage of the CAZy database is that it is manually-curated, meaning there is a much lesser probability of automatic assignation problems that can be found in other CAZy databases like CAT (Park et al., 2010), dbCAN (Yin et al., 2012) and its derivative (Ekstrom et al., 2014). In the CAZy database, a new family is created only when at least one of its members has been functionally characterised. All of this represents a powerful tool to infer the activity of uncharacterised enzymes using their sequences. However, since the identity of two sequences may vary considerably in the same family, other subcategories have been created to strengthen the prediction power of the database, the more precise the subcategories are:

- Superfamily: contains families with a shared common fold, and indicates the basic features of the reactions performed by its members (for example, creating or cutting α or β glycosidic bonds).
- Clan: contains families sharing sufficient common secondary structure elements to fold in the same way. The catalytic machinery can also be found on the same secondary structure elements.
- Family: contains all enzyme sequences sharing the same catalytic mechanism due to a sufficient sequence identity, harbouring conserved catalytic amino acid(s) and functional residues, positioned in the same space location in the catalytic or binding site, which implies a conservation of the fold.

- Subfamily: used for poly-specific families, it is family specific, depending on the available chemical data, and, as such, not present for all CAZy families.

Another advantage of the CAZy database can be seen for the interpretation of genomic and metagenomic data in regard to ecology. Indeed, the analysis of the CAZyme contents (or CAZomes), of one particular organism, brings crucial information about its metabolic abilities, and allows one to explain changes in organism behaviour during different stages of its life (Veneault-Fourrey et al., 2014).

In October 2015, the CAZy database lists 133 GH families (246,906 entries), 97 GT families (200,750 entries), 23 PL families (5,887 entries), 16 CE families (27,306 entries), 71 CBM families (56,890 families) and 13 AA families (10,106 entries), a number that continuously grows over time, with an increasing speed (André et al., 2014).

Bacteria, to more efficiently sequester, depolymerise and metabolise polysaccharides, have evolved two main specific strategies. The first one involves carbohydrate transporters and synergistic CAZymes, encoded by multigenic clusters specifically regulated, the polysaccharide utilisation loci (PUL) (Larsbrink et al., 2014) (Figure 9). Since they have been predominantly studied in Bacteroidetes species (Bjursell et al., 2006), the PUL denomination is specific to this phylum (Terrapon and Henrissat, 2014). It does not mean, however, that PUL-like systems have never been observed elsewhere: metagenomic studies, for example targeting gut bacteria, generated enough data to identify this kind of multigenic clusters (Cecchini et al., 2013; Tasse et al., 2010; Wang et al., 2013) in other genera, even if there is still no proof of simultaneous expression of the gene identified in these loci. The prototype of all PUL-like systems is the starch utilisation system or *Sus*, which is used to sense, bind and depolymerise complex carbohydrates.

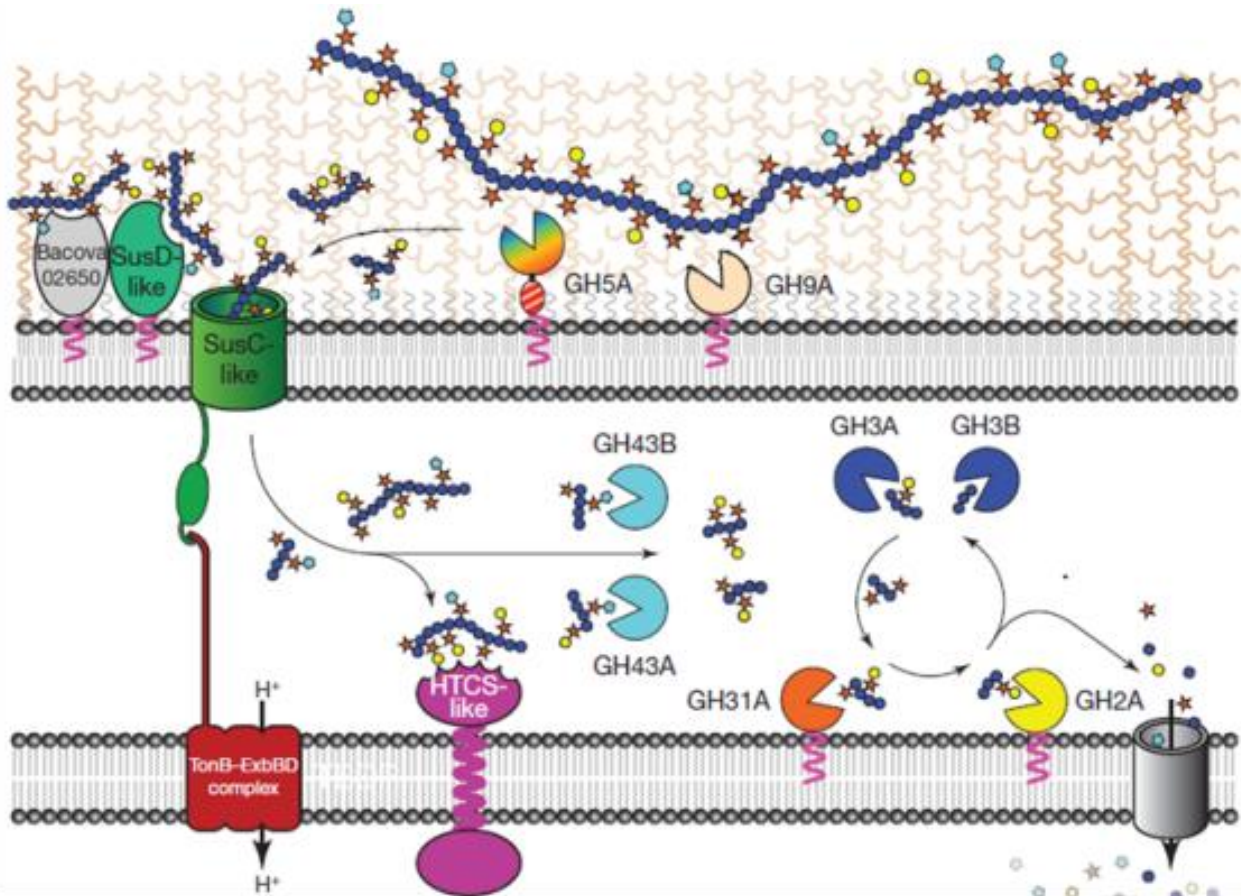


Figure 9: Bacteroidetes polysaccharide utilisation system: example of the xyloglucan degradation pathway of *Bacteroidetes ovatus* ATCC 8483 (adapted from Larsbrink et al., 2014). The xyloglucan oligosaccharide is made of a main chain composed of β -1,4-linked glucose units (blue spheres), extensively ramified with α -1,6-linked xylosyl units (red stars), β -1,2-galactosyl (yellow spheres) and α -1,2-L-fucosyl units (pale blue pentagons). Binding of the raw substrate and initial degradation steps are insured by the membrane anchored Bacova_02650, GH5A, GH9A, SusC and SusD-like proteins, allowing translocation into the periplasmic space of shorter xylo-glucooligosaccharides. The final deconstruction is ensured by the synergetic action of additional GH43A/B, GH31, GH2A and GH3A/B. Monosaccharides are then translocated to the cytosol through a final transporter. Regulation of the expression on the enzymes involved in this complex multi-enzymatic system is ensured by a hybrid two component system (HTCS)-like regulatory transmembrane protein.

In Firmicutes, the degradation of plant polysaccharides and notably cellulose (Bayer et al., 2004; White et al., 2014) and starch (Ze et al., 2015) is accomplished by a multi-functional complex consisting of the self-assembly of enzymatic and structural components named respectively the cellulosome (Figure 10), or amylosome.

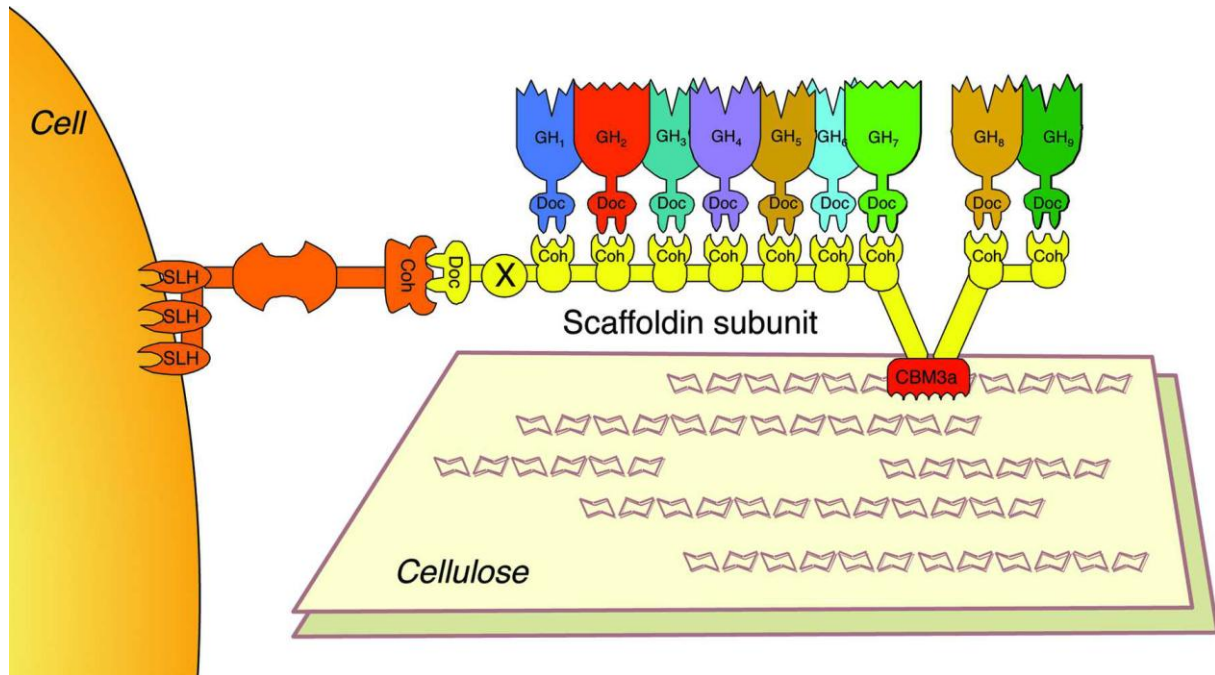


Figure 10: Schematic representation of the *C. thermocellum* cellulosome. The cellulosome is an extracellular multienzyme assembly attached to the cell wall via a surface layer homology (SLH)-like module of an anchoring scaffoldin. An assortment of the 70 possible glycoside hydrolases (GHs) can be incorporated into the cellulosome by binding to only nine available type I cohesin modules through the cohesin–dockerin interaction. The enzyme-borne dockerin (Doc) modules essentially exhibit the same specificity and can bind to any of the nine homologous cohesins (Coh). The cellulose-binding module (CBM3a) of the primary scaffoldin binds the cellulosome complex and the attached cell to the cellulosic substrate (from Yaniv et al., 2014)

Signatures of cellulosome-like assemblies were identified in several studies targeting highly prevalent bacterial species in the rumen, like *Ruminococcus flavefaciens* FD-1 (Berg Miller et al., 2009; Rincon et al., 2010), and the overall ruminal microbiome (Brulc et al., 2009). The high number of dockerin sequences in the genome of *Ruminococcus flavefaciens* FD-1 (225) highlighted the possibility of the cellulosome-like structures to assume different configurations depending on the targeted substrates (Berg Miller et al., 2009; Brulc et al., 2009). Furthermore, the cellulosome components of other ruminococcal strains were deeply analyzed (Purushe et al., 2010; Dassa et al., 2014).

As early as 2009, comparative metagenomics was used to analyse the ruminal metagenomic CAZome, and to compare its diversity in the fibre adherent and liquid microbiomes. In the pioneering study, the authors were able to identify 35 GH and 3 CBM families (Brulc et al., 2009). Hess et al further found 56 GH and 20 CBM families, using massive metagenome sequencing (Hess et al., 2011). Wang et al further studied the microbiome of biofilms adherent to fibre in the cow rumen, finding CAZy belonging to 42 different GH families, along with CBM, carbohydrate esterase and polysaccharide lyase

families (Wang et al., 2013). A comparison of these studies, performed by Wang et al, showed, for example, that between the studies of Hess and Wang, only 129 ORFs were identical, including 12 GH and 14 other CAZy encoding genes (Wang et al., 2013). This is attributed to the differences in host species, diet, and/or sampling and analysis methods. In any case, all these studies show the amazing CAZyme diversity in this ecosystem.

A comparison of the major GH families identified in some herbivores, ruminant or not, can be found in Table 4. It is clear, from these studies, that herbivore CAZymes are geared towards the degradation of oligosaccharides, since more than 50% of the GH encoding genes belong to families mostly containing glycosidases. The GH3 family is indeed the most represented. Endo-hemicellulases are by far second, not even reaching half the number of oligosaccharide degrading enzymes (15 to 24%). Cellulases and debranching enzymes follow (6 to 15%). Indeed, a large number of oligosaccharides are freed as by-products after initial breakdown of complex polymers. Also, the high proportion of oligosaccharide degrading enzymes results in rapid formation of simple fermentable carbohydrates. This is a necessary trait for a large amount of microbes which have to feed on carbohydrates, and is also an advantage for the host, as fiber fermentation leads to the production of volatile fatty acids which supply energy to the animal. The key ruminal enzymes will be described in the following paragraphs.

Table 4: Comparison of the abundance of major GH families in the gut metagenomes of herbivores, ruminant or not (adapted from Wang et al. 2013). The number of putative enzymes is followed by the percentages, in brackets, of each group relative to the total number of GHs identified in each metagenomic dataset.

GH family	Major activity	Wallaby¹	Reindeer²	Yak rumen³	Angus cows – pooled liquid⁴	Angus cows – fibre adherent⁴	Guernsey cows⁵	Jersey cows⁶
Cellulases								
GH5	Cellulases	10	287	1,302	7	7	1,451	24
GH6	Endoglucanase	0	0	0	NR	NR	0	1
GH7	Endoglucanase	0	0	0	NR	NR	1	0
GH9	Endoglucanase	0	109	767	7	6	795	8
GH44	Endoglucanase	0	5	0	0	0	99	0
GH45	Endoglucanase	0	0	13	NR	NR	115	0
GH48	Cellobiohydrolases	0	5	32	0	1	3	0
Total		10 (4%)	406 (8%)	2,114			2,464 (11%)	33 (13%)
Endo-hemicellulases								
GH8	Endoxylanase	1	35	174	8	2	329	0
GH10	Endo-1,4- β -xylanases	11	190	2,664	10	5	1,025	35
GH11	Xylanases	0	8	244	2	1	165	0
GH12	Xyloglucanases	0	0	0	NR	NR	0	1
GH26	B-mannanases & xylanases	5	153	537	2	5	369	1
GH28	Galacturonases	2	120	244	9	3	472	0
GH53	Endo-1,4- β -galactanases	9	125	1,066	15	17	483	18
Total		28 (11%)	631 (12%)	4,929			2,843 (12%)	55 (22%)
Xyloglucanases								
GH16	Xyloglucanases	4	116	563	0	1	483	0
GH74	Xyloglucanases	1	44	0	0	0	385	0
Total		5 (2%)	160 (3%)	563			868 (4%)	0

Debranching enzymes								
GH51	α -L-arabinofuranosidases	12	488	0	73	61	1,249	1
GH54	α -L-arabinofuranosidases	0	23	111	0	1	76	0
GH62	α -L-arabinofuranosidases	0	0	0	NR	NR	1	0
GH67	α -glucuronidases	5	74	1,090	NR	NR	120	0
GH78	α -L-rhamnosidases	25	313	426	41	31	1,260	13
Total		42 (15%)	898 (17%)	1,627			2,706 (12%)	14 (6%)
Oligosaccharide degrading enzymes								
GH1	β -glucosidases	61	122	331	7	10	253	10
GH2	β -glucosidases	24	716	942	218	176	1,436	16
GH3	β -glucosidases	72	844	5,448	207	166	2,844	48
GH29	α -L-fucosidases	2	268	899	31	26	939	3
GH35	β -galactosidases	3	39	468	21	9	158	2
GH38	α -mannosidases	3	116	90	22	15	272	1
GH39	β -xylosidases	1	76	159	2	2	315	18
GH42	β -galactosidases	8	95	207	10	12	374	0
GH43	Arabino/xylosidases	10	787	2,313	68	59	2,932	28
GH52	β -xylosidases	0	2	0	0	0	1	0
Total		184 (68%)	3,065 (60%)	10,857			9,524 (42%)	126 (51%)
Starch degrading enzymes								
GH13	α -Amylase	3	NR	2949	47	37	3442	15
GH77	Amylomaltase	0	NR	1739	0	1	943	2
Total		3 (1%)		4,688	47	38	4,385 (19%)	17 (7%)

Other domains associated with GHs								
Cohesion		0	52	51	0		80	0
Dockerin		41	92	516	8		188	2
SusC		36	1,122	NR	9		3,110	4
SusD		42	685	NR	11		1,889	3
Metagenomic size		0.054 Gb	0.030 Gb	9.4 Gb	0.024 Gb	0.027 Gb	268 Gb	0.28 Gb

¹ (Pope et al., 2010) sequencing of activity-screened fosmids using 454 GS FLX

² (Pope et al., 2012) shotgun metagenomic sequencing

³ (Dai et al., 2012) sequencing of activity screened BAC clones using 454 GS FLX Titanium

⁴ (Brulc et al., 2009) the number of fibre adherent samples was the average of three fibre adherent samples (shotgun metagenomic sequencing using 454 GS20)

⁵ (Hess et al., 2011) shotgun metagenomic sequencing using Illumina GAII and HiSeq2000

⁶ (Wang et al., 2013) pyrosequencing using Roche GS FLX Titanium

4.1.1. Cellulases

These enzymes are able to hydrolyse 1,4- β -D-glucosidic bonds in cellulose, to release cello-oligosaccharides. There are three major subcategories, depending on the structure and function of cellulases (Figure 11):

- Endo- β -1,4-glucanases : they randomly cleave internal bonds, and can be processive or not. GH9 is the most represented family in the bovine rumen.
- Exo- β -1,4-glucanases: they cleave the reducing or non-reducing end of cellulose, including crystalline cellulose. Cellobiohydrolases are processive exo- β -1,4-glucanases. GH48 is the most represented family.
- Cellobiases, also called β -D-glucosidases: they hydrolyse cellobiose, the major product of the endo- and exo-glucanase mixture, into glucose (Horn et al., 2012; Liu et al., 2011). GH5 and GH3 are the most represented families.

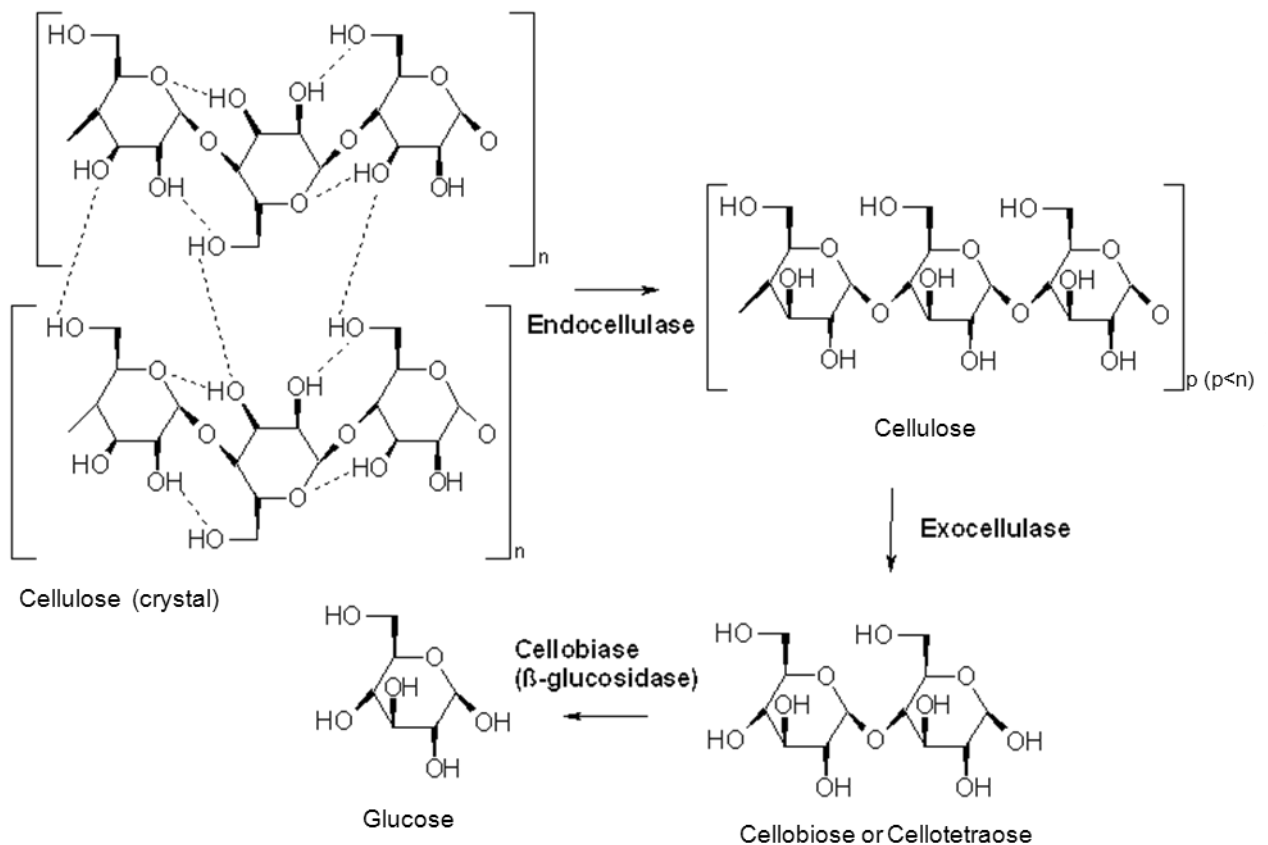


Figure 11: The three types of reactions catalyzed by cellulases on cellulose (Pradhan and Rathi, 2008)

Depending on the microbe producing them, cellulases can be free enzymes or included into cellulosomes, as previously explained. Cellulolytic activity has been found in rumen bacteria and also in rumen protozoa (Dassa et al., 2014).

4.1.2. Hemicellulases

This family includes many kinds of enzymes (Figure 12) acting on the hemicellulosic components surrounding cellulose, rendering it more accessible to cellulolytic enzymes:

- Mannanases: β -mannanases degrade mannan, an important part of hemicelluloses, to β -1,4-manno-oligomers, which are then reduced to mannose by β -mannosidases (Patel et al., 2015). GH26 is the most represented family.
- Arabinofuranosidases: bonded to xylan, arabinose is part of hemicelluloses. They are linked together on position α -1,2, 1,3 or 1,5, or can be linked to arabinoxylan on its C2 or C3. Arabinofuranosidases are able to hydrolyse the non-reducing terminal arabinofuranosyl residues (Ferrer et al., 2012). GH51 is the most represented family.
- Xylanases: they are able to hydrolyse xylan, and belong to two categories:
 - o Endoxylanases: they randomly cleave internal bonds. GH8 is the most represented family.
 - o β -xylosidases: they hydrolyse 1,4- β -D-xylans, to remove successive D-xylosyl residues from the non-reducing end of the xylan chain (Ferrer et al., 2012). GH52 is the most represented family.
- Ferulic acid esterases: they are a subclass of carboxylic ester hydrolases. They can produce ferulic acid by hydrolysing the link between hydroxycinnamates and carbohydrates (Cheng et al., 2012a). CE1 is the most represented family.
- Acetyl xylan esterases: they are able to deacetylate xylans and xylo-oligosaccharides (Cybinski et al., 1999). CE6 is the most represented family.
- α -glucuronidases: they cleave glucuronic acid side groups that protect β -1,4-linkages of the xylan backbone from degradation by xylan degrading enzymes. GH67 is the most represented family.

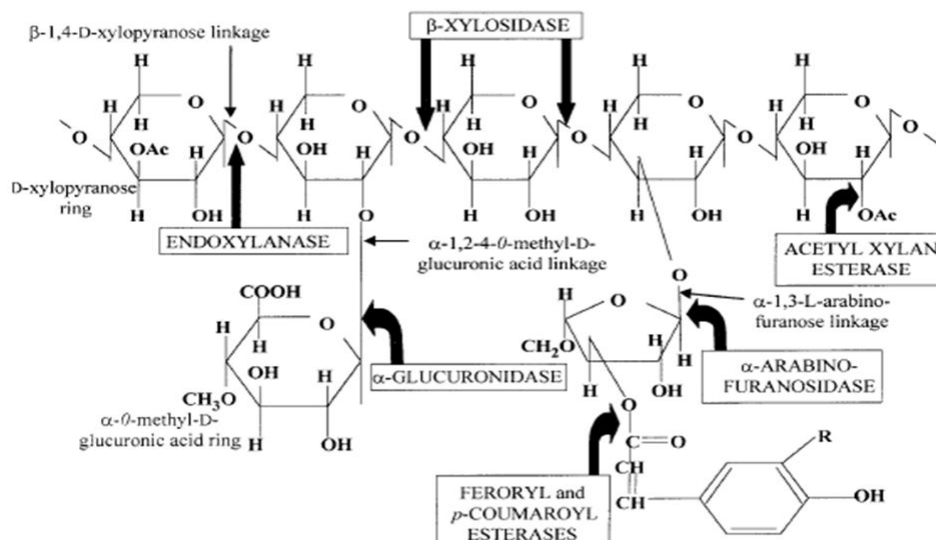


Figure 12: Schematic representation of plant xylan structure, with its different substituent groups, showing sites of attack by microbial xylanolytic enzymes. Ac: acetyl group; R -H: *p*-coumaric acid; R -OCH₃: ferulic acid. (Beg et al., 2001)

4.1.3. Pectinases

Different pectinolytic activities can be found in the ruminal microbiome (Figure 13):

- Pectin lyases: they cleave 1,4- α -D-galacturonan methyl esters into oligosaccharides with 4-deoxy-6-O-methyl- α -D-galactosyl-4-enuronosyl groups at their non-reducing ends. PL1 is the most represented family.
- Polygalacturonases: they hydrolyse α -1,4 glycosidic links between galacturonic acid residues. GH28 is the most represented family.
- Pectin methylesterases: they de-esterify pectin by ester bond hydrolysis of methylated α -1,4-linked D-galacturonosyl units, producing a negatively charged polymer and methanol (Sharma et al., 2013). CE8 is the most represented family.

Pectinases in the rumen are produced by protozoa, fungi and bacteria (Murad and Azzaz, 2011; Yuan et al., 2012). For example, one of the major pectinolytic bacterial species from the rumen is *Lachnospira multiparus*, which expresses a pectin lyase and a pectin methylesterase (Silly, 1985).

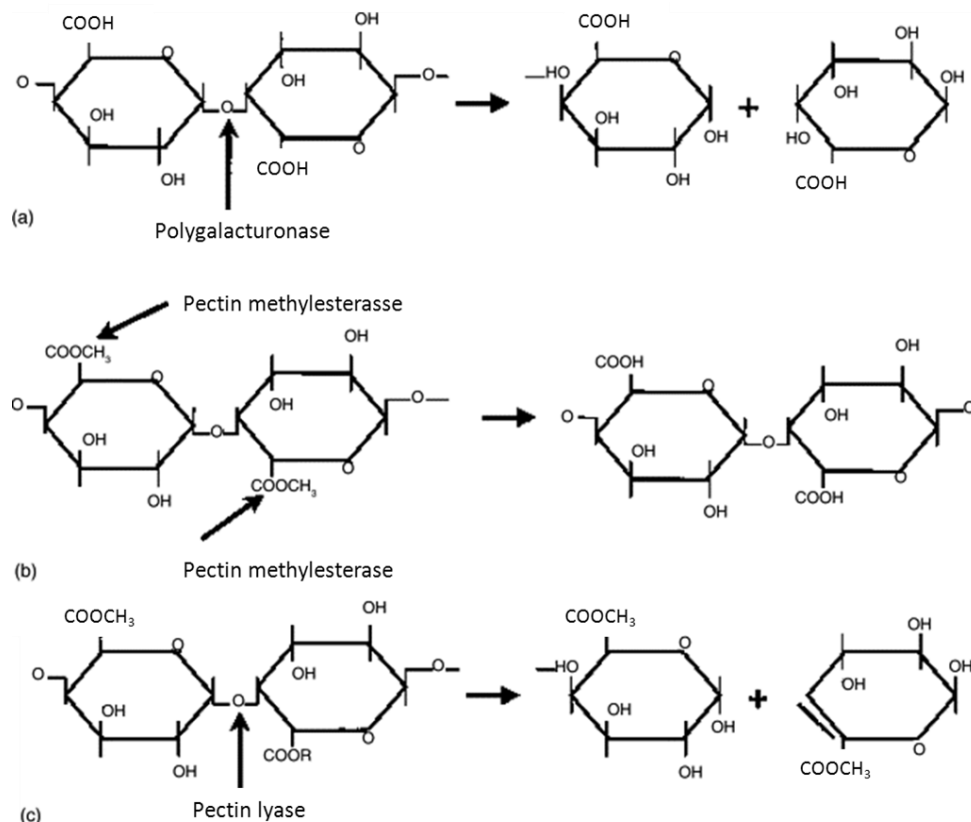


Figure 13: Pectinolytic activities. (a) polygaracturonase activity; (b) pectin methylesterase activity; (c) pectin lyase activity. Adapted from Jayani et al. 2005.

4.1.4. Auxiliary activities

CAZymes with auxiliary activities are also involved in the degradation of plant cell wall. AA members are redox enzymes acting on poly- and oligo-saccharides (cellulose and cellobiose, starch and glucooligosaccharides, chitin and chitooligosaccharides), but also on lignin, like laccases and lignin peroxidases. While these families have the potential to help processing plant cell wall, there is still no data regarding their presence and abundance in the bovine ruminal microbiome. This is due to the novelty of AA classification that has still to be exploited for mining metagenomes for such enzymes.

4.1.5. Amylases

Thanks to α -amylases (GH13, 57), starch degrading organisms degrade starch to glucose, which is then fermented in short chain fatty acids for the host (Zhao et al., 2010). Starch digestion is important for the health of bovine: if starch digestion is incomplete, the nutritional efficiency may be low, while a starch digestion that is too fast may lead to subacute ruminal acidosis because of the overproduction of volatile fatty acids. In the rumen, both protozoa and bacteria are known to degrade starch. Protozoa, like *Entodinium* spp., *Eremoplastron bovis*..., can ingest starch and soluble carbohydrates to digest them internally, or can ingest bacteria, including amylolytic ones, which would influence starch degradation in the rumen. Some bacteria, like *Streptococcus bovis*, *Butyrivibrio* spp. ..., are known to ferment starch, and some can produce extracellular or cell surface associated amylase activity. Unfortunately, the composition of active starch hydrolyzing bacteria and their enzymes in the rumen is still mostly unknown (Xia et al., 2015).

4.2. Oxidoreductases (not referenced in the CAZy database)

Lignin is not easily fermented by rumen bacteria, but is partly degraded by rumen anaerobic fungi (Ruiz-Dueñas and Martínez, 2009). One of the reasons could be that lignin degradation is an oxidative process, which has been described mostly in aerobic ecosystems where di-oxygen can act as electron acceptor. Regarding the rumen microbiota, exogenous fungal oxidoreductases can be used to supplement animal diet, or to pretreat wheat straw, in order to increase nutritional efficiency (Shrivastava et al., 2011). But not many oxidoreductases have been isolated from the bovine gastrointestinal tract itself, despite its high ability to degrade plant cell wall. In their study, Virkel et al showed that the bovine intestinal mucosa produces such enzymes (cytochrome P450, flavine containing monooxygenase) which can be involved in the oxidative transformation of xenobiotics, (Virkel et al., 2009), without implying an activity on lignin. Regarding bacterial enzymes, the only known, to our knowledge, oxidoreductases discovered from the bovine rumen microbiota by activity-based metagenomics is a bacterial laccase, a copper-containing oxidase enzyme (Beloqui et al., 2006), which was not tested on lignin. One has to mention that by using sequence-based metagenomics on anaerobic ecosystems, including cow rumen

microbiota, only few bacterial laccases were discovered, with a very low hit yield. None of these sequences covered any of the copper-binding regions within laccase sequence (Ausec et al., 2011). Further research is thus needed to deepen our understanding of the different lignin degradation mechanisms that occur in the bovine gastrointestinal tract.

4.3. Phytases

These enzymes are able to hydrolyse phytic acid in order to release inorganic phosphorus to be later used by the host and the ruminal microbes (Figure 14) (Huang et al., 2011). In their study, Huang et al were able to amplify dozens of sequences encoding cysteine phytases, highlighting the importance of this phytase family in the ruminal ecosystem.

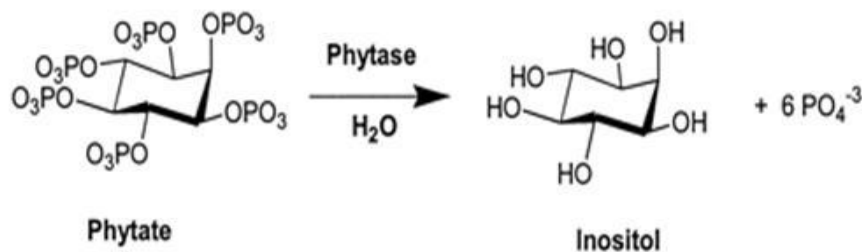


Figure 14: Mechanism of phytate hydrolysis by phytases (Mittal et al., 2011)

4.4. Tannases

Tannins are natural polyphenolic compounds present in vascular plants. They are the second most abundant group of phenols in nature. Tannins are known as antinutritional components of ruminant diet, since they can bind to polysaccharides, epithelial proteins, and digestive enzymes inside the rumen, which leads to inhibition of rumen microbial growth and activity, and finally, to nutrient uptake. Tannins are thus able to inhibit the growth and productivity of grazing animals. But tannins are also believed to act with a defensive role of plant proteins against ruminal digestion. This leads to an increased influx of essential amino acids into the small intestine, which results in an increase in microbial protein synthesis, in the use of endogenous nitrogen in the rumen and in the secretion of salivary glycoproteins. All of this leads to a decrease of parasite infestation and, as such, is important to prevent infection in livestock.

Parts of the tannin acyl hydrolases family, tannases (EC 3.1.1.20) hydrolyse the central ester bond between the two aromatic rings of substrates that contain at least two phenolic hydroxyl groups in the acid components (Figure 15). They have been found in the ruminal species *Selenomonas* spp. and *Streptococcus* spp. (Goel et al., 2005; Singh et al., 2001).

The multiple aromatic cycles that can be found in tannin structures, especially in their condensed form, and their ability to bind different minerals and macromolecules, such as

proteins, cellulose, starch, among others, reminds one of the complex lignin structure and its interactions with the different parts of the plant cell wall (Chávez-González et al., 2012). Furthermore, one of the tannin degradation products, digallic acid (Figure 15), contains both an ester bond and aromatic cycles, reminding one of some of our carbamate targets, later described in this study. Tannins could thus be considered as structural analogs of the compounds targeted in this thesis project.

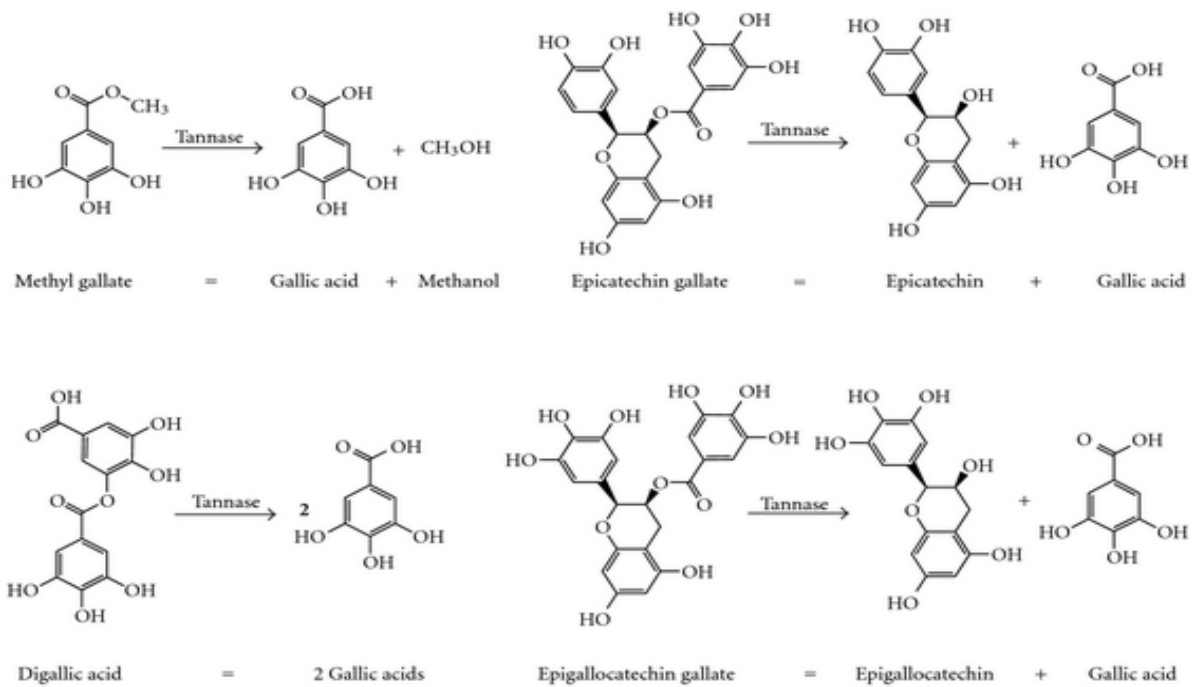


Figure 15: Hydrolytic reaction by tannase (Rodríguez-Durán et al., 2011)

4.5. *p*-coumaric acid esterases

In order to efficiently degrade lignocellulosic biomass, *p*-coumaric acid esterase is very important. Its role is to help breaking ester links connecting lignin to hemicelluloses, simultaneously releasing *p*-coumaric acid. Until now, this enzyme has been proven to be produced only by anaerobic fungi, explaining their importance in the rumen microbial ecosystem (Borneman et al., 1991).

4.6. Lipases

After feeding, dietary esterified lipids are hydrolysed by microbial lipases to free fatty acids and glycerol, along with a small amount of mono and diglycerides (Jenkins, 1993). Ruminant microbial lipases are extra cellular enzymes assembled in small beads, produced by a low number of microorganisms, and, as all enzymes, are highly specific to their physiologic substrates (Fay et al., 1990; Henderson, 1971). For example, while the *Butyrivibrio fibrisolvens* lipase hydrolyses phospholipids, that of *Anaerovibrio lipolytica* hydrolyses only di

and triglycerides. While lipase activity is predominant in ruminal bacteria, it is also possible at a lesser level by ciliate protozoa, but not by ruminal fungi. Free fatty acids can also be produced by some bacterial rumen galactosidases and phospholipases, after hydrolysis of plant galactolipids and phospholipids (Jenkins, 1993). Lipolysis, by producing free fatty acids, enables their biohydrogenation, explained in more details in the next paragraph.

4.7. Enzymes involved in polyunsaturated fatty acids biohydrogenation

Because of the toxic effect of the unsaturated fatty acids coming from the dietary compounds, ruminal microbes developed a defence mechanism based on the production of many different enzymes to hydrolyse and hydrogenate these compounds.

In the rumen, biohydrogenation is predominantly assured by bacteria (*Butyrivibrio fibrisolvens*, *Clostridium proteoclasticum*...), using a two-step process, involving isomerases and reductases (Figure 16). For example, the first step of biohydrogenation of linoleic acid is isomerisation, where the *cis*-12 bond is converted to a *trans*-11 bond. The second step is hydrogenation of *cis*-9, *trans*-11-conjugated linoleic acid to *trans*-11-18:1.

Apart from bacteria, anaerobic fungi have been showed to have some very slow (compared to bacteria) biohydrogenation activity. Protozoa, however, do not seem to be involved in biohydrogenation (Jenkins et al., 2008).

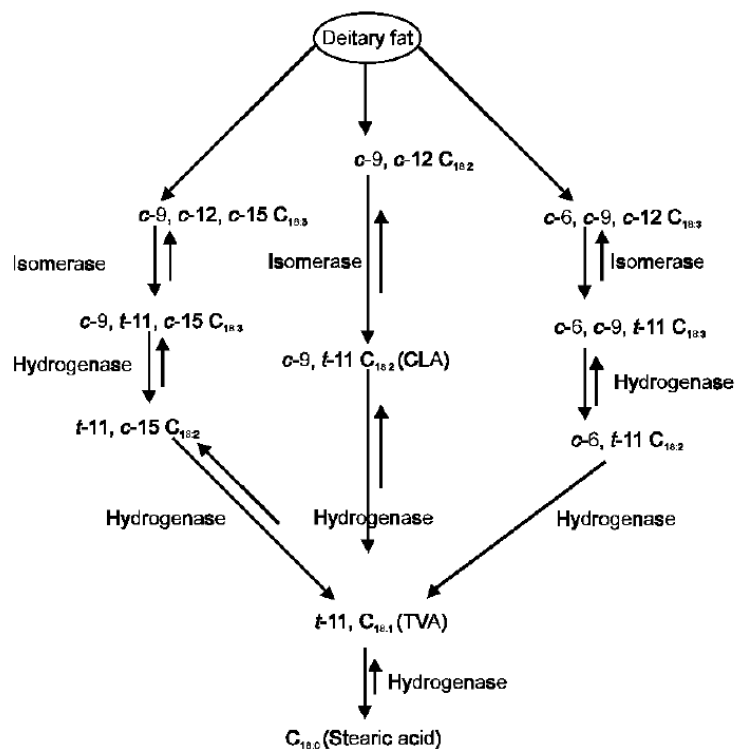


Figure 16: Predominant pathways of biohydrogenation of dietary linoleic and linolenic acids in the rumen (Khanal and Dhiman, 2004).

5. Application of rumen enzymes

Many of the different microbial enzymes present in the rumen are of interest for biotechnologies. Ruminal enzymes can be implemented in industrial processes either *in vivo*, by using directly the microbial strains that produce them, or *in vitro*, in enzymatic reactors or to supplement feeds.

5.1. Feed additives:

Because of the complexity of plant-based dietary compounds, and the lack of proper degradation pathway by non-ruminant hosts, which do not possess any complex stomach microbiota to assure this function, , addition of exogenous plant cell wall degrading enzymes (cellulases, pectinases, hemicellulases) to their feed improves digestion of the dietary raw materials (Cheng et al., 1999). This enables the decrease of anti-nutritional factors, high fibre content being the main one, which then help nutrient absorption and avoid intestinal disturbance by pathogenic microorganisms. These exogenous enzymes have to be very stable since they are, most of the time, added as granules. The coating is thus very important to protect the enzymes from heat treatment and from their deactivation by other food additives, like choline chloride (Bansal and Goel, 2015).

Another key element for animal health is the phosphorous content of their diet, which is important for the formation and upkeep of the skeletal structure. Addition of exogenous phytases to feeds, allows the release of inorganic phosphorous from the diet, and, as such, lowers the need of added orthophosphate. It also makes for less phosphorous manure that is detrimental to the environment. The presence of phytases in feeds also lowers the chelation properties of phytic acid which releases the different metals needed as nutrients by the animals (Cheng et al., 1999). For the same reasons, such plant cell wall degrading enzymes and phytases can also be used to supplement ruminant feeds (Murad and Azzaz, 2011).

5.2. Biorefineries:

Because of the depletion of non-renewable fossil fuel and the ever growing population on Earth, the need of alternatives to reach the necessary energy for all sectors implies the development of new sustainable processes of plant biomass bioconversion into biofuels, chemicals and other by-products. However, while their growth is exponential, the biorefinery industries are not yet economically competitive, since the raw materials used are costly, and the amounts of enzymes required to break down plant biomass too high. The use of highly efficient and stable enzymes and/or microbial strains, as well as agro-industrial by-products, like wheat straw (which is a lignocellulosic substrate that is not in competition with food industries), as substrates, already allows one to alleviate these barriers. With its exceptional functional diversity and its efficiency to break down plant materials, the rumen

microbiome constitutes a goldmine to search for robust enzymes that are naturally well adapted to industrial processes, as they act on crude materials in low water content conditions. This is of course the case for plant cell wall degrading enzymes, that are of particular interest for second generation biofuel production, but also for lipases for biodiesel synthesis (Bansal and Goel, 2015). However, their industrial uses are still limited because of the production of reaction by-products (glycerol for lipases, lignin derived aromatic compounds for cellulases and hemicellulases) that retro-inhibit the enzymes. As such, these processes are still being developed or, in the best cases, optimised to make the production of biofuels more efficient and cost-effective (Del Pozo et al., 2012; Moon et al., 2011).

5.3. Bio-based organic acids production:

Many organic acids and chemicals are used as synthons for the production of food additives, fragrances... Demands for these kinds of molecules, especially for biosourced ones, continuously increases. Thankfully, some processes have been developed to produce few of them from renewable resources, like for example, the production of succinic acid from different plant materials, depending on the geographical location (grass, maize...), through their fermentation by ruminal bacteria. As of now, to our knowledge, no individual enzymes are used, but strains like *Actinobacillus succinogenes* (Guettler et al., 1999), *Mannheimia succiniciproducens* (Kim et al., 2004), *Anaerobiospirillum succiniciproducens* (Nghiem et al., 1997) can be used to produce succinic acid (Sauer et al., 2012).

6. Meta-omic analyses of the rumen microbiota:

6.1. Metagenomics:

In order to explore the functional potential of complex ecosystems, high-throughput culture-independent genomic analyses of their microbial communities have been developed in the last decades. As mentioned in more details in the previous chapter (Ufarté et al., 2015), there are two different strategies for mining microbial ecosystems for novel enzymes and metabolic pathways: functional sequence-based and activity-based metagenomics, which have both been extensively used to study the bovine rumen microbiota (Morgavi et al., 2013).

6.1.1. Sequence-based metagenomics:

Sequence-based metagenomics is the analysis of the whole microbial genomes of an ecosystem, using large scale random sequencing of functional genes (in particular those encoding enzymes), or even using DNA/DNA or DNA/cDNA hybridisation (Ufarté et al., 2015).

There have been several large-scale metagenome sequencing projects targeting the rumen microbiome over the past few years. These studies were initiated in 2009, by Brulc et al., who revealed the diversity of CAZyme encoding genes (Brulc et al. 2009). However, not many of them also contain an experimental validation of enzymatic functions. Let's cite the study of Hess et al., which enabled identification of 27,755 CAZy encoding genes from 268 Gb of rumen metagenome. Fifty one CAZymes were isolated, belonging to known families specifically involved in lignocellulose degradation (Hess et al., 2011) after identification of the most original sequences within the known protein families, followed by PCR amplification and cloning of the gene targets. Other approaches, along with their advantages and disadvantages are detailed in a previous chapter (Gene-Targeted-metagenomics, microarrays, solution hybrid selection...) (Ufarté et al., 2015). In Table 5 are listed all the examples of sequence-based functional metagenomics studies which allowed PCR amplification of enzyme targets, which were further expressed in *Escherichia coli*, or in the case of Hou et al., in *Saccharomyces cerevisiae*, a strain of interest for bioethanol production. In this latter case, the recombinant strain contains a new xylose isomerase which renders it able to ferment xylose, and to benefit from the complete conversion of pentoses and hexoses in lignocellulosic feedstocks (Hou et al., 2015).

6.1.2. Activity-based metagenomics:

Activity-based metagenomics is the analysis of the whole microbial genome of an ecosystem, expressed in heterologous host. Briefly, total microbial DNA is extracted and fragmented, then cloned in a vector (plasmid, cosmid, fosmid, or bacterial artificial chromosome) for expression in a heterologous host (*Escherichia coli*, *Bacillus subtilis*...). These metagenomics libraries are then screened for various enzyme activities. Further details on the method can be found in a previous chapter (Ufarté et al., 2015).

The bovine rumen microbiome has been the target of many activity-based metagenomic studies targeting the discovery of new enzymes of biotechnological interest, especially plant cell wall degrading enzymes (Morgavi et al., 2013). The first study was published by Ferrer et al., a pioneering team in activity-based metagenomics, who found 11 esterases, one cyclodextrinase and 7 β -1,4-glucanases from a dairy cow rumen metagenomic library (Ferrer et al., 2005). Many other studies, found in Table 6, where published, mostly focused on CAZymes (glycoside hydrolases, carbohydrate esterases...). Only a handful of publications deals with the discovery of other kind of enzymes, like lipases (K. Liu et al.; Zhao et al., 2009 2009; Privé et al., 2015), a 3,5,6-trichloro-2-pyridinol degrading enzyme, targeting chlorpyrifos, an organophosphorous insecticide (Math et al., 2010), and a polyphenol oxidase with laccase activity (Beloqui et al., 2006).

6.2. Metatranscriptomics:

This method allows the measurement of the abundance of mRNA transcripts from the microbial ecosystem, along with their dynamic and regulation under different environmental stresses or different ages of the animals. In their study, Dai et al. gave a metatranscriptomic insight in the plant cell wall polysaccharide degradation in the cow rumen. The authors determined if and how actively the CAZy encoding genes are expressed and how the expression of these genes is coordinated in an efficient manner. Actively transcribed genes encoding putative glycoside hydrolases involved in cellulose, hemicellulose, and oligosaccharide degradation were identified, along with the rumen microorganisms responsible for their expression. They discovered that genes encoding GH48 proteins and cellulosome components were transcribed in a relatively high abundance (Dai et al., 2015). Metatranscriptomic studies of rumen samples are few, the only other found to our knowledge being the one of Qi and al, reporting fungal gene expression in the rumen of muskoxen (*Ovibos moschatus*) (Qi et al., 2011).

6.3. Metaproteomics:

This method is used to characterise the proteins encoded by the metagenome of the animals, along with their post-translational modifications and turnover. However, today only a small part of the proteins produced by an ecosystem as complex as the cow rumen can be characterised. Indeed, for complex microbial samples like rumen, the coverage of metaproteomic is low. Because of the high species diversity and cell density of the samples, only predominant proteins are identified while rarer species, which might produce highly efficient and/or unstable enzymes that assure key (but transient) metabolic functions, stay unseen (Deusch et al., 2015; Deusch and Seifert, 2015).

6.4. Metabolomics and meta-metabolomics:

Metabolomics and meta-metabolomics have been used these last years to get information on the relations between specific microbial strains (metabolomics), or the entire microbiota (meta-metabolomics), the proteins they produce, and the metabolites produced in the ecosystem (Saleem et al., 2013). For example, Zhao et al. analysed the ruminal metabolite profile of cows fed different types of roughage. They highlighted that the type of roughage can influence the ruminal microbial metabolome, especially regarding the metabolism of organic acids, amines and amino acids, which can affect health of the animals (Zhao et al., 2014).

7. Conclusion:

Meta-omics are very powerful tools to analyse complex ecosystems like bovine rumen. However, while they have improved much since their creation, the understanding of the functional traits of such microbiota at the ecosystem level, implies continuous amelioration of said tools to expand the coverage of functional diversity (at the gene, transcript, protein and metabolite levels), and to integrate meta-data. While powerful tools by themselves, metagenomics, metatranscriptomics, metaproteomics and meta-metabolomics allow a much better understanding of an ecosystem when being used together on the same sample. This is one of the reasons why, in the present study, a chapter will be devoted to metagenomic analyses of cow rumen libraries, at the functional and taxonomic levels, and to a comparison of different methods of enrichment for the same ecosystem. Indeed, while some studies describe a comparison of ruminal ecosystem with very different diets, the comparison of an *in vivo* and *in vitro* enrichment has never been done to our knowledge. The interest of such a comparison is quite high for industrial purposes: *in vitro* enrichment allows the use of substrates without the health requirement for the animal. Up until now, as seen in part 6.4., plant cell wall polysaccharide degrading enzymes are the most targeted with metagenomics studies, even though other enzymes of interest (oxidoreductases, xenobiotic degrading enzymes...) were discovered. As such, while many advances have been done thanks to meta-omics analyses, much more could be found by enlarging the scope of the studies, bringing hope for many new discoveries of unexpected, rare, and yet no less effective enzymes, of which biochemical properties have to be deeply characterized for the evaluation of their application potential.

Table 5: List of studies mentioning the discovery of enzymes from the cow rumen through sequence-based metagenomics, including an experimental proof of function

Ecosystem	Method	Library size	Number of positive sequences	Activity validated	Reference
Mehsani buffalo rumen	Shotgun sequencing	Unspecified	1 GH26	locust bean gum, beech wood xylan, CMCellulase, pectin	(Patel et al., 2015)
Swamp buffalo rumen	PCR amplification	Unspecified	1 further analysed	Cellulase	(Cheema et al., 2012)
Cow rumen (microbes adherent to plant fiber)	Shotgun sequencing	268 Gbp	51/90 tested, some multiactivities 34 14 11 18 19 17	Enzymes degrading: CMcellulase Xylan Switchgrass Miscanthus Avicel Lichenan	(Hess et al., 2011)
Cow rumen	PCR amplification	Unspecified	1	Xylose isomerase	(Hou et al., 2015)
Cow rumen	PCR amplification	Unspecified	49 clones, 9 distinct sequences	Cysteine phytase	(Huang et al., 2011)

Table 6: List of studies mentioning the discovery of enzymes from the cow rumen through activity-based metagenomics.

Ecosystem	Vector	Library size	Activity observed (primary screening)	Number of positive clones	Reference
Cow rumen	Fosmid	Unspecified size	Esterase	1	(Kambiranda et al., 2009)
Korean cow rumen	Fosmid	Unspecified size	Esterase	1	(Kim, 2012)
Cow rumen	Fosmid	27,500 clones (1.2Gb)	Esterase	3	(Rodríguez et al., 2015)

Holstein cow rumen	Fosmid	Strained ruminal fluid: 7,744 clones, solid-attached bacteria: 8,448 clones, liquid-associated bacteria: 7,680 clones (835 Mbp total)	Lipase/esterase	0 5 4	(Privé et al., 2015)
China Holstein cow rumen	Fosmid	30,000 clones (0.96 Gb)	Feruloyl esterase	157	(Cheng et al., 2012a)
			Xylanase	1 for further analysis	(Cheng et al., 2012b)
New Zealand dairy cow rumen	Fosmid	12,288 clones (490 Mb)	Glycoside hydrolases	8 (one containing a feruloyl esterase after further analysis)	(Ferrer et al., 2012)
Cow rumen	Fosmid	unspecified	3,5,6-trichloro-2-pyridinol degrading enzyme	1	(Math et al., 2010)
Jersey cow run: biofilm of feed particles	Fosmid	14,000 clones (0.56 Gb)	Degradation of: Cellulose Xylan Starch Esculin (a coumarin glucoside)	34 52 1 33	(Wang et al., 2013)
Korean cow rumen fluid	Fosmid	20,000 clones (600-800Mb)	Exo-cellulase	1	(Ko et al., 2013)
Ruminal fluid of cows	Fosmid	1.2 Gb	Cellulase	45	(Loaces et al., 2015)
Holstein cow rumen	Fosmid	16,896 clones (600 Mb)	β -glucosidase	4	(Del Pozo et al., 2012)
Xiangxi yellow cattle rumen	Fosmid	20,160 clones (600 Mb)	β -glucosidase	1	(Li et al., 2014)
Bovine rumen	Fosmid	70,000 clones (2.1 Gb)	CMCellulase	21/10,000 tested; 2 further analysed	(Rashamuse et al., 2013)

Steer rumen	Fosmid	20,000 clones	CMCellulase Sus-like PUL scaffolds	142 10	(Rosewarne et al., 2014)
Swamp buffalo rumen	Fosmid	10,000 clones	Cellulase Xylanase	93 4	(Nguyen et al., 2012)
Chinese yak rumen	Cosmid	4,000 clones (140 Mbp)	Endoglucanase	3	(Chang et al., 2011)
			β -1,4-glucanase exoglucanase	1 with both activities	(Bao et al., 2011)
			β -glucosidase, β -xylosidase	2 with both activities	(Bao et al., 2012)
Chinese yak	Cosmid	5,000 clones (200 Mbp)	β -glucosidase	50, 1 further analysed	(Zhou et al., 2012a)
			β -xylosidase	2 clones analysed	(Zhou et al., 2012b)
Buffalo rumen	Cosmid	15,000 clones (525 Mbp)	CMCellulase β -glucosidase	11 48	(Duan et al., 2009, 2010)
Buffalo rumen	Cosmid	15,600 clones (6.5 Gb)	β -glucosidase	41	(L. Liu et al., 2009)
Chinese holstein cow rumen	BAC	15,360 clones (838 Mbp)	Lipase	18 \rightarrow 2 further analysed	(K. Liu et al., 2009)
Dairy cow rumen (grass/hay diet)	BAC	6,912 clones (900 Mb)	CMCellulase β -glucosidase other glycoside hydrolases	10 9 7	(Gong et al., 2012, 2013; R. J. Gruninger et al., 2014)
Qinghai-Tibetan domesticated yak rumen	BAC	76,000 clones (4.2 Gbp) \rightarrow 9,600 clones	Cellulases, xylanase, carbohydrate esterase	223 (2 GH5 further analysed)	(Dai et al., 2012)
Chinese Holstein dairy cow rumen	BAC	15,360 clones (837 Mb)	Amylase Xylanase	10 18	(Zhao et al., 2010)
Yak rumen	BAC	-	Xylanase	14	(Wang et al., 2011)
Cow rumen	phagemid	14,000 clones (1.1 Gb)	Polyphenol oxidase with laccase activity	1	(Beloqui et al., 2006)
Cow rumen	λ phage	Unspecified	exo- α -1,5-L-arabinanase	1 further analysed	(Wong et al., 2008)
			endo- α -1,5-L-arabinanase	1 further analysed	(Wong et al., 2009)

			endo- β -1,4-glucanase	1 further analysed	(D. D. W. S. Wong et al., 2010)
			Exo-glucanase	1 further analysed	(D. W. S. Wong et al., 2010)
			Feruloyl esterase	1 further analysed	(Wong et al., 2013)
New Zealand dairy cow rumen	λ phage	(1.1 Gb) \rightarrow 14,000 clones (around .077 Gb)	Esterase Cyclodextrinase β -1,4-glucanase	11 1 7	(Ferrer et al., 2007, 2005; López-Cortés et al., 2007)
Chinese dairy cow rumen	λ phage	(0.21 Gbp)	Carboxymethyl cellulase β -glucosidase	2 1	(Wang et al., 2009)
Cow rumen	λ phage	Unspecified	endo-1,4- β -D-xylanase	1	(Palackal et al., 2007)
Cow rumen	λ phage	6,000 clones (90 Mb)	Lichenan degrading enzyme	35	(Shedova et al., 2009)
Cow rumen	λ phage	Unspecified	α -glucuronidase	1 further analysed	(Lee et al., 2012)
Holstein cows rumen protozoans	λ phage	Unspecified	Xylanase Cellulase	60 3	(Findley et al., 2011)
Dairy cow rumen	-	15,360 clones (922 Mb)	Lipase	18	(Zhao et al., 2009)

Artwork

Figure 4: Schematic representation of the bovine digestive tract.....	85
Figure 5: Composition and organisation of primary cell wall	87
Figure 7: Relative distribution of the most abundant phyla in three chambers of cow stomach	91
Figure 6: Phylum level composition depending on the age of the animals.....	91
Figure 8: Relative distribution of phyla of the cow abomasum	92
Figure 9: Bacteroidetes polysaccharide utilisation system: example of the xyloglucan degradation pathway of <i>Bacteroidetes ovatus</i> ATCC 8483.	101
Figure 10: Schematic representation of the <i>C. thermocellum</i> cellulosome.....	102
Figure 11: The three types of reactions catalyzed by cellulases on cellulose	107
Figure 12: Schematic representation of plant xylan structure, with its different substituent groups, showing sites of attack by microbial xylanolytic enzymes.	108
Figure 13: Pectinolytic activities.....	109

Figure 14: Mechanism of phytate hydrolysis by phytases 111
Figure 15: Hydrolytic reaction by tannase 112
Figure 16: Predominant pathways of biohydrogenation of dietary linoleic and linolenic acids
in the rumen..... 113

Table content

Table 4: Comparison of the abundancy of major GH families in the gut metagenomes of
herbivores, ruminant or not. 104
Table 5: List of studies mentioning the discovery of enzymes from the cow rumen through
sequence-based metagenomics, including an experimental proof of function..... 119
Table 6: List of studies mentioning the discovery of enzymes from the cow rumen through
activity-based metagenomics. 119

References

- Agarwal, N., Kamra, D.N., Chaudhary, L.C., 2015. Rumen Microbial ecosystem of domesticated ruminants, in: *Rumen Microbiology: From Evolution to Revolution*. Springer India, pp. 17–30.
- Agger, J.W., Isaksen, T., Varnai, A., Vidal-Melgosa, S., Willats, W.G.T., Ludwig, R., Horn, S.J., Eijsink, V.G.H., Westereng, B., 2014. Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc. Natl. Acad. Sci.* 111, 6287–6292. doi:10.1073/pnas.1323629111
- André, I., Potocki-Véronèse, G., Barbe, S., Moulis, C., Remaud-Siméon, M., 2014. CAZyme discovery and design for sweet dreams. *Curr. Opin. Chem. Biol.* 19, 17–24. doi:10.1016/j.cbpa.2013.11.014
- Ausec, L., Zakrzewski, M., Goesmann, A., Schlüter, A., Mandic-Mulec, I., 2011. Bioinformatic analysis reveals high diversity of bacterial genes for laccase-like enzymes. *PLoS One* 6, e25724. doi:10.1371/journal.pone.0025724
- Bansal, S., Goel, G., 2015. Commercial application of rumen microbial enzymes, in: *Rumen Microbiology: From Evolution to Revolution*. Springer India, pp. 281–291.
- Bao, L., Huang, Q., Chang, L., Sun, Q., Zhou, J., Lu, H., 2012. Cloning and Characterization of Two β -Glucosidase/Xylosidase Enzymes from Yak Rumen Metagenome. *Appl. Biochem. Biotechnol.* 166, 72–86. doi:10.1007/s12010-011-9405-x
- Bao, L., Huang, Q., Chang, L., Zhou, J., Lu, H., 2011. Screening and characterization of a cellulase with endocellulase and exocellulase activity from yak rumen metagenome. *J. Mol. Catal. B Enzym.* 73, 104–110. doi:10.1016/j.molcatb.2011.08.006
- Bayer, E.A., Belaich, J.-P., Shoham, Y., Lamed, R., 2004. The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides. *Annu. Rev. Microbiol.* 58, 521–554. doi:10.1146/annurev.micro.57.030502.091022
- Beg, Q.K., Kapoor, M., Mahajan, L., Hoondal, G.S., 2001. Microbial xylanases and their industrial applications: a review. *Appl. Microbiol. Biotechnol.* 56, 326–338. doi:10.1007/s002530100704
- Beloqui, A., Pita, M., Polaina, J., Martinez-Arias, A., Golyshina, O.V., Zumarraga, M., Yakimov, M.M., Garcia-Arellano, H., Alcalde, M., Fernandez, V.M., Elborough, K., Andreu, J.M., Ballesteros, A., Plou, F.J., Timmis, K.N., Ferrer, M., Golyshin, P.N., 2006. Novel Polyphenol Oxidase Mined from a Metagenome Expression Library of Bovine Rumen: BIOCHEMICAL PROPERTIES, STRUCTURAL ANALYSIS, AND PHYLOGENETIC RELATIONSHIPS. *J. Biol. Chem.* 281, 22933–22942. doi:10.1074/jbc.M600577200
- Berg Miller, M.E., Antonopoulos, D.A., Rincon, M.T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B., Coutinho, P.M., Borovok, I., Jindou, S., Lamed, R., Flint, H.J., Bayer, E.A., White, B.A., 2009. Diversity and Strain Specificity of Plant Cell Wall Degrading Enzymes Revealed by the Draft Genome of *Ruminococcus flavefaciens* FD-1. *PLoS ONE* 4, e6650. doi:10.1371/journal.pone.0006650
- Berg Miller, M.E., Yeoman, C.J., Chia, N., Tringe, S.G., Angly, F.E., Edwards, R.A., Flint, H.J., Lamed, R., Bayer, E.A., White, B.A., 2012. Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome: Rumen viral metagenome. *Environ. Microbiol.* 14, 207–227. doi:10.1111/j.1462-2920.2011.02593.x
- Bjursell, M.K., Martens, E.C., Gordon, J.I., 2006. Functional Genomic and Metabolic Studies of the Adaptations of a Prominent Adult Human Gut Symbiont, *Bacteroides thetaiotaomicron*, to the Suckling Period. *J. Biol. Chem.* 281, 36269–36279. doi:10.1074/jbc.M606509200
- Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J., 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769–781. doi:10.1042/BJ20040892
- Borneman, W.S., Ljungdahl, L.G., Hartley, R.D., Akin, D.E., 1991. Isolation and characterization of p-coumaroyl esterase from the anaerobic fungus *Neocallimastix* strain MC-2. *Appl. Environ. Microbiol.* 57, 2337–2344.

- Brulc, J.M., Antonopoulos, D.A., Berg Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., Edwards, R.E., Frank, E.D., Emerson, J.B., Wacklin, P., Coutinho, P.M., Henrissat, B., Nelson, K.E., White, B.A., 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci.* 106, 1948–1953. doi:10.1073/pnas.0806191105
- Burton, R.A., Gidley, M.J., Fincher, G.B., 2010. Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat. Chem. Biol.* 6, 724–732. doi:10.1038/nchembio.439
- Calabrò, S., 2015. Plant Secondary Metabolites, in: Puniya, A.K., Singh, R., Kamra, D.N. (Eds.), *Rumen Microbiology: From Evolution to Revolution*. Springer India, New Delhi, pp. 153–159.
- Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré, J., Henrissat, B., Remaud-Siméon, M., Monsan, P., Potocki-Véronèse, G., 2013. Functional Metagenomics Reveals Novel Pathways of Prebiotic Breakdown by Human Gut Bacteria. *PLoS ONE* 8, e72766. doi:10.1371/journal.pone.0072766
- Chang, L., Ding, M., Bao, L., Chen, Y., Zhou, J., Lu, H., 2011. Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. *Appl. Microbiol. Biotechnol.* 90, 1933–1942. doi:10.1007/s00253-011-3182-x
- Chaudhary, P.P., Dagar, S.S., Sirohi, S.K., 2012. Comparative quantification of major rumen microbial population in Indian Cattle and Buffalo fed on wheat straws based diet. *Prime J. Microbiol. Res.* 2, 105–108.
- Chávez-González, M., Rodríguez-Durán, L.V., Balagurusamy, N., Prado-Barragán, A., Rodríguez, R., Contreras, J.C., Aguilar, C.N., 2012. Biotechnological Advances and Challenges of Tannase: An Overview. *Food Bioprocess Technol.* 5, 445–459. doi:10.1007/s11947-011-0608-5
- Cheema, T.A., Jirajoenrat, K., Sirinarumitr, T., Rakshit, S.K., 2012. Isolation of a Gene Encoding a Cellulolytic Enzyme from Swamp Buffalo Rumen Metagenomes and Its Cloning and Expression in *Escherichia Coli*. *Anim. Biotechnol.* 23, 261–277. doi:10.1080/10495398.2012.722156
- Cheng, F., Sheng, J., Cai, T., Jin, J., Liu, W., Lin, Y., Du, Y., Zhang, M., Shen, L., 2012a. A Protease-Insensitive Feruloyl Esterase from China Holstein Cow Rumen Metagenomic Library: Expression, Characterization, and Utilization in Ferulic Acid Release from Wheat Straw. *J. Agric. Food Chem.* 60, 2546–2553. doi:10.1021/jf204556u
- Cheng, F., Sheng, J., Dong, R., Men, Y., Gan, L., Shen, L., 2012b. Novel Xylanase from a Holstein Cattle Rumen Metagenomic Library and Its Application in Xylooligosaccharide and Ferulic Acid Production from Wheat Straw. *J. Agric. Food Chem.* 60, 12516–12524. doi:10.1021/jf302337w
- Cheng, K.J., Lee, S.S., Bae, H.D., Ha, J.K., 1999. Industrial Applications of Rumen Microbes - Review -. *Asian-Australas. J. Anim. Sci.* 12, 84–92. doi:10.5713/ajas.1999.84
- Cheng, K.J., McCowan, R.P., Costerton, J.W., 1979. Adherent epithelial bacteria in ruminants and their roles in digestive tract function. *Am. J. Clin. Nutr.* 32, 139–148.
- Choudhury, P.K., Salem, A.Z.M., Kumar, R.J., Singh, R., Puniya, A.K., 2015. Rumen Microbiology: an overview, in: *Rumen Microbiology: From Evolution to Revolution*. Springer India, pp. 3–16.
- Cosgrove, D.J., 2005. Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* 6, 850–861. doi:10.1038/nrm1746
- Coutinho, P.M., Deleury, E., Davies, G.J., Henrissat, B., 2003. An Evolving Hierarchical Family Classification for Glycosyltransferases. *J. Mol. Biol.* 328, 307–317. doi:10.1016/S0022-2836(03)00307-3
- Cybinski, D.H., Layton, I., Lowry, J.B., Dalrymple, B.P., 1999. An acetylxylan esterase and a xylanase expressed from genes cloned from the ruminal fungus *Neocallimastix patriciarum* act synergistically to degrade acetylated xylans. *Appl. Microbiol. Biotechnol.* 52, 221–225. doi:10.1007/s002530051512
- Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., Liu, L., Luo, Y., Liu, D., Zheng, H., Wang, J., Dong, Z., Hu, S., Huang, L., 2015. Metatranscriptomic Analyses of Plant Cell Wall Polysaccharide

- Degradation by Microorganisms in the Cow Rumen. *Appl. Environ. Microbiol.* 81, 1375–1386. doi:10.1128/AEM.03682-14
- Dai, X., Zhu, Y., Luo, Y., Song, L., Liu, D., Liu, L., Chen, F., Wang, M., Li, J., Zeng, X., Dong, Z., Hu, S., Li, L., Xu, J., Huang, L., Dong, X., 2012. Metagenomic Insights into the Fibrolytic Microbiome in Yak Rumen. *PLoS ONE* 7, e40430. doi:10.1371/journal.pone.0040430
- Dassa, B., Borovok, I., Ruimy-Israeli, V., Lamed, R., Flint, H.J., Duncan, S.H., Henrissat, B., Coutinho, P., Morrison, M., Mosoni, P., Yeoman, C.J., White, B.A., Bayer, E.A., 2014. Rumen Cellulosomics: Divergent Fiber-Degrading Strategies Revealed by Comparative Genome-Wide Analysis of Six Ruminococcal Strains. *PLoS ONE* 9, e99221. doi:10.1371/journal.pone.0099221
- Davies, D.R., Theodorou, M.K., Lawrence, M.I.G., Trinci, A.P.J., 1993. Distribution of anaerobic fungi in the digestive tract of cattle and their survival in faeces. *J. Gen. Microbiol.* 139, 1395–1400. doi:10.1099/00221287-139-6-1395
- Del Pozo, M.V., Fernández-Arrojo, L., Gil-Martínez, J., Montesinos, A., Chernikova, T.N., Nechitaylo, T.Y., Waliszek, A., Tortajada, M., Rojas, A., Huws, S.A., Golyshina, O.V., Newbold, C.J., Polaina, J., Ferrer, M., Golyshin, P.N., 2012. Microbial β -glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. *Biotechnol. Biofuels* 5, 73. doi:10.1186/1754-6834-5-73
- Deusch, S., Seifert, J., 2015. Catching the tip of the iceberg - Evaluation of sample preparation protocols for metaproteomic studies of the rumen microbiota. *PROTEOMICS* n/a–n/a. doi:10.1002/pmic.201400556
- Deusch, S., Tilocca, B., Camarinha-Silva, A., Seifert, J., 2015. News in livestock research — use of Omics-technologies to study the microbiota in the gastrointestinal tract of farm animals. *Comput. Struct. Biotechnol. J.* 13, 55–63. doi:10.1016/j.csbj.2014.12.005
- Duan, C.-J., Liu, J.-L., Wu, X., Tang, J.-L., Feng, J.-X., 2010. Novel Carbohydrate-Binding Module Identified in a Ruminal Metagenomic Endoglucanase. *Appl. Environ. Microbiol.* 76, 4867–4870. doi:10.1128/AEM.00011-10
- Duan, C.-J., Xian, L., Zhao, G.-C., Feng, Y., Pang, H., Bai, X.-L., Tang, J.-L., Ma, Q.-S., Feng, J.-X., 2009. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J. Appl. Microbiol.* 107, 245–256. doi:10.1111/j.1365-2672.2009.04202.x
- Ekstrom, A., Taujale, R., McGinn, N., Yin, Y., 2014. PlantCAZyme: a database for plant carbohydrate-active enzymes. *Database* 2014, bau079–bau079. doi:10.1093/database/bau079
- Fay, J.P., Jakober, K.D., Cheng, K.-J., Costerton, J.W., 1990. Esterase activity of pure cultures of rumen bacteria as expressed by the hydrolysis of *p*-nitrophenylpalmitate. *Can. J. Microbiol.* 36, 585–589. doi:10.1139/m90-103
- Ferrer, M., Beloqui, A., Golyshina, O.V., Plou, F.J., Neef, A., Chernikova, T.N., Fernández-Arrojo, L., Ghazi, I., Ballesteros, A., Elborough, K., Timmis, K.N., Golyshin, P.N., 2007. Biochemical and structural features of a novel cyclodextrinase from cow rumen metagenome. *Biotechnol. J.* 2, 207–213. doi:10.1002/biot.200600183
- Ferrer, M., Ghazi, A., Beloqui, A., Vieites, J.M., López-Cortés, N., Marín-Navarro, J., Nechitaylo, T.Y., Guazzaroni, M.-E., Polaina, J., Waliczek, A., Chernikova, T.N., Reva, O.N., Golyshina, O.V., Golyshin, P.N., 2012. Functional Metagenomics Unveils a Multifunctional Glycosyl Hydrolase from the Family 43 Catalysing the Breakdown of Plant Polymers in the Calf Rumen. *PLoS ONE* 7, e38134. doi:10.1371/journal.pone.0038134
- Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A.P.M.D., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N., Golyshin, P.N., 2005. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora: Enzymatic diversity from bovine rumen metagenome. *Environ. Microbiol.* 7, 1996–2010. doi:10.1111/j.1462-2920.2005.00920.x
- Findley, S.D., Mormile, M.R., Sommer-Hurley, A., Zhang, X.-C., Tipton, P., Arnett, K., Porter, J.H., Kerley, M., Stacey, G., 2011. Activity-Based Metagenomic Screening and Biochemical

- Characterization of Bovine Ruminal Protozoan Glycoside Hydrolases. *Appl. Environ. Microbiol.* 77, 8106–8113. doi:10.1128/AEM.05925-11
- Fouts, D.E., Szpakowski, S., Purushe, J., Torralba, M., Waterman, R.C., MacNeil, M.D., Alexander, L.J., Nelson, K.E., 2012. Next Generation Sequencing to Define Prokaryotic and Fungal Diversity in the Bovine Rumen. *PLoS ONE* 7, e48289. doi:10.1371/journal.pone.0048289
- Gilbert, R.A., Klieve, A.V., 2015. Ruminal Viruses (Bacteriophages, Archaeaphages), in: *Rumen Microbiology: From Evolution to Revolution*. Springer India, pp. 3–16.
- Goel, G., Dagar, S.S., Raghav, M., Bansal, S., 2015. Rumen: an underutilised niche for industrially important enzymes, in: *Rumen Microbiology: From Evolution to Revolution*. Springer India, pp. 247–263.
- Goel, G., Puniya, A.K., Aguilar, C.N., Singh, K., 2005. Interaction of gut microflora with tannins in feeds. *Naturwissenschaften* 92, 497–503. doi:10.1007/s00114-005-0040-7
- Gong, X., Gruninger, R.J., Qi, M., Paterson, L., Forster, R.J., Teather, R.M., McAllister, T.A., 2012. Cloning and identification of novel hydrolase genes from a dairy cow rumen metagenomic library and characterization of a cellulase gene. *BMC Res. Notes* 5, 566. doi:10.1186/1756-0500-5-566
- Gong, X., Gruninger, R.J., Forster, R.J., Teather, R.M., McAllister, T.A., 2013. Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. *Appl. Microbiol. Biotechnol.* 97, 2423–2431. doi:10.1007/s00253-012-4088-y
- Gruninger, R.J., Gong, X., Forster, R.J., McAllister, T.A., 2014. Biochemical and kinetic characterization of the multifunctional β -glucosidase/ β -xylosidase/ α -arabinosidase, Bgxa1. *Appl. Microbiol. Biotechnol.* 98, 3003–3012. doi:10.1007/s00253-013-5191-4
- Gruninger, R.J., Puniya, A.K., Callaghan, T.M., Edwards, J.E., Youssef, N., Dagar, S.S., Fliegerova, K., Griffith, G.W., Forster, R., Tsang, A., McAllister, T., Elshahed, M.S., 2014. Anaerobic fungi (phylum *Neocallimastigomycota*): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiol. Ecol.* 90, 1–17. doi:10.1111/1574-6941.12383
- Guettler, M.V., Rumler, D., Jain, M.K., 1999. *Actinobacillus succinogenes* sp. nov., a novel succinic acid-producing strain from the bovine rumen. *Int. J. Syst. Bacteriol.* 49, 207–216. doi:10.1099/00207713-49-1-207
- Hall, J.B., Silver, S., 2009. *Nutrition and Feeding of the Cow-Calf Herd: Digestive System of the Cow*.
- Henderson, C., 1971. A Study of the Lipase Produced by *Anaerovibrio lipolytica*, a Rumen Bacterium. *J. Gen. Microbiol.* 65, 81–89. doi:10.1099/00221287-65-1-81
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., Davies, G., 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* 92, 7090–7094.
- Henrissat, B., Davies, G., 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644. doi:10.1016/S0959-440X(97)80072-3
- Hess, M., Szczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z., Rubin, E.M., 2011. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* 331, 463–467. doi:10.1126/science.1200387
- Highlander, S.K., 2014. *Encyclopedia of metagenomics: environmental metagenomics*. Springer, New York.
- Horn, S., Vaaje-Kolstad, G., Westereng, B., Eijsink, V.G., 2012. Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels* 5, 45. doi:10.1186/1754-6834-5-45
- Hou, J., Shen, Y., Jiao, C., Ge, R., Zhang, X., Bao, X., 2015. Characterization and evolution of xylose isomerase screened from the bovine rumen metagenome in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* doi:10.1016/j.jbiosc.2015.05.014
- Huang, H., Zhang, R., Fu, D., Luo, J., Li, Z., Luo, H., Shi, P., Yang, P., Diao, Q., Yao, B., 2011. Diversity, abundance and characterization of ruminal cysteine phytases suggest their important role in

- phytate degradation: Diversity and abundance of ruminal cysteine phytases. *Environ. Microbiol.* 13, 747–757. doi:10.1111/j.1462-2920.2010.02379.x
- Jami, E., Israel, A., Kotser, A., Mizrahi, I., 2013. Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J.* 7, 1069–1079. doi:10.1038/ismej.2013.2
- Jami, E., Mizrahi, I., 2012. Composition and Similarity of Bovine Rumen Microbiota across Individual Animals. *PLoS ONE* 7, e33306. doi:10.1371/journal.pone.0033306
- Janssen, P.H., Kirs, M., 2008. Structure of the Archaeal Community of the Rumen. *Appl. Environ. Microbiol.* 74, 3619–3625. doi:10.1128/AEM.02812-07
- Jayani, R.S., Saxena, S., Gupta, R., 2005. Microbial pectinolytic enzymes: A review. *Process Biochem.* 40, 2931–2944. doi:10.1016/j.procbio.2005.03.026
- Jenkins, T.C., 1993. Lipid Metabolism in the Rumen. *J. Dairy Sci.* 76, 3851–3863. doi:10.3168/jds.S0022-0302(93)77727-9
- Jenkins, T.C., Wallace, R.J., Moate, P.J., Mosley, E.E., 2008. Board-invited review: Recent advances in biohydrogenation of unsaturated fatty acids within the rumen microbial ecosystem. *J. Anim. Sci.* 86, 397–412. doi:10.2527/jas.2007-0588
- Kambiranda, D.M., Asraful-Islam, S.M., Cho, K.M., Math, R.K., Lee, Y.H., Kim, H., Yun, H.D., 2009. Expression of esterase gene in yeast for organophosphates biodegradation. *Pestic. Biochem. Physiol.* 94, 15–20. doi:10.1016/j.pestbp.2009.02.006
- Kamra, D.N., Agarwal, N., Chaudhary, L.C., 2015. Manipulation of Rumen Microbial Ecosystem for Reducing Enteric Methane Emission in Livestock, in: Sejian, V., Gaughan, J., Baumgard, L., Prasad, C. (Eds.), *Climate Change Impact on Livestock: Adaptation and Mitigation*. Springer India, New Delhi, pp. 255–272.
- Khanal, R.C., Dhiman, T.R., 2004. Biosynthesis of conjugated Linoleic Acid (CLA): A review. *Pak. J. Nutr.* 3, 72–81.
- Kiemle, S.N., Zhang, X., Esker, A.R., Toriz, G., Gatenholm, P., Cosgrove, D.J., 2014. Role of (1,3)(1,4)- β -Glucan in Cell Walls: Interaction with Cellulose. *Biomacromolecules* 15, 1727–1736. doi:10.1021/bm5001247
- Kim, D.Y., Yim, S.C., Lee, P.C., Lee, W.G., Lee, S.Y., Chang, H.N., 2004. Batch and continuous fermentation of succinic acid from wood hydrolysate by *Mannheimia succiniciproducens* MBEL55E. *Enzyme Microb. Technol.* 35, 648–653. doi:10.1016/j.enzmictec.2004.08.018
- Kim, M.K., 2012. Cloning and Identification of a New Group Esterase (Est5S) from Noncultured Rumen Bacterium. *J. Microbiol. Biotechnol.* 22, 1044–1053. doi:10.4014/jmb.1201.12070
- Ko, K.-C., Lee, J.H., Han, Y., Choi, J.H., Song, J.J., 2013. A novel multifunctional cellulolytic enzyme screened from metagenomic resources representing ruminal bacteria. *Biochem. Biophys. Res. Commun.* 441, 567–572. doi:10.1016/j.bbrc.2013.10.120
- Krause, D.O., Nagaraja, T.G., Wright, A.D.G., Callaway, T.R., 2013. Board-invited review: Rumen microbiology: Leading the way in microbial ecology. *J. Anim. Sci.* 91, 331–341. doi:10.2527/jas.2012-5567
- Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tausin, A.S., Spadiut, O., Klintner, S., Pudlo, N.A., Urs, K., Koropatkin, N.M., Creagh, A.L., Haynes, C.A., Kelly, A.G., Cederholm, S.N., Davies, G.J., Martens, E.C., Brumer, H., 2014. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* 506, 498–502. doi:10.1038/nature12907
- Lazuka, A., Auer, L., Bozonnet, S., Morgavi, D.P., O'Donohue, M., Hernandez-Raquet, G., 2015. Efficient anaerobic transformation of raw wheat straw by a robust cow rumen-derived microbial consortium. *Bioresour. Technol.* 196, 241–249. doi:10.1016/j.biortech.2015.07.084
- Lee, C.C., Kibblewhite, R.E., Wagschal, K., Li, R., Orts, W.J., 2012. Isolation of α -Glucuronidase Enzyme from a Rumen Metagenomic Library. *Protein J.* 31, 206–211. doi:10.1007/s10930-012-9391-z
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., Henrissat, B., 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6, 41. doi:10.1186/1754-6834-6-41

- Li, M., Zhou, M., Adamowicz, E., Basarab, J.A., Guan, L.L., 2012. Characterization of bovine ruminal epithelial bacterial communities using 16S rRNA sequencing, PCR-DGGE, and qRT-PCR analysis. *Vet. Microbiol.* 155, 72–80. doi:10.1016/j.vetmic.2011.08.007
- Li, R.W., Wu, S., Li, W., Huang, Y., Gasbarre, L.C., 2011. Metagenome Plasticity of the Bovine Abomasal Microbiota in Immune Animals in Response to *Ostertagia Ostertagi* Infection. *PLoS ONE* 6, e24417. doi:10.1371/journal.pone.0024417
- Liu, K., Wang, J., Bu, D., Zhao, S., McSweeney, C., Yu, P., Li, D., 2009. Isolation and biochemical characterization of two lipases from a metagenomic library of China Holstein cow rumen. *Biochem. Biophys. Res. Commun.* 385, 605–611. doi:10.1016/j.bbrc.2009.05.110
- Liu, L., Feng, Y., Duan, C.-J., Pang, H., Tang, J.-L., Feng, J.-X., 2009. Isolation of a gene encoding endoglucanase activity from uncultured microorganisms in buffalo rumen. *World J. Microbiol. Biotechnol.* 25, 1035–1042. doi:10.1007/s11274-009-9983-8
- Liu, Y.-S., Baker, J.O., Zeng, Y., Himmel, M.E., Haas, T., Ding, S.-Y., 2011. Cellobiohydrolase Hydrolyzes Crystalline Cellulose on Hydrophobic Faces. *J. Biol. Chem.* 286, 11195–11201. doi:10.1074/jbc.M110.216556
- Li, Y., Liu, N., Yang, H., Zhao, F., Yu, Y., Tian, Y., Lu, X., 2014. Cloning and characterization of a new β -Glucosidase from a metagenomic library of Rumen of cattle feeding with *Miscanthus sinensis*. *BMC Biotechnol.* 14, 85. doi:10.1186/1472-6750-14-85
- Loaces, I., Amarelle, V., Muñoz-Gutierrez, I., Fabiano, E., Martinez, A., Noya, F., 2015. Improved ethanol production from biomass by a rumen metagenomic DNA fragment expressed in *Escherichia coli* MS04 during fermentation. *Appl. Microbiol. Biotechnol.* doi:10.1007/s00253-015-6801-0
- Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M., Henrissat, B., 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J.* 432, 437–444. doi:10.1042/BJ20101185
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–495. doi:10.1093/nar/gkt1178
- López-Cortés, N., Reyes-Duarte, D., Beloqui, A., Polaina, J., Ghazi, I., Golyshina, O.V., Ballesteros, A., Golyshin, P.N., Ferrer, M., 2007. Catalytic role of conserved HQGE motif in the CE6 carbohydrate esterase family. *FEBS Lett.* 581, 4657–4662. doi:10.1016/j.febslet.2007.08.060
- Math, R.K., Asraful Islam, S.M., Cho, K.M., Hong, S.J., Kim, J.M., Yun, M.G., Cho, J.J., Heo, J.Y., Lee, Y.H., Kim, H., Yun, H.D., 2010. Isolation of a novel gene encoding a 3,5,6-trichloro-2-pyridinol degrading enzyme from a cow rumen metagenomic library. *Biodegradation* 21, 565–573. doi:10.1007/s10532-009-9324-5
- McCowan, R.P., Cheng, K.J., Bailey, C.B., Costerton, J.W., 1978. Adhesion of bacteria to epithelial cell surfaces within the reticulo-rumen of cattle. *Appl. Environ. Microbiol.* 35, 149–155.
- Mittal, A., Singh, G., Goyal, V., Yadav, A., Aneja, K.R., Gautam, S.K., Aggarwal, N.K., 2011. Isolation and biochemical characterization of acido-thermophilic extracellular phytase producing bacterial strain for potential application in poultry feed. *Jundishapur J. Microbiol.* 4.
- Moon, Y.H., Iakiviak, M., Bauer, S., Mackie, R.I., Cann, I.K.O., 2011. Biochemical Analyses of Multiple Endoxylanases from the Rumen Bacterium *Ruminococcus albus* 8 and Their Synergistic Activities with Accessory Hemicellulose-Degrading Enzymes. *Appl. Environ. Microbiol.* 77, 5157–5169. doi:10.1128/AEM.00353-11
- Morgavi, D.P., Forano, E., Martin, C., Newbold, C.J., 2010. Microbial ecosystem and methanogenesis in ruminants. *animal* 4, 1024–1036. doi:10.1017/S1751731110000546
- Morgavi, D.P., Kelly, W.J., Janssen, P.H., Attwood, G.T., 2013. Rumen microbial (meta)genomics and its application to ruminant production. *animal* 7, 184–201. doi:10.1017/S1751731112000419
- Murad, H.A., Azzaz, H.H., 2011. Microbial Pectinases and Ruminant Nutrition. *Res. J. Microbiol.* 6, 246–269. doi:10.3923/jm.2011.246.269

- Nghiem, N.P., Davison, B.H., Suttle, B.E., Richardson, G.R., 1997. Production of succinic acid by anaerobiospirillum succiniciproducens. *Appl. Biochem. Biotechnol.* 63-65, 565–576. doi:10.1007/BF02920454
- Nguyen, N.H., Maruset, L., Uengwetwanit, T., Mhuantong, W., Harnpicharnchai, P., Champreda, V., Tanapongpipat, S., Jirajaroenrat, K., Rakshit, S.K., Eurwilaichitr, L., Pongpattanakitshote, S., 2012. Identification and Characterization of a Cellulase-Encoding Gene from the Buffalo Rumen Metagenomic Library. *Biosci. Biotechnol. Biochem.* 76, 1075–1084. doi:10.1271/bbb.110786
- Palackal, N., Lyon, C.S., Zaidi, S., Luginbühl, P., Dupree, P., Goubet, F., Macomber, J.L., Short, J.M., Hazlewood, G.P., Robertson, D.E., Steer, B.A., 2007. A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Appl. Microbiol. Biotechnol.* 74, 113–124. doi:10.1007/s00253-006-0645-6
- Park, B.H., Karpinets, T.V., Syed, M.H., Leuze, M.R., Uberbacher, E.C., 2010. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* 20, 1574–1584. doi:10.1093/glycob/cwq106
- Patel, A.B., Patel, A.K., Shah, M.P., Parikh, I.K., Joshi, C.G., 2015. Isolation and characterization of novel multifunctional recombinant family 26 glycoside hydrolase from mehsani buffalo rumen metagenome. *Biotechnol. Appl. Biochem.* n/a–n/a. doi:10.1002/bab.1358
- Peng, S., Yin, J., Liu, X., Jia, B., Chang, Z., Lu, H., Jiang, N., Chen, Q., 2015. First insights into the microbial diversity in the omasum and reticulum of bovine using Illumina sequencing. *J. Appl. Genet.* 56, 393–401. doi:10.1007/s13353-014-0258-1
- Pope, P.B., Denman, S.E., Jones, M., Tringe, S.G., Barry, K., Malfatti, S.A., McHardy, A.C., Cheng, J.-F., Hugenholtz, P., McSweeney, C.S., Morrison, M., 2010. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc. Natl. Acad. Sci.* 107, 14793–14798. doi:10.1073/pnas.1005297107
- Pope, P.B., Mackenzie, A.K., Gregor, I., Smith, W., Sundset, M.A., McHardy, A.C., Morrison, M., Eijsink, V.G.H., 2012. Metagenomics of the Svalbard Reindeer Rumen Microbiome Reveals Abundance of Polysaccharide Utilization Loci. *PLoS ONE* 7, e38571. doi:10.1371/journal.pone.0038571
- Pradhan, S.S., Rathi, C.L., 2008. Novel compositions for biobleaching coupled with stone washing of indigo dyed denims and process thereof. WO2008023386 A2.
- Privé, F., Newbold, C.J., Kaderbhai, N.N., Girdwood, S.G., Golyshina, O.V., Golyshin, P.N., Scollan, N.D., Huws, S.A., 2015. Isolation and characterization of novel lipases/esterases from a bovine rumen metagenome. *Appl. Microbiol. Biotechnol.* 99, 5475–5485. doi:10.1007/s00253-014-6355-6
- Puniya, A.K., Singh, R., Kamra, D.N., 2015. *Rumen Microbiology: From Evolution to Revolution.* Springer India, New Delhi.
- Purushe, J., Morrison, M., White, B.A., Mackie, R.I., the North American Consortium for Rumen Bacteria, Coutinho, P.M., Henrissat, B., Nelson, K.E., 2010. Comparative Genome Analysis of *Prevotella ruminicola* and *Prevotella bryantii*: Insights into Their Environmental Niche. *Microb. Ecol.* 60, 721–729. doi:10.1007/s00248-010-9692-8
- Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., Selinger, L.B., Butler, G., Tsang, A., McAllister, T.A., Forster, R.J., 2011. Snapshot of the Eukaryotic Gene Expression in Muskoxen Rumen—A Metatranscriptomic Approach. *PLoS ONE* 6, e20521. doi:10.1371/journal.pone.0020521
- Ramirez, H.A.R., Nestor, K., Tedeschi, L.O., Callaway, T.R., Dowd, S.E., Fernando, S.C., Kononoff, P.J., 2012. The effect of brown midrib corn silage and dried distillers' grains with solubles on milk production, nitrogen utilization and microbial community structure in dairy cows. *Can. J. Anim. Sci.* 92, 365–380. doi:10.4141/cjas2011-133
- Rashamuse, K.J., Visser, D.F., Hennessy, F., Kemp, J., Roux-van der Merwe, M.P., Badenhorst, J., Ronneburg, T., Francis-Pope, R., Brady, D., 2013. Characterisation of Two Bifunctional

- Cellulase–Xylanase Enzymes Isolated from a Bovine Rumen Metagenome Library. *Curr. Microbiol.* 66, 145–151. doi:10.1007/s00284-012-0251-z
- Ricard, G., McEwan, N.R., Dutilh, B.E., Jouany, J.-P., Macheboeuf, D., Mitsumori, M., McIntosh, F.M., Michalowski, T., Nagamine, T., Nelson, N., Newbold, C.J., Nsabimana, E., Takenaka, A., Thomas, N.A., Ushida, K., Hackstein, J.H.P., Huynen, M.A., 2006. Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC Genomics* 7, 22. doi:10.1186/1471-2164-7-22
- Rincon, M.T., Dassa, B., Flint, H.J., Travis, A.J., Jindou, S., Borovok, I., Lamed, R., Bayer, E.A., Henrissat, B., Coutinho, P.M., Antonopoulos, D.A., Berg Miller, M.E., White, B.A., 2010. Abundance and Diversity of Dockerin-Containing Proteins in the Fiber-Degrading Rumen Bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS ONE* 5, e12476. doi:10.1371/journal.pone.0012476
- Rodríguez-Durán, L.V., Valdivia-Urdiales, B., Contreras-Esquivel, J.C., Rodríguez-Herrera, R., Aguilar, C.N., 2011. Novel Strategies for Upstream and Downstream Processing of Tannin Acyl Hydrolase. *Enzyme Res.* 2011, 1–20. doi:10.4061/2011/823619
- Rodríguez, M.C., Loaces, I., Amarelle, V., Senatore, D., Iriarte, A., Fabiano, E., Noya, F., 2015. Est10: A Novel Alkaline Esterase Isolated from Bovine Rumen Belonging to the New Family XV of Lipolytic Enzymes. *PLOS ONE* 10, e0126651. doi:10.1371/journal.pone.0126651
- Rosewarne, C.P., Pope, P.B., Cheung, J.L., Morrison, M., 2014. Analysis of the bovine rumen microbiome reveals a diversity of Sus-like polysaccharide utilization loci from the bacterial phylum Bacteroidetes. *J. Ind. Microbiol. Biotechnol.* 41, 601–606. doi:10.1007/s10295-013-1395-y
- Ross, E.M., Petrovski, S., Moate, P.J., Hayes, B.J., 2013. Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. *BMC Microbiol.* 13, 242. doi:10.1186/1471-2180-13-242
- Ruiz-Dueñas, F.J., Martínez, Á.T., 2009. Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.* 2, 164–177. doi:10.1111/j.1751-7915.2008.00078.x
- Sadet, S., Martin, C., Meunier, B., Morgavi, D.P., 2007. PCR-DGGE analysis reveals a distinct diversity in the bacterial population attached to the rumen epithelium. *animal* 1, 939. doi:10.1017/S1751731107000304
- Saleem, F., Bouatra, S., Guo, A.C., Psychogios, N., Mandal, R., Dunn, S.M., Ametaj, B.N., Wishart, D.S., 2013. The Bovine Ruminant Fluid Metabolome. *Metabolomics* 9, 360–378. doi:10.1007/s11306-012-0458-9
- Sauer, M., Marx, H., Mattanovich, D., 2012. From rumen to industry. *Microb. Cell Factories* 11, 121. doi:10.1186/1475-2859-11-121
- Sharma, N., Rathore, M., Sharma, M., 2013. Microbial pectinase: sources, characterization and applications. *Rev. Environ. Sci. Biotechnol.* 12, 45–60. doi:10.1007/s11157-012-9276-9
- Shedova, E.N., Berezina, O.V., Lunina, N.A., Zverlov, V.V., Schwarz, W.H., Velikodvorskaya, G.A., 2009. Cloning and characterisation of a large metagenomic DNA fragment containing glycosyl-hydrolase genes. *Mol. Genet. Microbiol. Virol.* 24, 12–16. doi:10.3103/S0891416809010030
- Shrivastava, B., Thakur, S., Khasa, Y.P., Gupte, A., Puniya, A.K., Kuhad, R.C., 2011. White-rot fungal conversion of wheat straw to energy rich cattle feed. *Biodegradation* 22, 823–831. doi:10.1007/s10532-010-9408-2
- Silley, P., 1985. A note on the pectinolytic enzymes of *Lachnospira multiparus*. *J. Appl. Bacteriol.* 58, 145–149. doi:10.1111/j.1365-2672.1985.tb01441.x
- Singh, B., Bhat, T.K., Sharma, O.P., 2001. Biodegradation of tannic acid in an in vitro ruminal system. *Livest. Prod. Sci.* 68, 259–262. doi:10.1016/S0301-6226(00)00227-X
- Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., Dore, J., Monsan, P., Remaud-Simeon, M., Potocki-Veronese, G., 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* 20, 1605–1612. doi:10.1101/gr.108332.110
- Terrapon, N., Henrissat, B., 2014. How do gut microbes break down dietary fiber? *Trends Biochem. Sci.* 39, 156–158. doi:10.1016/j.tibs.2014.02.005

- Tymensen, L.D., Beauchemin, K.A., McAllister, T.A., 2012. Structures of free-living and protozoa-associated methanogen communities in the bovine rumen differ according to comparative analysis of 16S rRNA and mcrA genes. *Microbiology* 158, 1808–1817. doi:10.1099/mic.0.057984-0
- Ufarté, L., Gabrielle Potocki-Veronese, Elisabeth Laville, 2015. Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology. *Front. Microbiol.* 6. doi:10.3389/fmicb.2015.00563
- Underwood, W., 2012. The Plant Cell Wall: A Dynamic Barrier Against Pathogen Invasion. *Front. Plant Sci.* 3. doi:10.3389/fpls.2012.00085
- Vaaje-Kolstad, G., Westereng, B., Horn, S.J., Liu, Z., Zhai, H., Sorlie, M., Eijsink, V.G.H., 2010. An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science* 330, 219–222. doi:10.1126/science.1192231
- van Vliet, P.C.J., Reijs, J.W., Bloem, J., Dijkstra, J., de Goede, R.G.M., 2007. Effects of Cow Diet on the Microbial Community and Organic Matter and Nitrogen Content of Feces. *J. Dairy Sci.* 90, 5146–5158. doi:10.3168/jds.2007-0065
- Veneault-Fourrey, C., Commun, C., Kohler, A., Morin, E., Balestrini, R., Plett, J., Danchin, E., Coutinho, P., Wiebenga, A., de Vries, R.P., Henrissat, B., Martin, F., 2014. Genomic and transcriptomic analysis of *Laccaria bicolor* CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. *Fungal Genet. Biol.* 72, 168–181. doi:10.1016/j.fgb.2014.08.007
- Virkel, G., Carletti, M., Cantiello, M., Della Donna, L., Gardini, G., Girolami, F., Nebbia, C., 2009. Characterization of xenobiotic metabolizing enzymes in bovine small intestinal mucosa: Xenobiotic metabolizing enzymes in bovine intestinal mucosa. *J. Vet. Pharmacol. Ther.* 33, 295–303. doi:10.1111/j.1365-2885.2009.01137.x
- Wang, F., Li, F., Chen, G., Liu, W., 2009. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. *Microbiol. Res.* 164, 650–657. doi:10.1016/j.micres.2008.12.002
- Wang, L., Hatem, A., Catalyurek, U.V., Morrison, M., Yu, Z., 2013. Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PloS One* 8, e78507. doi:10.1371/journal.pone.0078507
- Wang, M., Chen, F., Zhang, S., Zhu, Y., Dong, X., Huang, L., Tian, R., Dong, Z., Dai, X., 2011. [Analysis of xylanases derived from the metagenomic BAC clone library of yak rumen]. *Wei Sheng Wu Xue Bao* 51, 1364–1373.
- White, B.A., Lamed, R., Bayer, E.A., Flint, H.J., 2014. Biomass Utilization by Gut Microbiomes*. *Annu. Rev. Microbiol.* 68, 279–296. doi:10.1146/annurev-micro-092412-155618
- Whitney, S.E.C., Brigham, J.E., Darke, A.H., Reid, J.S.G., Gidley, M.J., 1998. Structural aspects of the interaction of mannan-based polysaccharides with bacterial cellulose. *Carbohydr. Res.* 307, 299–309. doi:10.1016/S0008-6215(98)00004-4
- Williams, A.G., Coleman, G.S., 1992. *The Rumen Protozoa*. Springer New York, New York, NY.
- Wong, D., Chan, V., McCormack, A., 2009. Functional Cloning and Expression of a Novel Endo- α -1,5-L-Arabinanase from a Metagenomic Library. *Protein Pept. Lett.* 16, 1435–1441. doi:10.2174/092986609789839313
- Wong, D.D.W.S., Chan, V.J., McCormack, A.A., Batt, S.B., 2010. A novel xyloglucan-specific endo- β -1,4-glucanase: biochemical properties and inhibition studies. *Appl. Microbiol. Biotechnol.* 86, 1463–1471. doi:10.1007/s00253-009-2364-2
- Wong, D.W.S., Chan, V.J., Batt, S.B., 2008. Cloning and characterization of a novel exo- α -1,5-L-arabinanase gene and the enzyme. *Appl. Microbiol. Biotechnol.* 79, 941–949. doi:10.1007/s00253-008-1504-4
- Wong, D.W.S., Chan, V.J., Liao, H., Zidwick, M.J., 2013. Cloning of a novel feruloyl esterase gene from rumen microbial metagenome and enzyme characterization in synergism with endoxylanases. *J. Ind. Microbiol. Biotechnol.* 40, 287–295. doi:10.1007/s10295-013-1234-1

- Wong, D.W.S., J. Chan, V., A. McCormack, A., B. Batt, S., 2010. Cloning and Characterization of an Exo-Xyloglucanase from Rumenal Microbial Metagenome. *Protein Pept. Lett.* 17, 803–808. doi:10.2174/092986610791190381
- Wu, S., Baldwin, R.L., Li, W., Li, C., Connor, E.E., Li, R.W., 2012. The Bacterial Community Composition of the Bovine Rumen Detected Using Pyrosequencing of 16S rRNA Genes. *Metagenomics* 1, 1–11. doi:10.4303/mg/235571
- Xia, Y., Kong, Y., Seviour, R., Yang, H.-E., Forster, R., Vasanthan, T., McAllister, T., 2015. *In situ* identification and quantification of starch-hydrolyzing bacteria attached to barley and corn grain in the rumen of cows fed barley-based diets. *FEMS Microbiol. Ecol.* 91, fiv077. doi:10.1093/femsec/fiv077
- Yaniv, O., Fichman, G., Borovok, I., Shoham, Y., Bayer, E.A., Lamed, R., Shimon, L.J.W., Frolow, F., 2014. Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti- σ^1 factors. *Acta Crystallogr. D Biol. Crystallogr.* 70, 522–534. doi:10.1107/S139900471302926X
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y., 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi:10.1093/nar/gks479
- Yuan, P., Meng, K., Wang, Y., Luo, H., Huang, H., Shi, P., Bai, Y., Yang, P., Yao, B., 2012. Abundance and Genetic Diversity of Microbial Polygalacturonase and Pectate Lyase in the Sheep Rumen Ecosystem. *PLoS ONE* 7, e40940. doi:10.1371/journal.pone.0040940
- Ze, X., Ben David, Y., Laverde-Gomez, J.A., Dassa, B., Sheridan, P.O., Duncan, S.H., Louis, P., Henrissat, B., Juge, N., Koropatkin, N.M., Bayer, E.A., Flint, H.J., 2015. Unique Organization of Extracellular Amylases into Amylosomes in the Resistant Starch-Utilizing Human Colonic *Firmicutes* Bacterium *Ruminococcus bromii*. *mBio* 6, e01058–15. doi:10.1128/mBio.01058-15
- Zhao, S., Wang, J., Bu, D., Liu, K., Zhu, Y., Dong, Z., Yu, Z., 2010. Novel Glycoside Hydrolases Identified by Screening a Chinese Holstein Dairy Cow Rumen-Derived Metagenome Library. *Appl. Environ. Microbiol.* 76, 6701–6705. doi:10.1128/AEM.00361-10
- Zhao, S., Wang, J., Liu, K., Zhu, Y., Bu, D., Li, D., Yu, P., 2009. [Screening and characterization of lipase from a metagenome library of dairy rumen microflora]. *Sheng Wu Gong Cheng Xue Bao Chin. J. Biotechnol.* 25, 869–874.
- Zhao, S., Zhao, J., Bu, D., Sun, P., Wang, J., Dong, Z., 2014. Metabolomics analysis reveals large effect of roughage types on rumen microbial metabolic profile in dairy cows. *Lett. Appl. Microbiol.* 59, 79–85. doi:10.1111/lam.12247
- Zhou, J., Bao, L., Chang, L., Liu, Z., You, C., Lu, H., 2012a. Beta-xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans: Characterization of a metagenomic GH3 β -glucosidase/xylosidase. *Lett. Appl. Microbiol.* 54, 79–87. doi:10.1111/j.1472-765X.2011.03175.x
- Zhou, J., Bao, L., Chang, L., Zhou, Y., Lu, H., 2012b. Biochemical and kinetic characterization of GH43 β -d-xylosidase/ α -l-arabinofuranosidase and GH30 α -l-arabinofuranosidase/ β -d-xylosidase from rumen metagenome. *J. Ind. Microbiol. Biotechnol.* 39, 143–152. doi:10.1007/s10295-011-1009-5
- Zhou, M., Chen, Y., Guan, L.L., 2015. Rumen Bacteria, in: *Rumen Microbiology: From Evolution to Revolution*. Springer India, pp. 79–95.

Thesis goals

As introduced in the previous chapters, the bovine rumen is a highly diverse ecosystem, able to degrade a large number of complex substrates, which has not been sufficiently explored up to now. Until now, the search for new enzymes was indeed mostly focused on the discovery of plant cell wall degrading enzymes, and more specifically glycoside-hydrolases. Nevertheless, if sequence-based and activity-based meta-omic approaches were indeed quite often used for hemicellulase and cellulase discovery, no novel CAZy family was discovered up to now from the ruminal microbiome, and the potential of the most original glycoside-hydrolases for crude plant cell wall breakdown was never investigated, because of the lack of high-throughput functional assays. Moreover, lignocelluloses, which constitute one of the main source of carbon for ruminal microbes, contain, in addition to plant cell wall polysaccharides, between 10 and 30 % of polyaromatic compounds, in the form of lignin, of which the metabolization by ruminal microbes has been only rarely described. In particular, the role of ruminal bacteria in lignin degradation has, to our knowledge, never been investigated. More generally, the discovery of the potential of lignin degradation by bacteria is very recent, and is mostly documented for aerobic ecosystems. Besides, esterases are known to play a critical role in metabolization of both fatty acids and esterified chains of lignocelluloses. Nevertheless, only few of them were characterized. Finally, while CAZymes are, of course, still enzymes of interest for biorefineries, other kind of enzymes (like bacterial oxidoreductases and esterases) are also of high interest both for green chemistry and for bioremediation.

In this context, the main goal of my thesis was to explore the functional diversity of the ruminal bacteriome, in order to find new enzymes able to degrade complex polymer structures, natural or synthetic ones.

Moreover, while some sequence-based analyses can be found regarding the effect of diet on the functional profile of the rumen microbiome, activity-based metagenomics was never applied on libraries constructed after two kinds of enrichments of the same ruminal microbiome. The comparison of the effects of two different enrichment methods (*in vivo*, by using diet shift, and *in vitro*, in bioreactors) was thus also part of the thesis. This study was performed in collaboration with Diego Morgavi (INRA Theix) for rumen microbiome sampling, Guillermina Hernandez-Raquet and Adèle Lazuka (LISBP) to conduct bioreactor enrichment, and Patrick Robe (Libragen SA) for metagenomic library construction.

At the beginning of this project, several questions arose, to which we were committed to answer during this thesis, along with several technological challenges that had to be taken up:

- Could the functional diversity of the bovine ruminal microbiome be exploited to find novel enzymes for pollutant bioremediation?

- How to rapidly investigate the substrate specificity of enzymes discovered by activity-based metagenomics towards pollutants of various structures, in order to rapidly evaluate their applicative potential?
- What could be the physiological role of these new enzymes in the rumen?
- Would there be major differences in the functional profile and taxonomic diversity between the consortia obtained after *in vivo* and *in vitro* enrichments on crude lignocellulosic biomass?
- How to boost the discovery of lignocellulases to alleviate the biomass pre-treatment bottlenecks, especially i) oxidoreductases capable of degrading lignin, and ii) enzymes acting synergistically to break down crude plant cell wall extracts?

The workflow of the general strategy and organization of the work conducted during this thesis is given below:

Initial diet of the cows: Standard dairy cow ration of corn silage, hay and concentrate		
Initial sampling of the rumen content Mix of ruminal samples from 2 animals		
<i>In vivo</i> enrichment (IVVE): Same 2 cows fed predominantly on wheat straw for seven weeks	<i>In vitro</i> enrichment (IVTE): Rumen sample (mix from 2 animals) with wheat straw as sole carbon source in a bioreactor (10 cycles of 7 days)	
DNA extraction for metagenomic library creation		
IVVE library	IVTE library	
<i>Primary screening</i>		
Esterases/lipases: Tween 20	Oxidoreductases: Lignin alkali	Carbohydrate Active Enzymes (CAZymes): Chromogenic polysaccharides
<i>Secondary screening</i>		
Esterases/lipases: <i>p</i> NP-acetate, <i>p</i> NP-butyrate, <i>p</i> NP-palmitate	Oxidoreductases: ABTS	CAZymes: 6 chromogenic polysaccharides of related structures (chemically modified celluloses, ramified or linear hemicelluloses)
<i>Characterization of substrate specificity and activity quantification</i>		
<i>IVVE library</i>		<i>IVVE and IVTE libraries</i>
Esterases/lipases: carbamate polymers and monomers	Oxidoreductases: Synthetic dyes, lignin derivatives	Wheat straw, Design of enzymatic cocktails
Sequencing, functional and taxonomical annotations	Sequencing, functional and taxonomical annotations	Sequencing, functional and taxonomical annotations
		Library comparison
New enzymes for biorefineries and bioremediation		

Chapter IV

Chapter V

Chapter VI & Annex I (Ufarté et al., Springer methods, 2016)

The work carried out for three years to answer these questions is presented in the three following chapters, and in the methodological paper published in Springer methods (2016, annex I).

Chapter IV:

Discovery of carbamate degrading enzymes by functional metagenomics

Ufarté L.^{1,2,3}, Laville E.^{1,2,3}, Duquesne S.^{1,2,3}, Morgavi D.⁴, Robe P.⁵, Klopp C.⁶, Rizzo A.^{1,2,3}, Bozonnet S.^{1,2,3} and Potocki-Veronese G.^{1,2,3}

¹Université de Toulouse, INSA, UPS, INP, LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

²INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France

³CNRS, UMR5504, F-31400 Toulouse, France

⁴INRA, UMR 1213 Herbivores, 63122 Saint-Genès Champanelle, France

⁵LibraGen S.A., F-31400 Toulouse, France

⁶Plateforme Bio-informatique Toulouse Genopole, UBIA INRA, BP 52627, F-31326 Castanet-Tolosan Cedex, France

To be submitted (2016)

Functional metagenomics is a powerful tool to explore the functional diversity of microbial ecosystems. As described in Chapter 1, the activity-based screening approach is the only one which enables the discovery of new enzyme families and new functions that would not be predictable by sequence analysis. There is indeed a drastic lack of biochemical data in sequence databases, functional characterization of enzymes requiring much more efforts than (meta)genome sequencing. While many kinds of ecosystems have been mined up to now (deep seas, human gut, insect gut, soil...) for novel enzymes through sequence or activity-based metagenomics, it was the cow rumen ecosystem and its possibilities that appealed to us.

The bovine ruminal metagenome has been mined mostly for glycoside-hydrolases involved in plant cell wall degradation. However, considering its high functional and taxonomical diversity, the potential of such an ecosystem has a much broader impact. A major concern that is found all around the world is the bioremediation of pollutants, such as plastics and phytosanitary compounds. Many of them, in particular carbamates, can be degraded by esterases. We decided to mine the ruminal bacteriome for such activities, in order to isolate a battery of novel esterases of biotechnological interest, with physiologic functions that can be diverted to break down unnatural substrates, like carbamates, thanks to their flexible substrate specificity. This work, which enabled the discovery of novel esterases acting on a major pollutant of coating industries (polyurethane) and on a widely used insecticide (fenobucarb), is described in the fourth chapter. This paper will be submitted in the next few weeks.

In addition of the team members, this part of the work was performed in collaboration with D. Morgavi (INRA Theix, rumen content sampling), P. Robe (Libragen SA, library construction), and C. Klopp (INRA Toulouse, read assembly).

Abstract:

Bioremediation of pollutants is a major concern worldwide, leading to the research of new processes to break down and recycle xenobiotics and environmental contaminating polymers. Among them, carbamates have a very broad spectrum of uses: insecticides, herbicides, elastomers... In this study, we mined the bovine rumen microbiome for enzymes capable of degrading different structures of carbamates. A fosmid *E. coli* metagenomic library, corresponding to 700 Mbp of bacterial DNA, was screened for esterase and protease activities that are known to be responsible for carbamate hydrolysis. Secondary screening was performed on Impranil DLN, a commercial polyurethane, and insecticide carbamates (fenobucarb, fenoxycarb, prosulfocarb), enabling the isolation of 57 clones of interest, exhibiting esterase/lipase activity, and able to degrade at least one of these compounds. Two clones were more deeply characterized. They indeed harbor very original carboxyl-ester hydrolase sequences, named CE1_Ubrb and CE2_Ubrb, belonging respectively to the lipolytic families V and IV. Both clones are able to grow using Impranil DLN as sole carbon source, and to affect polymer chain organization and/or structure. CE2_Ubrb, in particular, would be a highly polyspecific esterase of high interest for carbamate bioremediation. The metagenomic clone producing it is indeed active, in addition to Impranil, on Tween 20, *p*NP-acetate, butyrate and palmitate, and on the insecticide fenobucarb. This study highlights the potential of highly diverse microbiota such as the ruminal one to discover promiscuous enzymes, whose substrate flexibility could be exploited for industrial uses.

1. Introduction:

Carbamates are *N*-substituted esters of carbamic acid. They are organic compounds with general formula $R_1-NR_2-(C=O)O-R_3$, used in agriculture as insecticides, fungicides or herbicides (where R_1 is a methyl, benzimidazole and aromatic moiety, respectively). The synthesis and commercialisation of carbamate pesticides have been in progress since the 1950's. Nowadays, the most employed insecticides are organophosphorus compounds, carbamates and pyrethroids (Sogorb and Vilanova, 2002).

The carbamate pesticides have been in use in different kinds of crops all over the world. Their uses increased along with organophosphorus, to replace organochlorine pesticides. However, some carbamate compounds and derivatives have an acute toxicity for mammals and aquatic organisms. Carbamates, just like the well-known toxic organophosphorus compounds, are inhibitors of acetylcholinesterase and therefore cause very similar symptoms. Some of them, containing aromatic moieties, are also suspected carcinogens and mutagens since they break down to aniline based derivatives, which were shown to be carcinogenic for humans (Nunes and Barceló, 1999; Wang and Lemley, 2003).

For example, methyl carbamates are considered non genotoxic because of the inability of the methyl group to undergo metabolic degradation to an epoxide, while ethyl carbamates are considered as multispecies carcinogens causing malignancies in different tissues (lung, hematopoietic system, liver...) (Bemis et al., 2015). As such, environmental cleanup and wastewater treatment became a high priority, more than 10,000 tons of carbamates being used per year worldwide in the agriculture sector (data obtained from the Food and Agriculture Organization of the United Nations, <http://faostat3.fao.org/> and from (Grube et al., 2011)).

The elimination of many lipophilic environmental pollutants is based on their conversion to water soluble compounds and further microbial metabolisation. The first step in the metabolic degradation of pesticide/herbicide carbamates is hydrolysis (Sogorb and Vilanova, 2002; Smith and Bucher, 2012) catalysed by carboxyl ester hydrolases (EC 3.1.1). These enzymes are part of the wide group of ester hydrolases (EC 3.1) catalysing the hydrolysis of carboxylesters, thioesters (EC 3.1.2), phosphoric (EC 3.1.4/5/7/8) and sulfuric esters (EC 3.1.6). Carboxylesterases from various bacteria were reported to hydrolyse carbamates: *Blastobacter* (Hayatsu and Nagata, 1993), *Arthrobacter* (Pohlentz et al., 1992), *Pseudomonas* (Mulbry and Eaton, 1991), *Achromobacter* (Karns and Tomasek, 1991) and *Micrococcus* genera (Doddamani and Ninnekar, 2001). Some of them display important homologies with eukaryotic carboxylesterases also reported to degrade carbamates (Pohlentz et al., 1992; Russell et al., 2011).

Besides phytosanitary products, another type of carbamates, namely polyurethanes (PUs), causes huge environmental problems. These polymers are found in many industrial products: coatings, furniture, paints, synthetic skin, constructional materials, elastomers, and even implantable biomedical devices ... (Rahimi and Mashak, 2013; Biffinger et al., 2015). Step by step, they have replaced other polymers for many reasons: chlorinated rubber because of safety reasons; latex rubber because of the lower density and higher flexibility of PUs; plastics (such as PVC for automotive application for example) because of their resistance to water, oils and solvents... (Howard, 2002). PUs accounts for 6 to 7% of the total mass of plastic generated worldwide, with over half of the total global market for coating applications (Biffinger et al., 2015). PU worldwide production thus represents 15 Mtonnes per year. The major use of PUs in Europe is packaging, and while recycling has increased these last years, the first way to take care of the waste for most countries is still to use landfilling (PlasticsEurope, 2015). In particular, Impranil®DLN, the specific PU used in the current study, is a protected anionic aliphatic polyester-PU produced by Bayer Corporation for use in textile, leather and aircraft fabric coating (Biffinger et al., 2015).

PU biodegradation results from the activity of two main classes of enzymes (Howard, 2002; Loredó-Treviño et al., 2012): esterases (such as cholesterol esterases) and proteases (such as papain) that were shown to hydrolyse carbaryl and 4-nitrophenyl *N*-methyl-

carbamates (Kuhr and Dorough, 1976; Howard and Blake, 1998; Cregut et al., 2013). Other kinds of enzymes such as ureases (Phua et al., 1987; Nakajima-Kambe et al., 1999) along with enzymes named imbranilases, were shown to degrade PUs. Moreover, if their enzymes have not all been identified to date, PU bacterial degraders are numerous, mostly from the phyla Firmicutes, Actinobacteria, and Proteobacteria (Cregut et al., 2013). It was shown, however, that PUs are barely used as carbon or nitrogen sources for microorganisms in strict anaerobic conditions (Urgun-Demirtas et al., 2007). On the other hand, in aerobic conditions, only one bacterial strain, assigned to the *Comamonas* genus, has been shown to use polyester PU as carbon and nitrogen source (Nakajima-Kambe et al., 1995).

Despite these advances in the understanding of carbamate enzymatic degradation, there are still very few biological routes identified for the breakdown and recycling of carbamate-based xenobiotics and polymers. In particular, the carbamate degrading potential of the uncultivated bacterial fraction, which represents between 70 and 99 % of most microbial ecosystems, is underexplored. Only one study indeed allowed the discovery of a polyurethane (poly(diethylene glycol adipate)) degrading esterase from an uncultivated soil bacteria (Kang et al., 2011), while, to our knowledge, no metagenome was mined for activities of cleavage of the carbamate linkage of insecticides or pesticides (Ufarté et al., 2015).

The bovine rumen metagenome was chosen for this study because of its huge functional and microbial diversity (Wang et al., 2013), and because of its richness in esterases, including carbohydrate-esterases (Brulc et al., 2009; Wang et al., 2013) and lipases (Liu et al., 2009; Privé et al., 2015). In this dense ecosystem, these enzymes are involved both in metabolisation of lipids and of plant cell wall components, since carbohydrate-esterases catalyze the de-O or -N-acylation of substituted hemicelluloses and pectins (Faulds, 2010).

Here we report the discovery, by using a multi-step activity-based metagenomics strategy targeting the bovine rumen microbiome, of new esterases capable of hydrolysing two different structures of carbamates, more specifically Impranil and fenobucarb.

2. Results and discussion:

2.1. Screening of the bovine rumen metagenomic library

The metagenomic library was obtained from the rumen of two cows fed with wheat straw enriched food. It consisted of 19,968 *Escherichia coli* fosmid clones, each clone comprising a 30-40 kb DNA insert. The library covers around 700 Mbp of metagenomic DNA, which corresponds to the equivalent of more than one hundred bacterial genomes.

The library was first screened for the presence of chosen enzymatic activities: both esterases and proteases were searched for, since they were previously reported to cleave the carbamate linkage. The clone ability to degrade Tween 20 (polyethylene glycol sorbitan monolaurate, an ester with C12 chain length), would reveal the presence of esterases and lipases (Kulkarni et al., 2013), also renamed non lipolytic esterases and lipolytic esterases, respectively, according to their classification by Ben Ali et al. (Ben Ali et al., 2012). AZCL-casein was also used as primary screening substrate to search for proteases.

The screen on Tween 20 resulted in 26 positive clones, corresponding to a hit yield of 0.13%. This value is in the same range of those obtained from other activity-based metagenomics studies targeted on the bovine rumen microbiome. Indeed, from the rumen metagenome of New Zealand dairy cows fed on a forage-based diet, Ferrer et al. (2005) found 11 esterases (hit yield 0.08%) by using α -naphthyl acetate, a very small substrate, as the screening substrate. In another study, (Privé et al., 2015) found 9 positive clones (hit yield 0.04%) on tributyrin, a bigger substrate than α -naphthyl acetate, yet smaller than Tween 20, from the rumen microbiome of cows from Wales, fed with a mixture of grass silage, straw and sugar beet nuts. The high hit yield obtained in the present study indicates that Tween20 is an appropriate substrate to explore the functional diversity of bacterial carboxylesterases.

Another primary screen was used to explore the functional diversity of esterases from the ruminal microbiota, in order to specifically find those that, *in vivo*, are involved in plant cell wall breakdown. We based this screening approach on the fact that lignocellulose deconstruction requires a very diverse battery of enzymes, including carbohydrate-esterases, working in synergy to face the plant cell wall structural complexity. Many genomic (Terrapon and Henrissat, 2014) and metagenomic (Tasse et al., 2010) studies revealed that in bacteria, these enzymatic cocktails are encoded by multigenic clusters that can easily be fished in 30-40 kbp DNA fragments. We thus screened our library for plant cell wall polysaccharide (cellulose, barley β -glucan, xylan, carob galactomannan...) degradation activities. In total, 62 polysaccharide degrading clones were obtained (0.31% yield).

The screen on AZCL-casein did not allow the detection of positive clone, contrarily to what we previously observed in the same screening conditions with other metagenomic libraries issued from various composts (unpublished results). We cannot exclude that this could be due i) to differences in taxonomical diversity between compost and rumen samples, and thus to differences in the heterologous expression bias imposed by the activity-based metagenomic approach, and/or ii) to differences in specific requirements to screen for ruminal proteases, such as metalloproteases. Nevertheless, our result reflects the paucity of the rumen microbiome in proteases. Indeed, no activity-based metagenomic study having resulted in finding protease activity in the bovine rumen has been published. In addition, the sequence-based metagenomics studies targeting this ecosystem have been mostly focused on carbohydrate active enzymes (Brulc et al., 2009; Hess et al., 2011; Wang et al., 2013) and

do not give an estimation of the content in proteolytic enzymes. The sole knowledge of the proteolytic activities in the rumen come from activity assays on microbial consortia. These studies showed that protease activity is much lower than polysaccharide degrading activities (Rey et al., 2012), part of the initial protease activity being suspected to come from the plant residues present in the rumen (Zhu et al., 1999).

A secondary screening of the 85 clone hits was performed on two different kinds of substrate in parallel: Impranil, a commercial polyester polyurethane, and *p*NP chromogenic esters of various chain lengths (C2, C4 and C16) to discriminate and quantify lipolytic versus non lipolytic carboxylesterase activities.

Impranil degradation was screened by using both positive selection and by searching for clones, grown on rich medium, showing a polymer degradation halo (Howard and Blake, 1998). Among the 85 clones obtained in total after primary screening, three being positive on both Tween 20 and polysaccharides, 47 positive clones were obtained on mineral medium containing Impranil as sole carbon source, and 55 on the rich medium supplemented with polyurethane, for a total of 57 different clones. Forty five clones were positive on both Impranil-based assays (0.23% yield). They corresponded to those harbouring the largest halo, witness of high recombinant gene expression and/or enzyme efficiency and/or even secretion which may happen, even rarely, in *E. coli*. Thirteen and 35 out of the 45 of them were issued from primary screens on Tween 20 and polysaccharides, respectively. Furthermore, of the 57 clones active on both Impranil-based assays, 28 were active on *p*NPA, 22 on *p*NPB and 4 on *p*NPP, the latter substrate being specific to lipolytic carboxylesterases (Kim et al., 2006) (Figure 17).

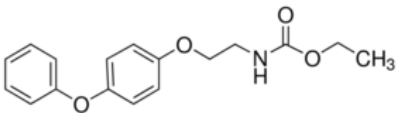
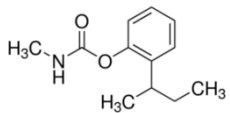
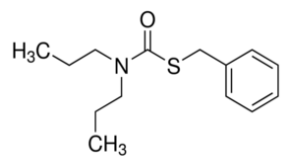
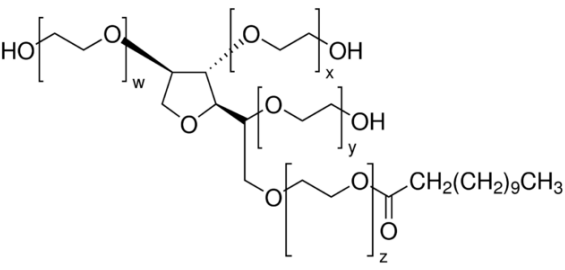
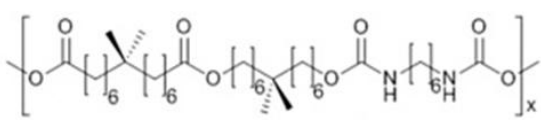


Figure 17: Functional profiles of the hit metagenomic clones, assayed on polysaccharides (dark blue), Tween 20 (green), Impranil (brown: minimal medium; orange: rich medium), *p*NPA (purple), *p*NPB (red), *p*NPP (bright blue). The clones mentioned in brackets are sequenced ones.

These results highlight the diversity of substrate specificity of ruminal bacterial esterases, but also the interest of screening fosmid libraries for plant cell wall degrading activities in order to find these esterases. Discrimination screening results allowed for clusterisation of the hit clones, based on their functional diversity (Figure 17). From the 45 clones highly active on Impranil, 21 were chosen for further sequencing in order to identify the novel enzymes active on carbamates. These hits were chosen belonging to various functional clusters, in order to explore the largest sequence and functional diversity as possible. Among these 21 clones, 2 clones, issued from each primary screen (Tween 20 / polysaccharides), which were among the most efficient to break down Impranil on both screening media (as they produced the largest degradation halos) and which harboured quantifiable activity on 3 and 1 of the tested *p*NP-fatty acids, respectively, were selected for in depth characterization of their activity on carbamates of various structures (Table 7). Clone 44I12 is indeed active on the three tested *p*NP-substrates, and also on Impranil with both assays. Clone 12F23 is weakly active on *p*NPB (Table 8), along with being able to degrade Impranil with both assays. This last characteristic is especially important since it shows the ability of the clone to use Impranil as a carbon source, since recent studies have

spread some doubts on the specificity of the screening methods based on the observation of a degradation halo on solid media for the discovery of Impranil degrading enzymes and strains (Biffinger et al., 2015).

Table 7: Structures of Tween 20 and of the carbamates used in this study

Fenoxycarb	Fenobucarb	Prosulfocarb
		
Tween 20		Impranil (deduced structure (Biffinger et al., 2015))
		

2.2. Polyurethane degradation

In order to investigate deeper the action of clones 12F23 and 44I12 onto polyurethanes, concentrated cytoplasmic extracts of the clones were incubated in the presence of Impranil. After two days of reaction at 30°C and pH 7.0, the 12F23 reaction medium became limpid, and a small aggregate could be seen at the bottom of the tube. The opposite was observed for 44I12: the medium became more compact and more viscous (Figure 18). We compared these observations with those of Biffinger et al. (2015), who used various commercial enzymes to degrade Impranil: an esterase from *Pseudomonas fluorescens* with a broad substrate specificity (from lysophospholipase to amidase activity), Savinase™, a commercial serine-type protease originally produced by *Bacillus* sp., and a triacylglycerol lipase from *Pseudomonas* sp.. They observed that 2 days incubation at 30°C of Impranil with esterase and lipase resulted in a decrease of the reaction medium opacity without aggregate formation, while Impranil aggregated when incubated with protease. Here, we checked that none of the ORFs of 12F23 was annotated as protease, using the RAST annotation server. In addition, no protease activity, which could have been due to an ORF annotated as hypothetical protein, was observed when tested on Hide Remazol Blue, AZO- and AZCL-casein (data not shown).

From these results, we deduce that the Impranil structural modifications induced by clones 12F23 and 44I12 are different, due to differences in substrate specificity of the enzymes encoded by the two metagenomic inserts.


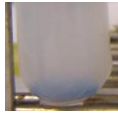




Clones	Before reaction	After reaction
negative control (empty fosmid)		
12F23		
44I12		

Figure 18: Reaction media of 12F23 and 44I12 cytoplasmic extracts incubated with Impranil at 30°C during 48 h. For 44I12 after 48 h reaction, the reaction medium, in and out of the tube, is showed, as a solid aggregate could be obtained.

2.2.1. HPSEC analyses

In order to check that Impranil was depolymerized, 44I12 and 12F23 reaction media were analysed by HPSEC, after lyophilisation of the whole media and dissolution in THF. Dilution in PBS buffer, lyophilisation and dissolution in pure THF strongly impacted polymer chain organization, as a precipitated fraction appeared in the samples, both with the clones and the Impranil controls (Figure S1). This phenomenon did not occur when the Impranil standard solution was only diluted twice in THF before injection. As such, only the soluble fraction of the reaction media could be analysed. The analyses are thus more qualitative than quantitative. Nevertheless, we noticed that after reaction, the HPSEC recovery yields decreased by 58, 37 and 12% respectively for clones 12F23, 44I12 and negative control, compared to the samples at initial reaction times (Figure 19). In addition, the molecular mass profiles were significantly modified, especially for clone 12F23, for which the peak eluted after around 13 min and corresponding to high molecular masses, disappeared, while a third peak eluted around 17.5 min on the profile corresponding to smaller molecular masses. Biffinger et al. (2015) also used HPSEC to analyse degradation products of Impranil by commercial enzymes. With the three tested enzymes, some large oligomers were still observable on the chromatograms, along with a change of profile in the degradation products for the reaction with the esterase from *Pseudomonas fluorescens*, just as the one we observed here.

Even if we cannot conclude to Impranil depolymerisation by clones 12F23 and 44I12, since, because of the insolubility of the aggregates produced during reaction we did not analyse all reaction products, these results confirm that the metagenomic clones drastically

impact the polymer chain structure and/or organisation, which explains reaction media clarification and gelification, respectively.

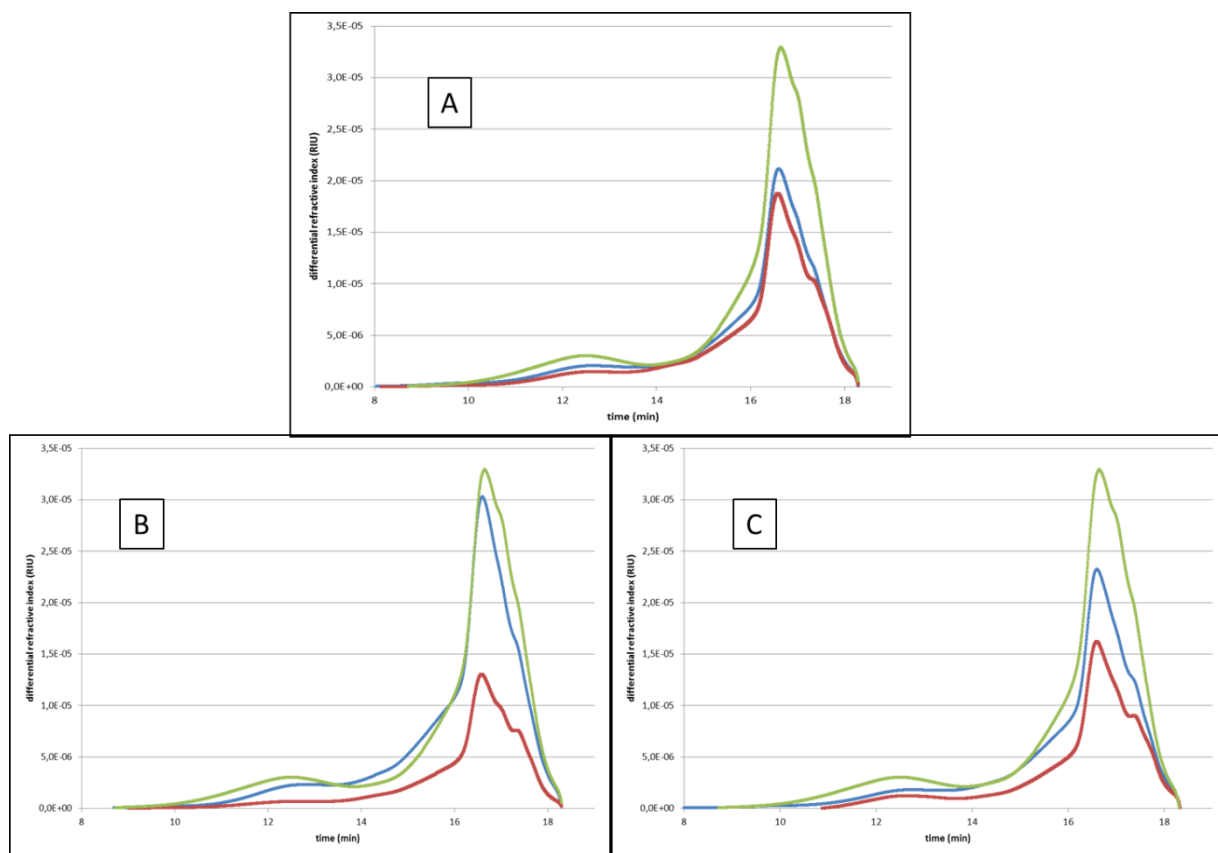


Figure 19: HPSEC chromatogram of the reaction between Impranil and the clones of interest. The green curve corresponds to Impranil DLN standard, 3 g/L. The blue curves are the T0, the red T24 of **A.** negative control; **B.** clone 12F23; **C.** clone 44I12

2.2.2. MALDI-TOF analyses

In order to analyse the molecular mass distribution of the Impranil substrate and reaction products, MALDI-TOF analyses were performed on reaction media before and after reaction. Only the spectra obtained with clone 44I12 presented a significant difference before and after reaction. A different molar mass distribution was observed, with the apparition of peaks at m/z 682.4, 683.4, 763.4 and 782.4 after reaction (Figure 20). Unfortunately, there is no known structure of Impranil (Howard and Blake, 1998), even if a hypothetical one has been proposed by Biffinger et al. (2015). According to this hypothetical structure, the predominant new peaks could correspond to the following degradation products: the one at m/z 682.4 is the closest in size to $C_{36}H_{68}O_8N_2$ with an atom of Na; the one at 782 could be the motif $C_{36}H_{66}O_8N_2$ with an atom of I but nothing seems to match with only additions of Na^+ or I^- . Other peaks, present for both reaction times, may correspond to the polymerized Impranil motives with low polymerization degrees. These results constitute a definite proof, along with primary screening and HPSEC analysis, that the enzymes produced by clones 12F23 and 44I12 impact the structure of the commercial polyurethane Impranil.

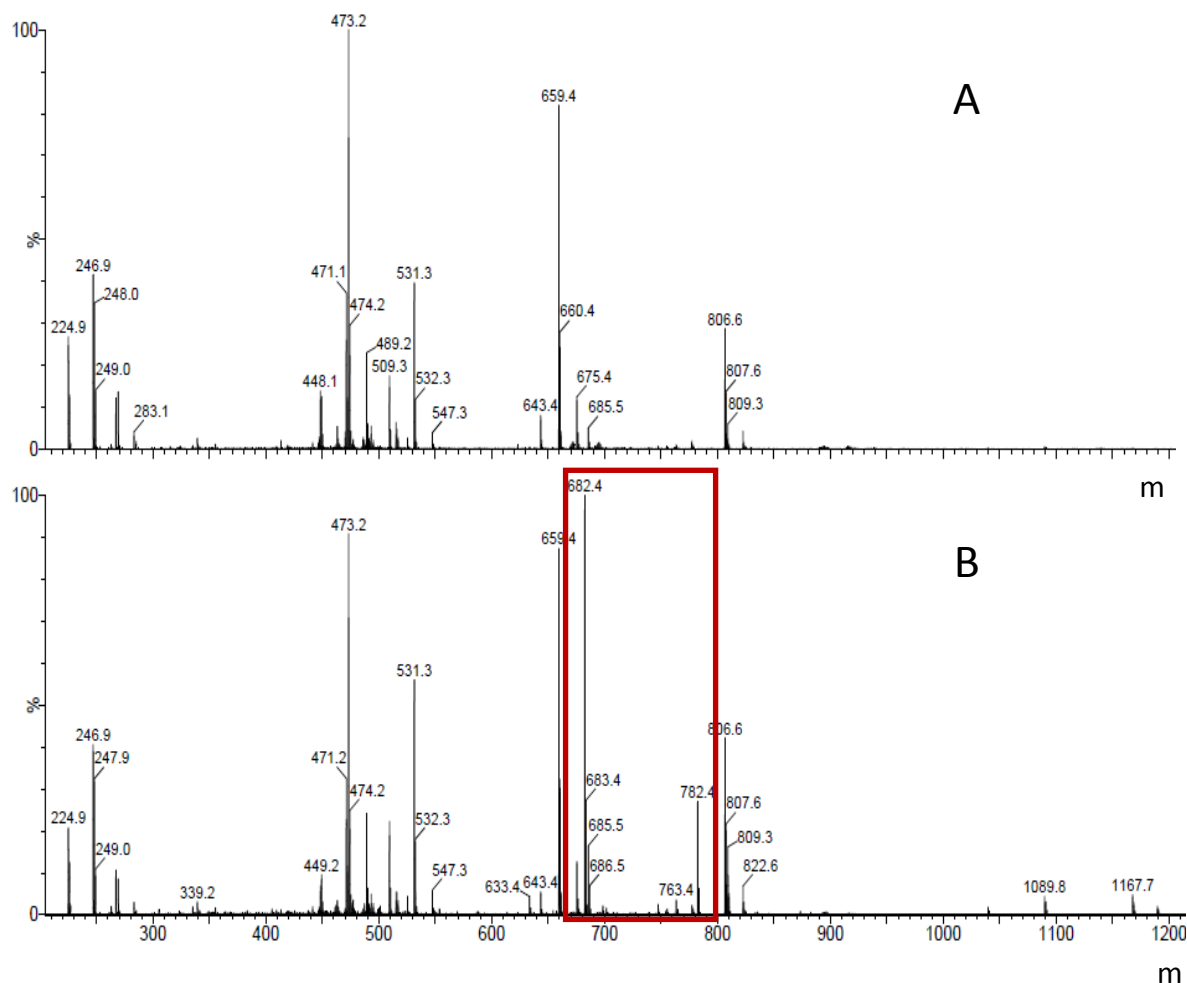


Figure 20: MALDI-TOF spectra of reaction media (A) before and (B) after 24h reaction with clone 44I12. The red rectangle shows the new peaks observed after reaction

2.3. Insecticide/herbicide carbamate degradation

Clones 12F23 and 44I12 were then tested on smaller carbamate pollutants of known structures (Table 7).

An example of carbamate insecticide, mentioned in the current study, is fenobucarb, a synthetic insecticide with a high potential for bio-concentration, acting as an acetylcholinesterase inhibitor. Fenoxycarb is also a synthetic insecticide, acting by mimicking a juvenile hormone from the insects to keep them in an immature state, of which the bioaccumulation potential is moderate. Prosulfocarb, however, is a synthetic thiocarbamate herbicide which acts as a lipid synthesis inhibitor. It is very persistent in water and sediments (University of Hertfordshire (2013)). While fenoxycarb and prosulfocarb have been and still

are authorised in many European countries, including France, it is not the case for fenobucarb. The latter is used mainly in paddy fields, in order to control rice thrips, leafhoppers and plant hoppers, in many rice growing countries worldwide. However, the biodegradation of fenobucarb is not well known. Indeed, to our knowledge, isolation of fenobucarb-degrading bacteria is described in only one recent study (Kim et al., 2014).

The cytoplasmic extracts of clones 12F23 and 44I12 were incubated during three days at 30°C with these 3 compounds independently, and reaction media were analysed by LC-MS. Nothing significant was observed for the reactions on fenoxycarb and prosulfocarb, while the UV profile of the reaction medium with fenobucarb was significantly changed after reaction with clone 44I12 (Figure 21). The fenobucarb peak at retention time 6.4 min decreased by 33 ± 4 % after reaction, to the profit of a second peak at retention time 7.7 minutes. This peak, which is present in very little amount on the commercial fenobucarb UV profile, corresponds to the degradation product 2-sec-butylphenol (Kim et al., 2014), which was checked with a commercial standard. Moreover, mass spectrum analysis allowed us to confirm that peak at 6.4 min corresponds to fenobucarb (m/z 207) while peak at 7.7 min presents MS peaks at masses (m/z 150 and 121) that correspond to 2-sec-butylphenol. Reference masses were taken from the work of Wang and Lemley (2003), who studied fenobucarb degradation using membrane anodic Fenton treatment. These results demonstrate that clone 44I12 is able to degrade fenobucarb by cutting the ester link, releasing a product (2-sec-butylphenol) which is less toxic than fenobucarb itself (U.S. National Library of Medicine, 2014). Nevertheless, clone 44I12 is unable to degrade fenoxycarb and prosulfocarb. This trend was already observed by Kim et al. (2014), who discovered bacterial strains from the genera *Sphingobium* and *Novosphingobium*, which are able to degrade fenobucarb and carbaryl, but not other carbamates like fenoxycarb.

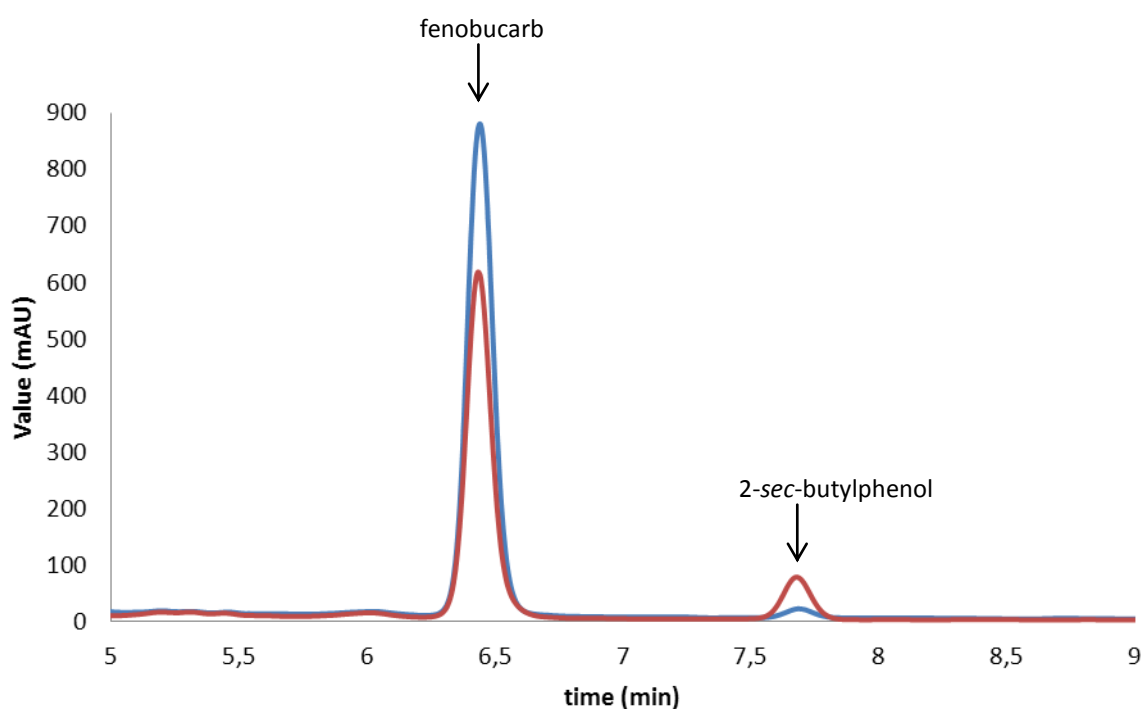


Figure 21: HPLC analysis from the reaction of fenobucarb with 44I12 enzymatic extract at the beginning of the reaction (blue) and after 90 hours reaction (red)

2.4. Sequence analysis

The metagenomic DNA inserts of the 21 hit clones chosen as previously described in this paper were sequenced. For each clone, a random selection of reads was performed to obtain a mean depth coverage of 40 x. For each clone, the largest contig, sizing between 21 and 26 kbp, was also chosen for ORF detection and annotation. They contain 14 to 31 predicted ORFs, depending on the clone.

The contig taxonomic annotation, based on ORF Megan analysis, is presented in table 2. All sequences came from bacteria of which the genome has not been sequenced yet. No contig was assignable to a particular phylum, since all of them contain ORFs which are not all assigned to the same phylum. This highlights the originality of these sequences, which are all issued from bacteria i) that are probably very distant from those of which the genome has already been sequenced, and/or ii) of which the genome evolved *via* the frequent horizontal gene transfers occurring between mammal gut bacteria (Tasse et al., 2010).

All ORFs were assigned to bacteria, except one in contigs 49M8 and 37F15, which were misassigned to an eukaryotic organism, and ten in contig 33M21 which were unassignable. Many ORFs were assigned to bacteria from environmental samples without more taxonomic information. This is explained by the strong enrichment of databases through massive metagenome sequencing projects of the last decade and by the lack of knowledge on organisms that colonize these ecosystems. When ORFs were assigned to bacterial organisms, it was only at the phylum level and for the great majority to Firmicutes, which are known to be the major phylum in the bovine rumen microbiota (Peng et al., 2015). Few ORFs were however assigned to Proteobacteria (clones 45P14, 42M18, 33M21, 32D12, 21K7), Actinobacteria (33M21), and Spirochaete (33M21). Regarding the two clones that caught our attention for further characterization, clone 44I12 had a typical Firmicute profile, since 20 out of the 27 ORFs were assigned to this phylum, 10 being further assigned to the Clostridia class and one to the *Thermobacillus composti* species (Watanabe et al., 2007). Conversely, sequences from clone 12F23 are highly distant to those of sequenced cultured bacteria and only matches to sequences retrieved from bacterial ecosystems, even if three ORFs were assigned to the Clostridia class of the Firmicute phylum.

ORFs were then functionally annotated by using the RAST software, and the results were confronted to those of BLASTP analysis against the NCBI NR database, and of the CAZy mining. Three clones present a high redundancy (45P14, 25I16 and 50E19), their sequence overlapping, thus creating one bigger contig. Such sequence redundancy frequently occurs between sequences issued from activity-based screening approaches due to the high selection pressure imposed by this strategy (Tasse et al., 2010; Cecchini et al., 2013).

Table 8: ORFs of interest from the sequenced clones active on Tween 20, Impranil and/or pNPA/B/P, annotated as putative esterases, ureases or proteases. Taxonomic annotation according to MEGAN analysis: number of ORFs assigned to F, Firmicutes; P, Proteobacteria; A, Actinobacteria; ESB, bacteria from environmental samples.

Clone ID and ORF number	Degraded Substrates	RAST annotation	Best BLAST hit against the nr NCBI database (accession number) Sequence coverage, sequence identity	NCBI best BLAST hit with the annotated activity of interest	Taxonomic annotation
7D9 32	GM, Impranil (both media), pNPA (9.4×10^{-2} U/mL _{extract}), pNPB (3.5×10^{-3} U/mL _{extract})	Hypothetical protein	putative uncharacterized protein [Clostridium sp. CAG:510] (CDA68954.1) Cov. 97%, Id. 60%	gluconolactonase [Lachnoclostridium phytofermentans] (WP_051650879.1) Cov. 96%, Id. 60%	24F 1ESB
10N15 28	XYL, Impranil (both media)	Hypothetical protein	NHL repeat-containing protein [uncultured bacterium Contig33] (AHF25005.1) Cov. 100%, Id. 90%	gluconolactonase [Paenibacillus sp. VT-400] (WP_047842212.1) Cov. 89%, Id. 37%	6F 22ESB
12F23 20	XYL, Impranil (both media), pNPB (3.3×10^{-2} U/mL _{extract})	Lysophospholipase ; Monoglyceride lipase	lysophospholipase L2 [Clostridium sp. CAG:678] (CCY68436.1) Cov. 98%, Id. 52%	lysophospholipase L2 [Clostridium sp. CAG:678] (CCY68436.1) Cov. 98%, Id. 52%	3F 1P 13ESB
13M17 25	Tween 20, Impranil (both media)	Hypothetical protein	hypothetical protein [Ruminococcus callidus] (WP_021683508.1) Cov. 99%, Id. 53%	metallopeptidase [Anaerostipes hadrus] (WP_044924108.1) Cov. 98%, Id. 52%	20F
		Esterase/lipase	hypothetical protein [Butyrivibrio fibrisolvens] (WP_051216853.1) Cov. 100%, Id. 73%	esterase [Clostridiales bacterium S7-1-4] (WP_034547640.1) Cov. 99%, Id. 39%	
		Hypothetical protein	hypothetical protein [Intestinimonas butyriciproducens] (WP_052082726.1) Cov. 85%, Id. 52%	PREDICTED: mucin-22-like [Bactrocera dorsalis] (XP_011198142.1) Cov. 68%, Id. 28%	
14O19 26	GM, Impranil (both media)	Lipolytic enzyme, G-D-S-L precursor	lipase [Shingomonas taxii] (WP_043059301.1) Cov. 99%, Id. 85%	lipase [Shingomonas taxii] (WP_043059301.1) Cov. 99%, Id. 85%	22F 1P 2ESB
21K7 33	XYL, Impranil (both media), pNPA (3.7×10^{-1} U/mL _{extract})	Hypothetical protein	hydrolase, carbon-nitrogen family [Parvimonas sp. oral taxon 110] (WP_009355038.1) Cov. 97%, Id. 44%	hydrolase, carbon-nitrogen family [Parvimonas sp. oral taxon 110] (WP_009355038.1) Cov. 97%, Id. 44%	10F 4P 13ESB
25I16 29	XYL, Tween 20, Impranil (both media), pNPA (3.0×10^{-2} U/mL _{extract}), pNPB (3.3×10^{-2} U/mL _{extract})	Hypothetical protein	putative transglutaminase/protease [uncultured bacterium Contig1767] (AHF23837.1) Cov. 100%, Id. 100%	putative transglutaminase/protease [uncultured bacterium Contig1767] (AHF23837.1) Cov. 100%, Id. 100%	1F 27ESB
		Putative esterase	putative esterase [uncultured bacterium Contig1767] (AHF23826.1) Cov. 100%, Id. 99%	putative esterase [uncultured bacterium Contig1767] (AHF23826.1) Cov. 100%, Id. 99%	
		Hypothetical protein	hypothetical protein [Lachnospiraceae bacterium C6A11] (WP_035628054.1) Cov. 99%, Id. 46%	diguanylate phosphodiesterase [Butyrivibrio proteoclasticus] (WP_013280495.1) Cov. 99%, Id. 28%	
29D17 29	Tween 20, Impranil (both media), pNPA (7.6×10^{-3} U/mL _{extract}), pNPB (9.2×10^{-3} U/mL _{extract})	Esterase/lipase	esterase [Lactobacillus reuteri] (WP_003670556.1) Cov. 97%, Id. 50%	esterase [Lactobacillus reuteri] (WP_003670556.1) Cov. 97%, Id. 50%	20F
		Hypothetical protein	hypothetical protein [Erysipelotrichaceae bacterium NK3D112] (WP_051665678.1) Cov. 97%, Id. 52%	peptidase family M20/M25/M40 [Firmicutes bacterium CAG:129] (CCZ45637.1) Cov. 96%, Id. 49%	

Discovery of carbamate degrading enzymes by functional metagenomics

30F10 32	XYL, Impranil (both media)	Hypothetical protein	hypothetical protein [Lachnospiraceae bacterium JC7] (WP_044995538.1) Cov. 100%, Id. 25%	peptidase M16 [Mannheimia varigena] (WP_025235452.1) Cov. 47%, Id. 29%	2F 1P 25ESB
		Hypothetical protein	signal peptidase I [uncultured bacterium Contigcl_1764b] (AHF26082.1) Cov. 99%, Id. 62%	signal peptidase I [uncultured bacterium Contigcl_1764b] (AHF26082.1) Cov. 99%, Id. 62%	
		FIG01032054: hypothetical protein	NHL repeat-containing protein [uncultured bacterium Contigcl_1764b] (AHF26075.1) Cov. 100%, Id. 69%	gluconolactonase [Thermobacillus composti] (WP_041853885.1) Cov. 89%, Id. 35%	
32D12 37	Impranil (both media)	Endoglucanase E precursor (EC 3.2.1.4) (EgE) (Endo-1,4-beta-glucanase E) (Cellulase E)	acetyl-xylan esterase [uncultured bacterium Contig1549a] (AHF23770.1) Cov. 100%, Id. 99%	acetyl-xylan esterase [uncultured bacterium Contig1549a] (AHF23770.1) Cov. 100%, Id. 99%	5F 3P 23ESB
		Putative xylanase	acyl-CoA thioesterase I-like protein [uncultured bacterium Contig1549a] (AHF23762.1) Cov. 100%, Id. 90%	acyl-CoA thioesterase I-like protein [uncultured bacterium Contig1549a] (AHF23762.1) Cov. 100%, Id. 90%	
		Lysophospholipase (EC 3.1.1.5); Monoglyceride lipase (EC 3.1.1.23); putative	lysophospholipase L2 [Clostridium sp. CAG:678] (CCY68436.1) Cov. 99%, Id. 53%	lysophospholipase L2 [Clostridium sp. CAG:678] (CCY68436.1) Cov. 99%, Id. 53%	
33M21 63	BGLU, Impranil (both media), pNPA (9.6×10^{-3} U/ mL _{extract})	Hypothetical protein	putative uncharacterized protein [uncultured bacterium Ad_125_D08] (AIF26054.1) Cov. 99%, Id. 81%	short-chain dehydrogenase [Paenibacillus massiliensis] (WP_025679574.1) Cov. 17%, Id. 31%	25F 7P 3A 1Spirochaete 2ESB
		Prophage Clp protease-like protein	hypothetical protein [Clostridium cylindrosporum] (WP_048569386.1) Cov. 92%, Id. 33%	protease-like protein [Vibrio phage X29] (YP_009043932.1) Cov. 88%, Id. 32%	
35C9 32	XYL, Impranil (both media)	Hypothetical protein	signal peptidase I [uncultured bacterium Contigcl_51] (AHF25760.1) Cov. 100%, Id. 98%	signal peptidase I [uncultured bacterium Contigcl_51] (AHF25760.1) Cov. 100%, Id. 98%	3F 26ESB
36E1 34	XYL, Impranil (both media), pNPA (6.0×10^{-3} U/ mL _{extract}), pNPP (7.9×10^{-3} U/ mL _{extract})	Hypothetical protein	metallophosphoesterase [Lachnospiraceae bacterium AC2014] (WP_031543010.1) Cov. 73%, Id. 53%	metallophosphoesterase [Lachnospiraceae bacterium AC2014] (WP_031543010.1) Cov. 73%, Id. 53%	23F 2ESB
		Hypothetical protein	metallophosphoesterase [Subdoligranulum sp. 4_3_54A2FAA] (WP_009325564.1) Cov. 97%, Id. 57%	metallophosphoesterase [Subdoligranulum sp. 4_3_54A2FAA] (WP_009325564.1) Cov. 97%, Id. 57%	
37F15 35	Tween 20, Impranil (rich medium), pNPA (1.9×10^{-2} U/ mL _{extract}), pNPB (9.0×10^{-3} U/ mL _{extract})	putative alpha-dextrin endo-1, 6-alpha-glucosidase	carbohydrate esterase [Streptococcus henryi] (WP_026183046.1) Cov. 100%, Id. 49%	carbohydrate esterase [Streptococcus henryi] (WP_026183046.1) Cov. 100%, Id. 49%	26F 1ESB
		Esterase/lipase-like protein	esterase [Lactobacillus reuteri] (WP_003666275.1) Cov. 96%, Id. 50%	esterase [Lactobacillus reuteri] (WP_003666275.1) Cov. 96%, Id. 50%	
39F1 21	Tween 20, Impranil (both media), pNPA (2.0×10^{-2} U/ mL _{extract}), pNPB (1.6×10^{-2} U/ mL _{extract})	FIG004556: membrane metalloprotease	m50 family membrane metalloprotease [Faecalibacterium sp. CAG:74] (CDE51946.1) Cov. 99%, Id. 46%	m50 family membrane metalloprotease [Faecalibacterium sp. CAG:74] (CDE51946.1) Cov. 99%, Id. 46%	8F 1P 10ESB

Discovery of carbamate degrading enzymes by functional metagenomics

42N18 34	XYL, Impranil (both media), pNPA (8.3×10^{-3} U/ mL_{extract}), pNPB (3.9×10^{-3} U/ mL_{extract})	Hypothetical protein	signal peptidase I [uncultured bacterium Contigcl_51] (AHF25760.1) Cov. 100%, Id. 98%	signal peptidase I [uncultured bacterium Contigcl_51] (AHF25760.1) Cov. 100%, Id. 98%	1F 25ESB
44I12 27	Tween 20, Impranil (both media), pNPA (1.5×10^{-2} U/ mL_{extract}), pNPB (9.6×10^{-3} U/ mL_{extract}), pNPP (5.0×10^{-3} U/ mL_{extract})	probable lipase/esterase	hypothetical protein [[Clostridium] stercorarium] (WP_015360182.1) Cov. 95%, Id. 46%	esterase [Lactobacillus vaginalis] (WP_003717503.1) Cov. 95%, Id. 43%	22F (10 Clostridiales) 1 seul Thermobacillus) 2ESB
45P14 34	XYL, Impranil (both media)	Putative esterase	putative esterase [uncultured bacterium Contig1767] (AHF23826.1) Cov. 100%, Id. 99%	putative esterase [uncultured bacterium Contig1767] (AHF23826.1) Cov. 100%, Id. 99%	7F 13P 13ESB
49M8 31	BGLU, Impranil (both media), pNPA (2.1×10^{-2} U/ mL_{extract})	Hypothetical protein	hypothetical protein [Ruminococcus champanellensis] (WP_015558043.1) Cov. 97%, Id. 38%	CAAX amino terminal protease family protein [[Clostridium] methylpentosum DSM 5476] (EEG30438.1) Cov. 62%, Id. 25%	22F 2P 1ESB
		N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25)	mucin-desulfating sulfatase [Clostridium sp. CAG:1024] (CCX42339.1) Cov. 49%, Id. 56%	N-acetylglucosamine-6-phosphate deacetylase [Paenibacillus sp. URHA0014] (WP_036741135.1) Cov. 47%, Id. 45%	
		Beta-xylosidase (EC 3.2.1.37)	hypothetical protein [Ruminococcus sp. HUN007] (WP_049962627.1) Cov. 93%, Id. 58%	esterase [Cellulosilyticum ruminicola JCM 14822] (ACZ98648.1) Cov. 51%, Id. 43%	
		Rhomboid family protein	peptidase S54 [Clostridium homopropionicum] (WP_052222902.1) Cov. 90%, Id. 58%	peptidase S54 [Clostridium homopropionicum] (WP_052222902.1) Cov. 90%, Id. 58%	
		Hypothetical protein	transglutaminase-like enzymes putative cysteine proteases [Ruminococcus sp. CAG:624] (CDF02576.1) Cov. 94%, Id. 27%	transglutaminase-like enzymes putative cysteine proteases [Ruminococcus sp. CAG:624] (CDF02576.1) Cov. 94%, Id. 27%	
		Hypothetical protein	hydrolase alpha/beta domain protein [Faecalibacterium sp. CAG:74] (CDE51046.1) Cov. 91%, Id. 36%	esterase/lipase-like protein [Epulopiscium sp. 'N.t. morphotype B'] (WP_010168052.1) Cov. 91%, Id. 26%	
50E19 28	XYL, Tween 20, Impranil (both media), pNPA (3.7×10^{-2} U/ mL_{extract}), pNPB (4.2×10^{-2} U/ mL_{extract})	Putative esterase	putative esterase [uncultured bacterium Contig1767] (AHF23826.1) Cov. 100%, Id. 99%	putative esterase [uncultured bacterium Contig1767] (AHF23826.1) Cov. 100%, Id. 99%	27ESB
50I3 29	Tween 20, Impranil (rich medium), pNPA (5.8×10^{-3} U/ mL_{extract}), pNPB (7.6×10^{-3} U/ mL_{extract})	N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25)	n-acetylglucosamine-6-phosphate deacetylase [Ruminococcus sp. CAG:382] (CDD02602.1) Cov. 99%, Id. 55%	n-acetylglucosamine-6-phosphate deacetylase [Ruminococcus sp. CAG:382] (CDD02602.1) Cov. 99%, Id. 55%	18F 4ESB
		Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	putative uncharacterized protein [Blautia sp. CAG:237] (CDB78797.1) Cov. 90%, Id. 45%	putative esterase [Clostridium sp. KLE 1755] (WP_021635315.1) Cov. 91%, Id. 46%	

		Putative esterase	hypothetical protein [Hungatella hathewayi] (WP_006780815.1) Cov. 96%, Id. 57%	tributylin esterase [Clostridiales bacterium VE202-28] (WP_025484029.1) Cov. 96%, Id. 56%	
--	--	-------------------	--	---	--

All the hit clone sequences contain at least one gene encoding putative esterases, ureases or proteases. Among the 21 sequenced hit clones, 10 contain at least two putative enzyme encoding genes that may act synergistically on the targeted substrates. In total, 26 putative esterases, 11 putative proteases and 3 putative ureases encoding ORFs were found (Table 8). The fact that no protease hit clone was found after primary and secondary screening, indicates either that the putative protease encoding genes were not properly expressed in *E. coli*, or, if well expressed, that these enzymes are unable to break down the particular structure of the substrates used in this study.

According to Gabor et al. (2004), *E. coli* is predicted to readily express around 40% of environmental genes, with strong variations depending on the group of organisms. One example of this variation is the work of McMahon et al. (2012), who screened a metagenomic library created in *Streptomyces lividans*, where they found many hits that were not bioactive in *E. coli*, known to express no more than around 7% of the high GC% Actinomycete DNA. Here, our activity-based screening approach allowed to select sequences mainly derived from Firmicutes or uncultured bacteria from environmental samples, highlighting the potential of *E. coli* to express genes issued from bacteria that are very distant from Proteobacteria.

Sequences of clones 12F23 and 44I12 were further analysed. Each of these metagenomic inserts encodes one putative esterase, considered as the best candidate genes to be responsible for the screened activities.

The best BLAST hit of the 12F23 putative esterase, which presents 52% identity (98% sequence coverage) with its protein sequence, is the lysophospholipase L2 from *Clostridium* sp. CAG: 678. That of the 44I12 putative esterase target (46% identity, 95% coverage) is a hypothetical protein from *Clostridium stercorarium*. These results highlight the originality of our hit sequences, which are very distant from any characterized enzymes. Further analyses of these interesting sequences were thus performed to get information on the protein families and subfamilies they belong to.

The best characterised homolog of the esterase from clone 12F23, that we named carboxyl-esterase 1 from unknown bovine ruminal bacterium (CE1_Ubrb), is the PLA-depolymerase of an uncultured bacterium issued from the adherent soil of PLA disks buried in compost (Mayumi et al., 2008), with which it shares 39 % identity for 92 % sequence coverage. This enzyme is able to degrade different kinds of polyesters but it was never tested on carbamates, to our knowledge. It is more active on *p*NPB compared to *p*NPA, which is also the case for the 12F23 enzymatic extract. Four other characterised enzymes

were found among the 6,127 sequences of the NCBI nr database presenting more than 25 % identity with CE1_Ubrb, on more than 50 % of sequence length and with an e-value inferior to 0.001. These other characterised enzymes are annotated as monoglyceride lipase or as their larger appellation acylglycerol lipase. They are part of the glycerolipid metabolism, since they hydrolyse the triglycerides stored in adipocytes or other cells to fatty acid and glycerol. They are members of the serine hydrolase family, and all contain the GXSXG consensus sequence. The phylogenetic tree (Figure S2), made from the first 250 best BLAST hits against the nr NCBI database and the sequence of characterised enzymes, highlights the great distance between these four other sequences, and that of the PLA depolymerase. Moreover, the branch containing CE1_Ubrb and 34 other sequences is a furnished one. It contains sequences annotated as uncharacterised proteins, putative lysophospholipases, and putative hydrolases with alpha/beta fold (or domain). It is interesting to note that they predominantly belong to the Firmicute phylum (*Clostridium*, *Eubacterium*, *Anaerobacter*, *Subdoligranulum* genera), the second most abundant phylum in the rumen of adult cow and also a phylum known to contain polyurethane degrading bacteria (Cregut et al., 2013). This is in accordance with the taxonomical annotation profile of the 12F23 contig, which harbours three genes assigned to the Clostridia class of the Firmicute phylum.

According to Arpigny and Jaeger (1999) bacterial lipolytic enzymes can be classified in eight families, depending on the conserved sequence motif and their biological properties. According to a multiple sequences alignment performed using Clustal Omega, CE1_Ubrb would belong to family V (Figure 22) (Tirawongsaroj et al., 2008; Hu et al., 2010). This alignment shows a conservation of the GXSXG consensus sequence containing the catalytic serine, along with the two other catalytic amino acids, an aspartic acid and a histidine, and with many other conserved amino acids. Lipolytic family V is supposed to contain enzymes from mesophilic bacteria or from cold-adapted or heat-adapted organisms. Our target probably belongs to a mesophilic organism, considering the cow rumen normal temperature varies between 38 and 41°C. Enzymes from family V also share significant homology with other bacterial enzymes like epoxide hydrolases, dehalogenases, and haloperoxidases that also possess the typical α/β hydrolase fold. To our knowledge, no enzymes from this lipolytic family has been found active on carbamates, family VII being the one mostly known for this ability (Hausmann and Jaeger, 2010). The best BLAST hit found against the ESTHER database presents 43% of identity with CE1_Ubrb. It corresponds to *Janibacter* sp. Lysophospholipase L2, a member of the Monoglyceridelipase_lyso phospholip family, and part of the X block. In this block, tannases (enzymes also present in the rumen), PLA depolymerases and other enzymatic families can be found. InterProScan also predicts an α/β hydrolase fold for CE1_Ubrb, and a domain (residues 18 to 278 or 292) related to phospholipases, monoglyceride lipases or serine aminopeptidases. This is not surprising, as lysophospholipases and monoacylglycerol lipases (also known as monoglyceride lipases) are part of the glycerophospholipid and glycerolipid metabolism pathways, respectively, of bovine ruminal bacteria, for instance belonging to the Clostridia class of the Firmicute

phylum (Shingfield and Wallace, 2014). CE1_Ubrb might thus originally be involved in the lipolysis happening in the rumen. Lipolysis leads to biohydrogenation, the conversion of dietary fats to saturated fatty acids, which consumes a small part (1 to 2%) of the H₂ produced by complex carbohydrate fermentation. Microorganisms involved in such mechanism very often carry out other functions (Lourenço et al., 2010), which might explain the presence in the 12F23 metagenomic insert of genes encoding enzymes belonging to families 10 and 13 of glycoside-hydrolases, that would be involved in the degradation of plant polysaccharides. The presence of these CAZymes gives credence to the choice of the primary screening strategy, especially for this clone. 12F23 was indeed isolated from screening on AZCL-xylan, but was inactive on Tween 20, even being able to degrade Impranil.

CE1_Ubrb	-----MCE-	3
gi 3342450 gb AAC67392.1	-MNRD-----AIKILGAGAIAGALATRIPLSISQTPPIQINY----SNLVPKEFGWAPV	49
gi 747876 emb CAA47949.1	MLLKRLCFAALFSLSMVGCCTNAPNA-----LAVNTTQKIIQYERNKSDLEIKSLTLASG	54
gi 44523 emb CAA37863.1	MLLKRLGLAALFSLSMVGCCTTAPNT-----LAVNTTQKIIQYERSKSDLEVKSLTLASG	54
CE1_Ubrb	ISAHYITPDDGFFETVMVIHHGMAEHQKRYEEFIRFLCDHGIAYVMHDMANHGESCQ---	60
gi 3342450 gb AAC67392.1	NNIHIIYYEIIYGSCEP-LIMIMGYLGNLESWGP I IINGLASQYEV I IFDNRGTGRSGTVGQ	108
gi 747876 emb CAA47949.1	DKMVAENGNVAGEP-LLLIHGFGGNKDNFTRIAR--QLEGYHLIIPDLLGFGESSKPMMS	111
gi 44523 emb CAA37863.1	DKMVAENDNVTGEP-LLLIHGFGGNKDNFTRIAD--KLEGYHLIIPDLLGFGNSKSPMT	111
	. * : : * . . : : : * * *	
CE1_Ubrb	DPELKGWFGKKGWGNALIEDYREVMVR-ARQEHAPACRLIVMGHSMGFSFIVRMYVSRYPKD	119
gi 3342450 gb AAC67392.1	DPLHDA-----LTYTIPLYASDTIGLLNVLG-YSNLNVLGWSMGGFVAQQIAIDYPSY	160
gi 747876 emb CAA47949.1	AD-----YRSEAQRTRLHELQAKGLASNIHVGGNSMGGAISVAYAAKYPKD	158
gi 44523 emb CAA37863.1	AD-----YRADAQATRLHELMQAKGLASNTHVGGNSMGGAISVAYAAKYPKE	158
	. . * * * * . . * *	
CE1_Ubrb	GFAGAVFMGTGGPNPAAGAGKAV--SAMIGAVRGKKHRSSLMNTMAFGTYGNKFEGRTPF	177
gi 3342450 gb AAC67392.1	VNKLVL--CTAPNIYLYPPKVSPQSIITGFTASD-----PTVT---VETIIPYLVP	208
gi 747876 emb CAA47949.1	VKSLWLV--DSAGFWSAGIP---KSL-EGATLEN-----NPLL---IKSNEDFYK--	199
gi 44523 emb CAA37863.1	IKSLWLV--DTAGFWSAGVP---KSL-EGATLEN-----NPLL---INSKEDFYK--	199
	: . . * * . . : . .	
CE1_Ubrb	DWLTRDKAIVDRYIADPYCGFMFPVQG--MNDLVRLNA-----ASNTAAWARD	223
gi 3342450 gb AAC67392.1	DWLQHPDVAKYVLFITL---KYPISYTSVLKQ-----TNALATFNSVVGQ	250
gi 747876 emb CAA47949.1	-----MYDFVMYKPP---YLPKSVKAVFAQERIKNKELDKILEQIVTDNVEERA	248
gi 44523 emb CAA37863.1	-----MYDFVMYKPP---YIPKSVKAVFAQERINNKALDTKILEQIVTDNVEERA	248
	: . : * . : : . .	
CE1_Ubrb	-IPAEPLILLVSGGLDPVGDYAGVQKVGELLAETGHRVITTKIYPECR-----HE	273
gi 3342450 gb AAC67392.1	LQNTAPTLLVIGGSDLL-----PPQNSQYLAENIPNAQLYIFSPDAGHG	296
gi 747876 emb CAA47949.1	IAQYKIPTLVVWGDKDQII-----KPETVNLIKKIIPQAQVIMME-DVGHV	293
gi 44523 emb CAA37863.1	IAKYNIPTLVVWGDKDQVI-----KPETTELIIKIIIPQAQVIMMN-DVGHV	293
	* * : * * : . : * : *	
CE1_Ubrb	VLNEINRDEVMNDITEWIRGI----	294
gi 3342450 gb AAC67392.1	LIYQ-YPTQFINLVTSFLG-----	314
gi 747876 emb CAA47949.1	PMVE-ALDETADNYKAFRSILEAQR	317
gi 44523 emb CAA37863.1	PMVE-AVKDTANDYKAFRDGLKK--	315
	: : : : :	

Figure 22: Sequence alignment between CE1_Ubrb and enzymes from the lipolytic family V, chosen from the studies of Hu et al. (2010) and Tirawongsaroj et al. (2008): a lipolytic enzyme from *Sulfolobus acidocaldarius* (AAC67392.1), a triacylglycerol lipase from *Psychrobacter immobilis* (CAA47949.1) and a triacylglycerol lipase *Moraxella* sp (CAA37863.1). The red boxes are catalytic amino acids, and the blue box the GX SXG consensus sequence. "*" is for fully conserved residue, ":" shows the conservation between groups of residues with strongly similar properties, and "." between groups with weakly similar properties.

The best characterized homolog of the carboxyl-esterase from clone 44I12, that we named CE2_Ubrb, is the esterase LC-Est2 from an uncultured bacterium, discovered from a leaf-branch compost metagenome (Okano et al., 2015). LC-Est2 shares 32 % identity on 94 % of sequence length with CE2_Ubrb, and is active on tributyrin, a substrate specific to esterases and lipases. This is in accordance with CE2_Ubrb being active on all *p*NP esters tested. LC-Est2 has never been tested on carbamates, to our knowledge.

Seven other characterised enzymes were found among the 5,662 sequences of the NCBI nr database presenting more than 25 % identity with CE2_Ubrb, on more than 50 % of sequence length and an e-value inferior to 0.001. The characterised enzymes are annotated as lipases/esterases, or lipolytic enzymes. All of them are part of the serine hydrolase family, containing the GX SXG consensus sequence, and take part in prokaryotic and eukaryotic lipolysis.

The phylogenetic tree (Figure S3), made from the first 250 best BLAST hit against the nr database of NCBI, along with the sequence of characterised enzymes, highlights the distance between CE2_Ubrb sequence and those of the other characterized enzymes. The branch closest to the CE2_Ubrb's one with a characterised enzyme contains EstJ, an alkaliphilic and moderately thermophilic esterase from the lipolytic family IV, identified from a soil metagenome. The CE2_Ubrb branch also contains many sequences from Firmicutes, especially from the *Lactobacillus* genus, which is in agreement with the taxonomic assignation of the 44I12 metagenomic insert. According to MEGAN analysis, one ORF of the 44I12 contig was indeed assigned to the Bacilli class of the Firmicute phylum, while ten were assigned to the Clostridia class of the same phylum. According to a multiple sequences alignment performed using Clustal Omega (Figure 23), CE2_Ubrb would belong to the lipolytic family IV, composed of many esterases from distantly related prokaryotes (Jeon et al., 2012; Li et al., 2014). The enzymes of this family are remarkably similar to the family of mammalian hormone-sensitive lipases (HSL), containing the S, D, H catalytic residues, along with a G that is involved in hydrogen bonding interaction, stabilizing an oxyanion hole formed during enzymatic reaction and promoting the catalysis (Hausmann and Jaeger, 2010). The alignment shows conservation in the CE2_Ubrb sequence of the GX SXG consensus motif and of the three catalytic amino acids, along with the amino acids that are part of the oxyanion hole. Bacterial HSL supposedly show activity only towards short fatty acid chain length substrates (tributyrin, vinyl propionate) (Chahinian et al., 2005) compared to mammalian HSL that have a broad substrate specificity (olive oil, short chain esters, trioctanoin...). The CE2_Ubrb activity profile shows a closer similarity to mammalian HSL than to bacterial ones, our target being active on short chain esters but also on Impranil and its long motif.

CE2_Ubrb	MSIR-----VIPNNPEFKGLVEMLGDIEFAAPDGHALKL---	34
gi 305849870 gb ADA70028.2	MDNRMIRLAAFLVILMASAGCLTSRVAAQELLPVSP-GVTVLRDRTPSPVGVSSSENTAI	59
gi 268634712 gb ACZ16565.1	MKTRFFQRLGI-SAAVSLACFQ-AAVAQELPPIFP-GVTTLGERITPAPVGLSPEMLEV	57
gi 170179986 gb ACB11219.1	-----MTSQESTT-----GNKSRE-TVWRIGPRILPPPTGGSAELRDA	37
gi 219957612 gb ACL67837.1	-----MTNQKSSST-----DNQAQE-NTWHIGPRTLPSFAGASDVLYNS	37
	. : : * * : :	
CE2_Ubrb	-----H	35
gi 305849870 gb ADA70028.2	IASRQIPADL---PLPTTTEGWLEFQRVDFEPGGELARQGAEHLGATYEVQEIAGVRTY	115
gi 268634712 gb ACZ16565.1	VKERQIPVVI---PAPETTEGWLELQALFDVAGEELGRAAAAAYNEVTYEVKEIAGVRTY	113
gi 170179986 gb ACB11219.1	IASIPQDPAT-MKIEPQSEEWLAVIAQLDAGKVEANQALREQLSVSVTQEEIEGINVY	96
gi 219957612 gb ACL67837.1	VMNTPTPDPAPNLALDPQSEAEWLAVIAQLDEGKVDMMAREISKQLSASGEHDTIEGVNVI	97
	:	
CE2_Ubrb	IIRPWKRT-TRGYPLVVFVIGSSWT-----TPDQYWEIPQLSQAARGFVVATVTHRS	87
gi 305849870 gb ADA70028.2	VVTPRMINKRFADRVFVHVHGGAFVFGGDSAMREAIWAAHGL-----GVKVISVDYRR	169
gi 268634712 gb ACZ16565.1	IVTPKELNERWGRDRFFVYHGGAFVFGGDSALREAVWAAHGL-----GVKVVSVDYRR	167
gi 170179986 gb ACB11219.1	HLTPAEVDPRHEDHLFVYHLGGAFVNLAGEAGLAESIIIAHRL-----KMRLMHIDYRM	150
gi 219957612 gb ACL67837.1	HVTPAEIDPGLDEKLFHHTGGAFILNGGVAGTIEALVIASLA-----KVRVLSIDYRM	151
	: * . . . : : : : :	
CE2_Ubrb	CFEAKAPAFITDVKAAIRFLKVKASEYDIDPDRVCANGTSSGGNAALLVGMTGDDPAFES	147
gi 305849870 gb ADA70028.2	PPTYPPFAAIDDVIAVWEEV-----TKDQSAGETALFGTSAGGNITLATVCLKKE---LG	221
gi 268634712 gb ACZ16565.1	PPLHPFPAAIEDALAVWREV-----TKDQDAAGTALFGTSAGGNIALATTLCLKKE---LG	219
gi 170179986 gb ACB11219.1	PKKYPTPAGRDDVVTYVYKHL-----LKERPARSMAMGGTSSGGANMTMATVQRLE---LG	202
gi 219957612 gb ACL67837.1	PPLHPAPAGRDDVMTVYVYQHL-----LTQRPAQKIAMGSSGGANLTMGMVQLLIK---QG	203
	** * . . . : : . * . * . * : :	
CE2_Ubrb	ADWPGVSSRVQAVVDCFGPTDLMRMMEVQYADQLDEDGALFYALGGGESADACRAAMT--	205
gi 305849870 gb ADA70028.2	KPLPGAVFAG-----TPATDLENTIDTWTYLEGGLDP-LGKREGLIST	262
gi 268634712 gb ACZ16565.1	ESLPGAVFAG-----TPATDLENISDTWTYLEGGLDP-LGARDGMLAA	260
gi 170179986 gb ACB11219.1	LDVPGVLYLG-----TPGTDVSNITGDSWIINEGIDHLLITREGMLEA	244
gi 219957612 gb ACL67837.1	IDVPGALFLG-----TPGADMSKTGDSYIINDGIDRNIITYDGIIEA	245
	** . . . : : * : : :	
CE2_Ubrb	-----RISPIAYAQPGRDFPFLMLHGDADPVVLYEDTERLYRRLVLELGYA	251
gi 305849870 gb ADA70028.2	TFTLYAAGEDLSNPMLSP--VNGDLKDFPPTILISGTRD--LLLSDTVMMHRAALRAAGVV	318
gi 268634712 gb ACZ16565.1	TFDIYARGADPSNPPLSP--INGDLGFPPTILISGTRD--LLLSDTVMMHRAALRKVGVE	316
gi 170179986 gb ACB11219.1	CVAVYAAGRDPKDPVWSP--YVGDVHGFPPPTLLITGTRD--MLLSSTARHTRIKLRQAGVV	300
gi 219957612 gb ACL67837.1	SVRLYANGRDLKDPVWSP--LYGDFHGFPPPTFLVTGTRD--LLLSATVTRTHIKLRQTGVV	301
	: ** *** : : * : * . * * : * *	
CE2_Ubrb	ADLVRVSGAEHGSFWSE-----TVLEIIFAFIEEKIGGARNTGTFEND	295
gi 305849870 gb ADA70028.2	ADLHVYDGGQAHG DYLQNLIRHVPEAEDAQRELF E FFDKHLK-----	359
gi 268634712 gb ACZ16565.1	ADLHVYDGGQTHADYMQNLIRYVPESEDAQRELYSFFDKHLE-----	357
gi 170179986 gb ACB11219.1	ADILVYDGVSHGDYIFVM--NAPESAHAYAE LNAFL LQHLQ-----	339
gi 219957612 gb ACL67837.1	ADILVYEGMAHADYMDL--TAPETQHAFAE LNAFL LQHLQ-----	340
	** : * * : : : * : : :	

Figure 23: Sequence alignment between CE2_Ubrb and enzymes from the lipolytic family IV chosen from the studies of Jeon et al. (2011) and Li et al. (2014): a lipolytic enzyme precursor from uncultured marine bacterium (ADA70028.2), a putative esterase from uncultured bacterium (ACZ16565.1), an esterase from uncultured bacterium (ACB11219.1) and a lipolytic enzyme from uncultured bacterium (ACL67837.1). The red boxes are catalytic amino acids, the green box the oxyanion hole and the blue box the GXSXG consensus sequence. "*" is for single, fully conserved residue, ":" shows the conservation between groups of residues bearing strongly similar properties, and "." between groups of weakly similar properties.

Against the ESTHER database, the best BLAST hit was indeed found to be *Lactobacillus reuteri* 100-23 HSL, sharing 43% of identity. This enzyme is a member of the HSL-like family, which according to the ESTHER database, corresponds to family IV of the

classification of lipolytic enzymes from Arpigny and Jaeger (1999) for bacterial enzymes. This concurs with the results of alignment analysis previously described. The ESTHER database also mentions the affiliation of these HSL-like family to the block H, containing, among others, plant carboxylesterases and bacterial tannases, the first possibly being present in the rumen through food components, the second being produced by ruminal bacteria. It also contains kynurenine-formamidases, a pattern found in the sequence of CE2_Ubrb using the InterProScan software, lending credence to the possible affiliation of this enzyme to the group IV of lipolytic enzymes. As a possible HSL-like enzyme, it could potentially hydrolyse all acylglycerols (triacylglycerol, diacylglycerol and monoacylglycerol) as well as cholesteryl esters, steroid fatty acid esters, retinyl esters and *p*-nitrophenyl esters.

3. Material and methods:

3.1. Chemicals

Impranil DLN was supplied by Bayer Corporation.

Tween 20, Pefabloc, *para*-nitrophenyl acetate (*p*NPA), *para*-nitrophenyl butyrate (*p*NPB), *para*-nitrophenyl palmitate (*p*NPP), *para*-nitrophenol (*p*NP), tetrahydrofuran (THF), 2-*sec*-butylphenol, (2-Butan-2-ylphenyl), N-methylcarbamate (fenobucarb), Ethyl 2-(4-phenoxyphenoxy)ethylcarbamate (fenoxycarb) and S-Benzyl dipropylthiocarbamate (prosulfocarb) were purchased from Sigma Aldrich (France).

AZCL-xylan (XYL), AZCL-casein (CAS), AZCL-Barley β -glucan (BGLU), AZO-CM Cellulose (CMC), AZO-Carob Galactomannan (GM) were purchased from Megazyme (Ireland).

Lysonase™ Bioprocessing Reagent was purchased from Novagen.

3.2. Metagenomic DNA sampling and library construction

Before enrichment, the non-producing Holstein dairy cows used in this study were reared according to the national standards fixed by the legislation on animal care (Certificate of Authorization to Experiment on Living Animals, No. 004495, Ministry of Agriculture, France). In particular, the cows were fed a standard dairy cow ration composed of corn silage (64% DM), hay (6% DM) and concentrate (30% DM). They were fed *ad libitum* once a day in the morning. The two cows received during 7 weeks a diet of 80% wheat straw, 12% concentrate and 8% beetroot molasses before sampling of their rumen content. Rumen contents were taken from various parts of the rumen and manually homogenized.

The metagenomic library was constructed in *Escherichia coli* strain from the mix of rumen bacterial DNAs extracted from both cows. Briefly, total metagenomic DNA was extracted and fragmented by using a linear sucrose gradient. Fragments sizing between 30 and 40 kb were isolated as previously described (Tasse et al., 2010) and cloned into pCC1FOS fosmid (Epicentre Technologies). EPI100 *E. coli* cells were then transfected to obtain a library of 19,968 clones from the rumen samples. Recombinant clones were transferred to 52 384-

well microtiter plates containing Luria Bertani (LB) medium, supplemented with 12.5 mg/L chloramphenicol and 8% glycerol. After 22 hours of growth at 37°C without shaking, the plates were long-time stored at -80°C.

3.3. Primary high-throughput screening of the metagenomic library

The resulting library was gridded on 22 cm x 22 cm trays containing solid agar medium supplemented with 12.5 mg/L chloramphenicol, using an automated microplate gridded (K2, KBiosystem, Basildon, UK), which allowed to perform 140,000 assays in 2 days.

Carboxylesterase screening medium was constituted by LB medium supplemented with 1% (w/v) final concentration of Tween 20. The assay plates were incubated during 3 days at 37°C until apparition of a powdery halo, signifying the presence of a positive clone.

Protease activity screening medium was LB medium, supplemented with 0.2% (w/v) final concentration of AZCL-casein. Protease screening plates were incubated during 3 weeks at 37°C.

Polysaccharide degrading activity screening media were LB media, supplemented with chromogenic substrates (AZCL-xylan, AZCL-Barley β -glucan, AZO-CM Cellulose, AZO-Carob Galactomannan) at final concentration of 0.2% (w/v). Plates were incubated at 37°C for 1 to 10 days. For plates containing AZCL-substrates, positive clones were visually detected by the presence of a blue halo resulting from the production of coloured oligosaccharides of which colour diffused around the bacterial colonies. For AZO-polysaccharides, clear halos were observed around the positive clones.

Each positive clone was picked, and streaked on Petri dishes containing solid agar medium supplemented with 12.5 mg/L chloramphenicol. Three isolated colonies were picked and grown in liquid LB medium and stored in microplates containing LB medium with 8% of glycerol, to be used for the validation process.

3.4. Discrimination screening

Validated hit clones obtained after primary screening were grown in microplates containing 200 μ L of LB medium for pre-culture, at 37°C overnight. These pre-cultures were used to sow deep well micro-plates containing 1.6 mL of LB liquid medium supplemented with 12.5 mg/L chloramphenicol, which were subsequently incubated overnight at 37°C. Cells were pelleted by centrifugation for 30 minutes at 4°C and 1,760 x g. Each pellet was re-suspended in 250 μ L of lysis buffer (100 mM HEPES pH 7.5, 1 mM pefabloc and 666.7 μ L/L lysonase), and incubated for one hour at 37°C and 600 rpm (shaking throw 25 mm) in a Multitron Pro shaker (INFORS HT) with a freeze/unfreeze cycle at -80°C in order to perform cell lysis. Cellular extracts were isolated by centrifugation for 30 minutes at 4°C and 1,760 x g, and stored at 4°C for less than 24 h before use.

Carboxylesterase activity of the cytoplasmic extracts was estimated by monitoring the hydrolysis of *p*NP esters into the corresponding acid and *p*-nitrophenol. Assays were performed in 96-well microtiter plates on *p*NPA, *p*NPB and *p*NPP at 1 mM final concentration. *p*NP was used for standard curve. For *p*NPA and *p*NPB, 195 μ L of cellular extract were mixed with 5 μ L of substrate dissolved at 40 mM in 2-methyl-2-butanol (2M2B). For *p*NPP, 175 μ L of cellular extract were mixed with 25 μ L of substrate dissolved at 8 mM in isopropanol. Enzymatic activity was determined by following the absorbance increase at 405 nm during 30 min at 30°C, in a microplate spectrophotometer (BioTek™ Eon™ Microplate Spectrophotometers, Colmar, France).

3.5. Sequencing and data analysis

Fosmid DNA of the positive clones was extracted with the NucleoBond Xtra Midi kit from Macherey-Nagel (France) following supplier recommendations. Fosmids were then individually labelled to avoid any further miscontigation (Lam et al., 2014), and sequenced with the MiSeq technology, at the Genotoul platform (<http://get.genotoul.fr/>). Read assembly was carried out using Masurca (<http://www.genome.umd.edu/masurca.html>). Contigs were cleaned from the vector pCC1FOS sequence using Crossmatch (<http://bozeman.mbt.washington.edu/phredphrapconsed.html>). ORF detection and functional annotation was performed using RAST software (Aziz et al., 2008).

Carboxylesterase encoding genes, as annotated by RAST, were confirmed by BLASTP analysis of the predicted ORFs against the NCBI non-redundant protein sequences database (E-value $\leq 10^{-8}$, identity $\geq 35\%$, query length coverage $\geq 50\%$). Further sequence analysis was performed with other BLASTP analysis against the ESTerases and alpha/beta-Hydrolase Enzymes and Relatives (ESTHER) (Lenfant et al., 2013). The protein signature was detected using the InterProscan software (Jones et al., 2014). Sequence alignments and phylogenetic trees were designed using the MEGA6 software (Tamura et al., 2013). The first 250 sequences from the BLASTP results as well as homologs (E-value $\leq 10^{-3}$, identity $\geq 25\%$, query length coverage $\geq 50\%$) experimentally characterised were aligned with ClustalW from the MEGA6 software. A distance matrix was generated from the multiple sequence alignment using the BLOSUM62 amino acid residue substitution matrix. The output result file was subjected to hierarchical clustering using Ward's method, and the resulting tree was visualized using DENDROSCOPE 3 (Huson and Scornavacca, 2012).

Sequence alignments using Clustal Omega were performed between CE1_Ubrb and CE2_Ubrb gene sequences and sequences known to be part of lipolytic families.

The most probable common ancestor of the organism from which the metagenomic DNA insert came from was retrieved using MEGAN v5.10.6 (Huson et al., 2007), based on BLASTX analysis against the non-redundant NCBI database (min score = 35, min support = 1).

3.6. Functional characterisation of enzymatic activities

3.6.1. Carbamate and polyurethane degradation reactions

The positive clones were grown in tubes containing 3 mL of LB medium for preliminary culture, at 37°C overnight. These cultures were used to sow 20 mL cultures grown in LB medium at 37°C, with orbital shaking at 120 rpm (shaking throw 25 mm). After 24 h, cells were harvested by centrifuging for 5 min at 12,857 x g, and re-suspended in activity buffer (PBS pH 7.0) to obtain a final OD_{600nm} of 80. Cell lysis was done using sonication. Cell debris were centrifuged at 12,857 x g for 10 min and the cytoplasmic extracts were filtered with a 0.20 µm Minisart RC4 syringe filter.

Enzymatic reactions were carried out at 30°C in hemolysis tubes, containing 500 µL of cytoplasmic extract, mixed with 500 µL of the substrate solubilised in the activity buffer (PBS pH 7.0) to obtain a final concentration of 6 g/L for the Impranil dispersion, or 2 g/L fenoxycarb, fenobucarb or prosulfocarb (for structures see Table 7).

The progress of the reaction was followed by taking samples at initial time and after 24h reaction for HPLC analysis. Samples were incubated at 95°C for 5 minutes to stop the reaction for the small carbamates. Impranil reactions were directly frozen at - 80°C to stop the reaction.

3.6.2. High pressure size exclusion chromatography (HP-SEC): polyurethane characterisation

Samples were lyophilised before being re-suspended in 400 µL of HPSEC mobile phase (THF). The SEC experiments were performed at the Toulouse White Biotechnology analytical platform (Toulouse, France) on an HPLC system connected to a RI detector (Wyatt), with in-line degasser (Shimadzu). Size-exclusion chromatography of polyurethane samples was performed on two PLgel 5 µm MIXED-C columns (300 x 7.5 mm) with THF as the mobile phase, pumped at a flow rate of 1 mL/min. Samples were observed with RI detector. Data were recorded and chromatographic peaks were treated using ASTRA software.

3.6.3. Matrix assisted laser desorption/ionization - time-of-flight (MALDI-TOF) mass spectrometry

Samples were lyophilised before being re-suspended in 400 µL of THF. Samples were then analyzed onto a commercial MALDI-TOF mass spectrometer MALDI-TOF Micro Mx (Waters), Laser N2 (337 nm), acceleration voltage 12 kV, reflectron mode, dithranol matrix and NaI salts, in the linear and positive mode for a range of m/z 200-1,200. Spectra were acquired automatically using a standard procedure at the Service Commun de Spectrométrie de Masse at Université Paul Sabatier in Toulouse, France.

3.6.4. High performance liquid chromatography – mass spectrometry

Samples were dissolved in 1.5 volumes of acetonitrile, in order to recover a ratio of 60% acetonitrile, 40% water. Analyses were performed on an HPLC system (Ultimate 3000, Dionex) with in line degasser. LC of small carbamate samples was performed on a SUPELCOSIL™ LC-18 HPLC Column 5 µm particle size (250× 4.6 mm) equipped with a SUPELCOSIL™ LC-18 Supelguard™ Cartridge 5 µm particle size (20 × 4.0 mm) pre-column. The mobile phase (60% acetonitrile-40% water) was pumped at a flow rate of 1 mL/min, and UV detection was carried out at 205 nm, 210 nm and 235 nm for fenobucarb, prosulfocarb and fenoxycarb, respectively.

The LC system was hyphenated with a mass spectrophotometer (MSQ PLUS, Thermo Scientific) operated in positive mode and equipped with an electrospray ionisation (ESI) source for 30 min with a scan time of 0.10 seconds over the range m/z 100-1,500 with nitrogen (N₂) as drying gas. A cone of 60 V was used, and the probe temperature was set at 450°C. 40% of the sample goes through the RI detector, 60% to the mass spectrometer. Chromeleon software was used for data acquisition and analysis.

4. Conclusion

In this study, functional metagenomics was used to discover new enzymes from the bovine rumen microbiota, that are able to degrade various carbamate structures, including polyurethane and insecticide ones.

Two of the isolated enzymes were more deeply studied. They present very original sequences, and belong to the lipolytic families IV and V, respectively. They are the first members of these families that act on carbamates, since all the other known carbamate degrading enzymes belong to family VII. The substrate specificity of CE2_Ubrb enzyme is highly flexible, this enzyme being able to degrade many substrates (namely Tween 20, Impranil, *p*NPA, B and P, and fenobucarb), while CE1_Ubrb specificity was restricted, among the tested esterase substrates, to Impranil and *p*NPB.

This work is the first metagenomic study that enabled the discovery of fenobucarb and Impranil degrading enzymes. The diversity of esterase sequences retrieved in this study from the bovine rumen metagenome highlights again the interest of this complex microbiome for the mining of promiscuous enzymes of biotechnological interest, of which the physiologic functions can be diverted to break down unnatural substrates, thanks to their flexible substrate specificity.

Further works will be focused on the in depth investigation of the relationships between CE1_ and CE2_Ubrb structure and functional flexibility. This will allow us to

estimate their potential for bioremediation of polyurethanes and carbamate insecticides, in terms of efficiency, stability, yield of recombinant expression, and toxicity/recovery facility/biodegradability and recyclability of the released products.

Acknowledgements

This research was funded by the French Ministry of Education and Research (Ministère de l'Enseignement supérieur et de la Recherche) and the INRA metaprogramme M2E (project Metascreen).

The high throughput screening work was carried out at the Laboratory for BioSystems & Process Engineering (Toulouse, France) with the equipments of the ICEO facility, dedicated to the screening and the discovery of new and original enzymes. ICEO is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund and the Institut National de la Recherche Agronomique (INRA).

Artwork

Figure 17: Functional profiles of the hit metagenomic clones, assayed on polysaccharides (dark blue), Tween 20 (green), Impranil (brown: minimal medium; orange: rich medium), pNPA (purple), pNPB (red), pNPP (bright blue).	149
Figure 18: Reaction media of 12F23 and 44I12 cytoplasmic extracts incubated with Impranil at 30°C during 48 h.....	151
Figure 19: HPSEC chromatogram of the reaction between Impranil and the clones of interest.	152
Figure 20: MALDI-TOF spectra of reaction media (A) before and (B) after 24h reaction with clone 44I12.....	153
Figure 21: HPLC analysis from the reaction of fenobucarb with 44I12 enzymatic extract at the beginning of the reaction (blue) and after 90 hours reaction (red)	154
Figure 22: Sequence alignment between CE1_Ubrb and enzymes from the lipolytic family V	161
Figure 23: Sequence alignment between CE2_Ubrb and enzymes from the lipolytic family IV	163

Table content

Table 7: Structures of Tween 20 and of the carbamates used in this study.....	150
Table 8: ORFs of interest from the sequenced clones active on Tween 20, Impranil and/or pNPA/B/P, annotated as putative esterases, ureases or proteases. Taxonomic annotation according to MEGAN analysis: number of ORFs assigned to F, Firmicutes; P, Proteobacteria; A, Actinobacteria; ESB, bacteria from environmental samples.	156

Supplementary information

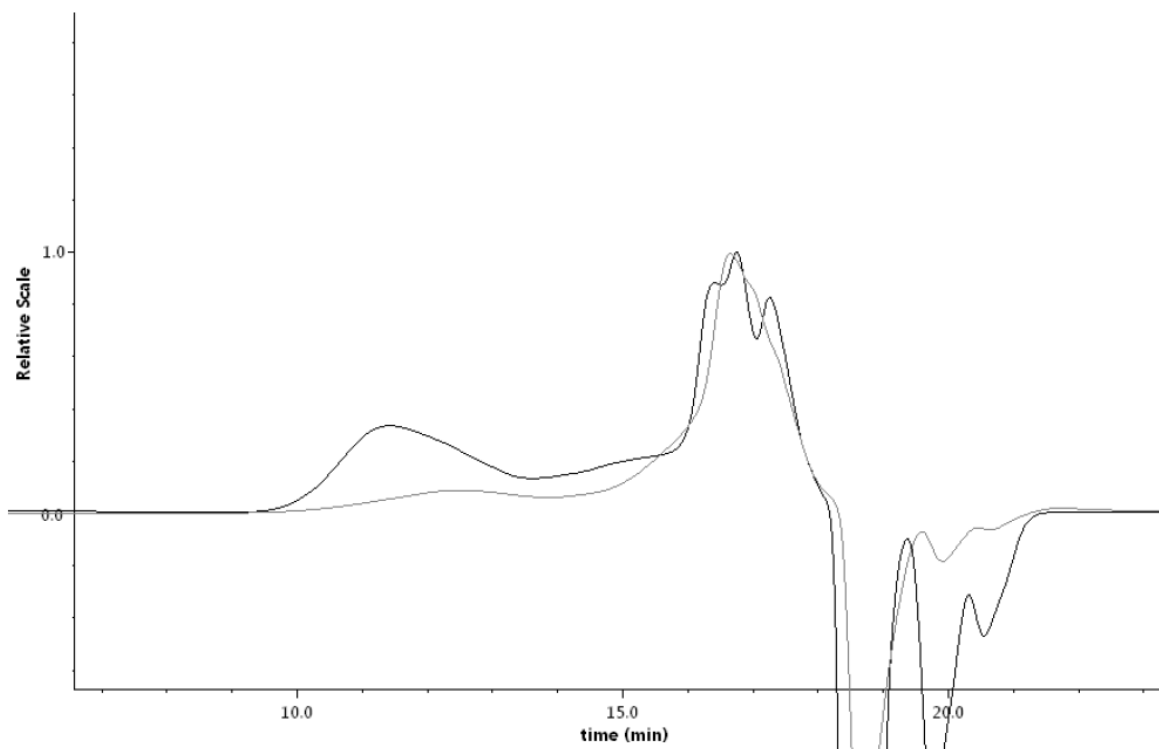


Figure S1: HPSEC chromatogram of the standard Impranil at 3 g/L. In black is the standard without lyophilisation, in grey the one after lyophilisation and resuspension in THF.

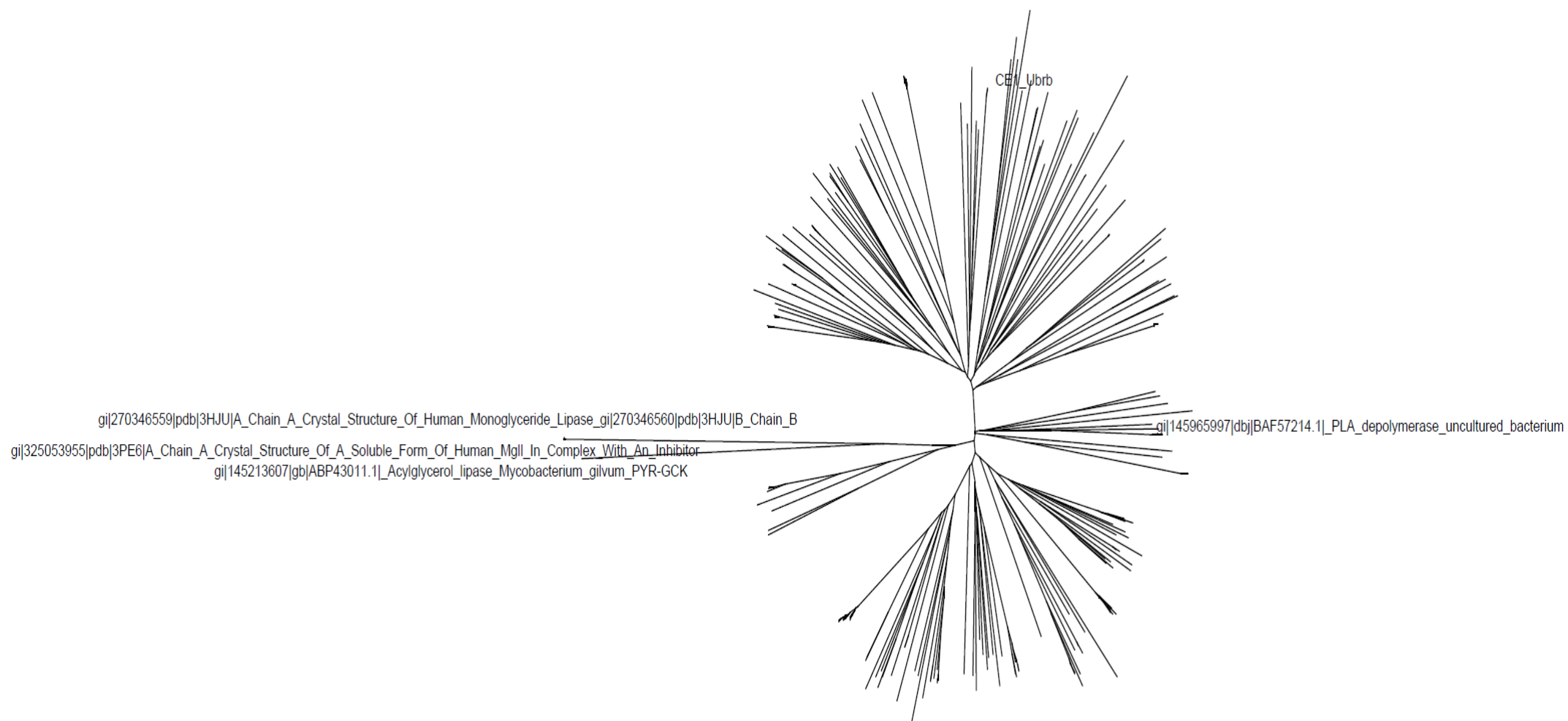


Figure S2: Phylogenetic tree of the 250 best blast hits against the nr database of NCBI of CE1_Ubrb, along with the blast hits biochemically characterized, with an e-value < 0.001.

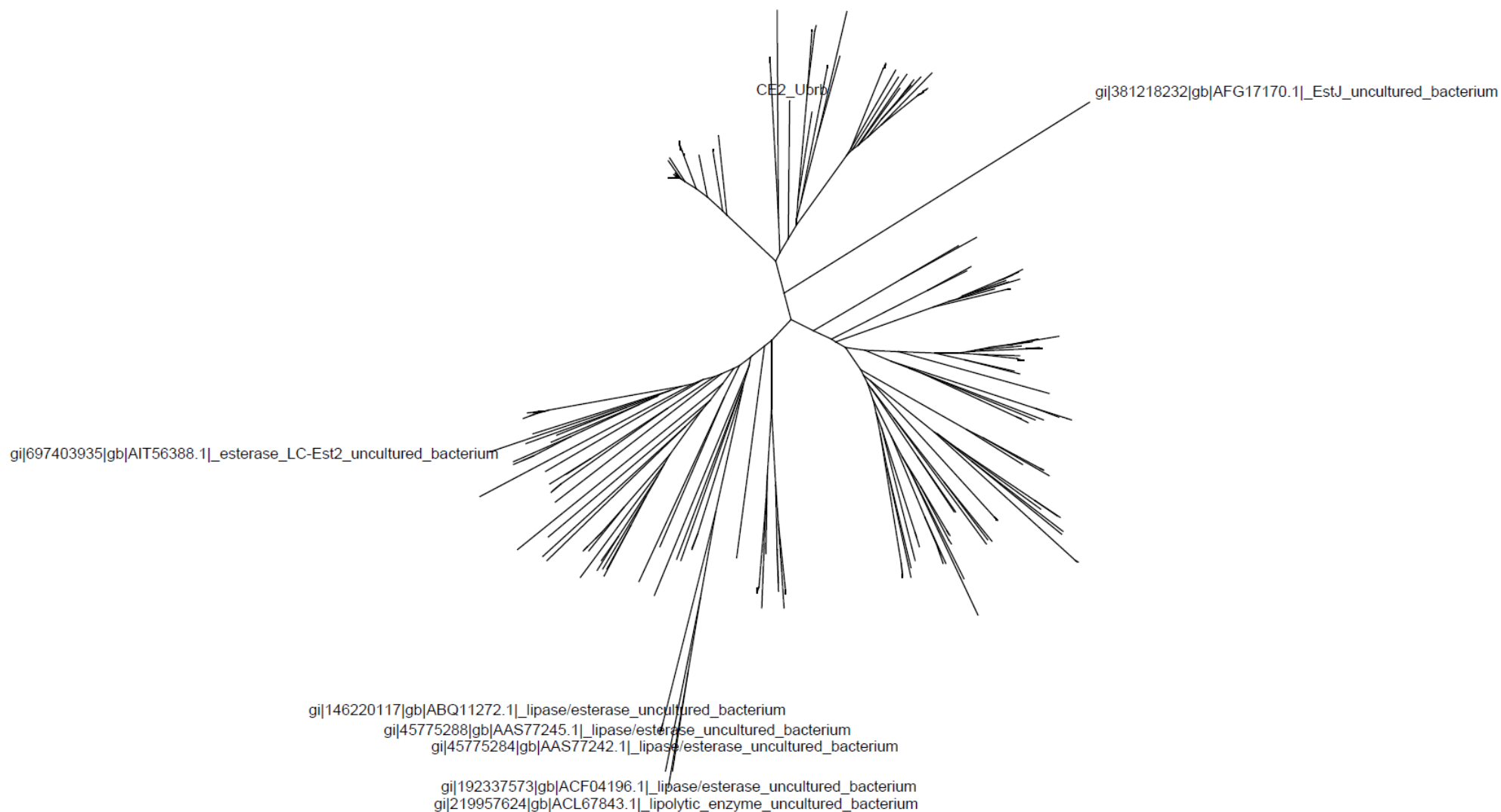


Figure S3: Phylogenetic tree of the 250 best blast hits against the nr database of NCBI of CE2_Ubrb, along with the blast hits biochemically characterized, with an e-value < 0.001.

References :

- Arpigny, J.L., Jaeger, K.E., 1999. Bacterial lipolytic enzymes: classification and properties. *Biochem. J.* 343 Pt 1, 177–183.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. doi:10.1186/1471-2164-9-75
- Bemis, J.C., Labash, C., Avlasevich, S.L., Carlson, K., Berg, A., Torous, D.K., Barragato, M., MacGregor, J.T., Dertinger, S.D., 2015. Rat Pig-a mutation assay responds to the genotoxic carcinogen ethyl carbamate but not the non-genotoxic carcinogen methyl carbamate. *Mutagenesis* 30, 343–347. doi:10.1093/mutage/geu084
- Ben Ali, Y., Verger, R., Abousalham, A., 2012. Lipases or Esterases: Does It Really Matter? Toward a New Bio-Physico-Chemical Classification, in: Sandoval, G. (Ed.), *Lipases and Phospholipases*. Humana Press, Totowa, NJ, pp. 31–51.
- Biffinger, J.C., Barlow, D.E., Cockrell, A.L., Cusick, K.D., Hervey, W.J., Fitzgerald, L.A., Nadeau, L.J., Hung, C.S., Crookes-Goodson, W.J., Russell, J.N., 2015. The applicability of Impranol®DLN for gauging the biodegradation of polyurethanes. *Polymer Degradation and Stability* 120, 178–185. doi:10.1016/j.polymdegradstab.2015.06.020
- Brulc, J.M., Antonopoulos, D.A., Berg Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., Edwards, R.E., Frank, E.D., Emerson, J.B., Wacklin, P., Coutinho, P.M., Henrissat, B., Nelson, K.E., White, B.A., 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences* 106, 1948–1953. doi:10.1073/pnas.0806191105
- Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré, J., Henrissat, B., Remaud-Siméon, M., Monsan, P., Potocki-Véronèse, G., 2013. Functional Metagenomics Reveals Novel Pathways of Prebiotic Breakdown by Human Gut Bacteria. *PLoS ONE* 8, e72766. doi:10.1371/journal.pone.0072766
- Chahinian, H., Ali, Y., Abousalham, A., Petry, S., Mandrich, L., Manco, G., Canaan, S., Sarda, L., 2005. Substrate specificity and kinetic properties of enzymes belonging to the hormone-sensitive lipase family: Comparison with non-lipolytic and lipolytic carboxylesterases. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1738, 29–36. doi:10.1016/j.bbalip.2005.11.003
- Cregut, M., Bedas, M., Durand, M.-J., Thouand, G., 2013. New insights into polyurethane biodegradation and realistic prospects for the development of a sustainable waste recycling process. *Biotechnology Advances* 31, 1634–1647. doi:10.1016/j.biotechadv.2013.08.011
- Doddamani, H.P., Ninnekar, H.Z., 2001. Biodegradation of Carbaryl by a *Micrococcus* Species. *Current Microbiology* 43, 69–73. doi:10.1007/s002840010262
- Faulds, C.B., 2010. What can feruloyl esterases do for us? *Phytochemistry Reviews* 9, 121–132. doi:10.1007/s11101-009-9156-2
- Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A.P.M.D., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N., Golyshin, P.N., 2005. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora: Enzymatic diversity from bovine rumen metagenome. *Environmental Microbiology* 7, 1996–2010. doi:10.1111/j.1462-2920.2005.00920.x
- Gabor, E.M., Alkema, W.B.L., Janssen, D.B., 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology* 6, 879 – 886. doi:10.1111/j.1462-2920.2004.00640.x
- Grube, A., Donaldson, D., Kiely, T., Wu, L., 2011. *Pesticides Industry Sales and Usage 2006 and 2007 Market Estimates*.

- Hausmann, S., Jaeger, K.-E., 2010. Lipolytic Enzymes from Bacteria, in: Timmis, K.N. (Ed.), Handbook of Hydrocarbon and Lipid Microbiology. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1099–1126.
- Hayatsu, M., Nagata, T., 1993. Purification and Characterization of Carbaryl Hydrolase from *Blastobacter* sp. Strain M501. *Appl. Environ. Microbiol.* 59, 2121–2125.
- Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z., Rubin, E.M., 2011. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* 331, 463–467. doi:10.1126/science.1200387
- Howard, G.T., 2002. Biodegradation of polyurethane: a review. *International Biodeterioration & Biodegradation* 49, 245–252. doi:10.1016/S0964-8305(02)00051-3
- Howard, G.T., Blake, R.C., 1998. Growth of *Pseudomonas fluorescens* on a polyester–polyurethane and the purification and characterization of a polyurethanase–protease enzyme. *International Biodeterioration & Biodegradation* 42, 213–220. doi:10.1016/S0964-8305(98)00051-1
- Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. *Genome Research* 17, 377–386. doi:10.1101/gr.5969107
- Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology* 61, 1061–1067. doi:10.1093/sysbio/sys062
- Hu, Y., Fu, C., Huang, Y., Yin, Y., Cheng, G., Lei, F., Lu, N., Li, J., Ashforth, E.J., Zhang, L., Zhu, B., 2010. Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. *FEMS Microbiology Ecology* 72, 228–237. doi:10.1111/j.1574-6941.2010.00851.x
- Jeon, J.H., Lee, H.S., Kim, J.T., Kim, S.-J., Choi, S.H., Kang, S.G., Lee, J.-H., 2012. Identification of a new subfamily of salt-tolerant esterases from a metagenomic library of tidal flat sediment. *Applied Microbiology and Biotechnology* 93, 623–631. doi:10.1007/s00253-011-3433-x
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi:10.1093/bioinformatics/btu031
- Kang, C.-H., Oh, K.-H., Lee, M.-H., Oh, T.-K., Kim, B., Yoon, J.-, 2011. A novel family VII esterase with industrial potential from compost metagenomic library. *Microbial Cell Factories* 10, 41. doi:10.1186/1475-2859-10-41
- Karns, J.S., Tomasek, P.H., 1991. Carbofuran hydrolase - purification and properties. *Journal of Agricultural and Food Chemistry* 39, 1004–1008. doi:10.1021/jf00005a041
- Kim, I., Kim, D.-U., Kim, N.-H., Ka, J.-O., 2014. Isolation and characterization of fenobucarb-degrading bacteria from rice paddy soils. *Biodegradation* 25, 383–394. doi:10.1007/s10532-013-9667-9
- Kim, Y.-J., Choi, G.-S., Kim, S.-B., Yoon, G.-S., Kim, Y.-S., Ryu, Y.-W., 2006. Screening and characterization of a novel esterase from a metagenomic library. *Protein Expression and Purification* 45, 315–323. doi:10.1016/j.pep.2005.06.008
- Kuhr, R.J., Dorough, H.W., 1976. Carbamate insecticides: chemistry, biochemistry, and toxicology. CRC Press, Cleveland.
- Kulkarni, S., Patil, S., Satpute, S., 2013. Microbial Esterases: An overview. *International Journal of Current Microbiology and Applied Sciences* 2, 135–146.
- Lam, K.N., Hall, M.W., Engel, K., Vey, G., Cheng, J., Neufeld, J.D., Charles, T.C., 2014. Evaluation of a Pooled Strategy for High-Throughput Sequencing of Cosmid Clones from Metagenomic Libraries. *PLoS ONE* 9, e98968. doi:10.1371/journal.pone.0098968
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., Chatonnet, A., 2013. ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res.* 41, D423–429. doi:10.1093/nar/gks1154
- Li, P.-Y., Ji, P., Li, C.-Y., Zhang, Y., Wang, G.-L., Zhang, X.-Y., Xie, B.-B., Qin, Q.-L., Chen, X.-L., Zhou, B.-C., Zhang, Y.-Z., 2014. Structural Basis for Dimerization and Catalysis of a Novel Esterase from

- the GTSAG Motif Subfamily of the Bacterial Hormone-sensitive Lipase Family. *Journal of Biological Chemistry* 289, 19031–19041. doi:10.1074/jbc.M114.574913
- Liu, K., Wang, J., Bu, D., Zhao, S., McSweeney, C., Yu, P., Li, D., 2009. Isolation and biochemical characterization of two lipases from a metagenomic library of China Holstein cow rumen. *Biochemical and Biophysical Research Communications* 385, 605–611. doi:10.1016/j.bbrc.2009.05.110
- Loredo-Treviño, A., Gutiérrez-Sánchez, G., Rodríguez-Herrera, R., Aguilar, C.N., 2012. Microbial Enzymes Involved in Polyurethane Biodegradation: A Review. *Journal of Polymers and the Environment* 20, 258–265. doi:10.1007/s10924-011-0390-5
- Lourenço, M., Ramos-Morales, E., Wallace, R.J., 2010. The role of microbes in rumen lipolysis and biohydrogenation and their manipulation. *animal* 4, 1008–1023. doi:10.1017/S175173111000042X
- Mayumi, D., Akutsu-Shigeno, Y., Uchiyama, H., Nomura, N., Nakajima-Kambe, T., 2008. Identification and characterization of novel poly(DL-lactic acid) depolymerases from metagenome. *Appl. Microbiol. Biotechnol.* 79, 743–750. doi:10.1007/s00253-008-1477-3
- McMahon, M.D., Guan, C., Handelsman, J., Thomas, M.G., 2012. Metagenomic Analysis of *Streptomyces lividans* Reveals Host-Dependent Functional Expression. *Applied and Environmental Microbiology* 78, 3622–3629. doi:10.1128/AEM.00044-12
- Mulbry, W.W., Eaton, R.W., 1991. Purification and characterization of the N-methylcarbamate hydrolase from *Pseudomonas* strain CRL-OK. *Appl. Environ. Microbiol.* 57, 3679–3682.
- Nakajima-Kambe, T., Onuma, F., Kimpara, N., Nakahara, T., 1995. Isolation and characterization of a bacterium which utilizes polyester polyurethane as a sole carbon and nitrogen source. *FEMS Microbiology Letters* 129, 39–42. doi:10.1016/0378-1097(95)00131-N
- Nakajima-Kambe, T., Shigeno-Akutsu, Y., Nomura, N., Onuma, F., Nakahara, T., 1999. Microbial degradation of polyurethane, polyester polyurethanes and polyether polyurethanes. *Applied Microbiology and Biotechnology* 51, 134–140. doi:10.1007/s002530051373
- Nunes, G., Barceló, D., 1999. Analysis of carbamate insecticides in foodstuffs using chromatography and immunoassay techniques. *TrAC Trends in Analytical Chemistry* 18, 99–107. doi:10.1016/S0165-9936(98)00076-4
- Okano, H., Hong, X., Kanaya, E., Angkawidjaja, C., Kanaya, S., 2015. Structural and biochemical characterization of a metagenome-derived esterase with a long N-terminal extension: Characterization of Metagenome-Derived Esterase. *Protein Science* 24, 93–104. doi:10.1002/pro.2591
- Peng, S., Yin, J., Liu, X., Jia, B., Chang, Z., Lu, H., Jiang, N., Chen, Q., 2015. First insights into the microbial diversity in the omasum and reticulum of bovine using Illumina sequencing. *Journal of Applied Genetics* 56, 393–401. doi:10.1007/s13353-014-0258-1
- Phua, S.K., Castillo, E., Anderson, J.M., Hiltner, A., 1987. Biodegradation of a polyurethane in vitro. *Journal of Biomedical Materials Research* 21, 231–246. doi:10.1002/jbm.820210207
- PlasticsEurope, 2015. *Plastics – the Facts 2014/2015 An analysis of European plastics production, demand and waste data.*
- Pohlentz, H.D., Boidol, W., Schüttke, I., Streber, W.R., 1992. Purification and properties of an *Arthrobacter oxydans* P52 carbamate hydrolase specific for the herbicide phenmedipham and nucleotide sequence of the corresponding gene. *J. Bacteriol.* 174, 6600–6607.
- Privé, F., Newbold, C.J., Kaderbhai, N.N., Girdwood, S.G., Golyshina, O.V., Golyshin, P.N., Scollan, N.D., Huws, S.A., 2015. Isolation and characterization of novel lipases/esterases from a bovine rumen metagenome. *Applied Microbiology and Biotechnology* 99, 5475–5485. doi:10.1007/s00253-014-6355-6
- Rahimi, A., Mashak, A., 2013. Review on rubbers in medicine: natural, silicone and polyurethane rubbers. *Plastics, Rubber and Composites* 42, 223–230. doi:10.1179/1743289811Y.0000000063

- Rey, M., Enjalbert, F., Monteils, V., 2012. Establishment of ruminal enzyme activities and fermentation capacity in dairy calves from birth through weaning. *Journal of Dairy Science* 95, 1500–1512. doi:10.3168/jds.2011-4902
- Russell, R.J., Scott, C., Jackson, C.J., Pandey, R., Pandey, G., Taylor, M.C., Coppin, C.W., Liu, J.-W., Oakeshott, J.G., 2011. The evolution of new enzyme function: lessons from xenobiotic metabolizing bacteria versus insecticide-resistant insects: Evolution of new enzyme function. *Evolutionary Applications* 4, 225–248. doi:10.1111/j.1752-4571.2010.00175.x
- Shingfield, K.J., Wallace, R.J., 2014. CHAPTER 1. Synthesis of Conjugated Linoleic Acid in Ruminants and Humans, in: Sels, B., Philippaerts, A. (Eds.), *RSC Catalysis Series*. Royal Society of Chemistry, Cambridge, pp. 1–65.
- Smith, M.J., Bucher, G., 2012. Tools to study the degradation and loss of the N-phenyl carbamate chlorpropham — A comprehensive review. *Environment International* 49, 38–50. doi:10.1016/j.envint.2012.08.005
- Sogorb, M.A., Vilanova, E., 2002. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. *Toxicol. Lett.* 128, 215–228.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi:10.1093/molbev/mst197
- Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., Dore, J., Monsan, P., Remaud-Simeon, M., Potocki-Veronese, G., 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* 20, 1605–1612. doi:10.1101/gr.108332.110
- Terrapon, N., Henrissat, B., 2014. How do gut microbes break down dietary fiber? *Trends in Biochemical Sciences* 39, 156–158. doi:10.1016/j.tibs.2014.02.005
- Tirawongsaroj, P., Sriprang, R., Harnpicharnchai, P., Thongaram, T., Champreda, V., Tanapongpipat, S., Pootanakit, K., Eurwilaichitr, L., 2008. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *Journal of Biotechnology* 133, 42–49. doi:10.1016/j.jbiotec.2007.08.046
- Ufarté, L., Laville, É., Duquesne, S., Potocki-Veronese, G., 2015. Metagenomics for the discovery of pollutant degrading enzymes. *Biotechnology Advances* 33, 1845–1854. doi:10.1016/j.biotechadv.2015.10.009
- University of Hertfordshire (2013), 2006. The Pesticide Properties DataBase (PPDB) developed by the Agriculture & Environment Research Unit (AERU) [WWW Document]. URL <http://sitem.herts.ac.uk/aeru/ppdb/en/index.htm> (accessed 9.16.15).
- Urgun-Demirtas, M., Singh, D., Pagilla, K., 2007. Laboratory investigation of biodegradability of a polyurethane foam under anaerobic conditions. *Polymer Degradation and Stability* 92, 1599–1610. doi:10.1016/j.polymdegradstab.2007.04.013
- U.S. National Library of Medicine, 2014. ChemIDplus Advanced - Chemical information with searchable synonyms, structures, and formulas [WWW Document]. URL <http://chem.sis.nlm.nih.gov/chemidplus/> (accessed 1.6.16).
- Wang, L., Hatem, A., Catalyurek, U.V., Morrison, M., Yu, Z., 2013. Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS ONE* 8, e78507. doi:10.1371/journal.pone.0078507
- Wang, Q., Lemley, A.T., 2003. Competitive Degradation and Detoxification of Carbamate Insecticides by Membrane Anodic Fenton Treatment. *Journal of Agricultural and Food Chemistry* 51, 5382–5390. doi:10.1021/jf034311f
- Watanabe, K., Nagao, N., Yamamoto, S., Toda, T., Kurosawa, N., 2007. *Thermobacillus composti* sp. nov., a moderately thermophilic bacterium isolated from a composting reactor. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* 57, 1473–1477. doi:10.1099/ijs.0.64672-0
- Zhu, W.-Y., Kingston-Smith, A.H., Troncoso, D., Merry, R.J., Davies, D.R., Pichard, G., Thomas, H., Theodorou, M.K., 1999. Evidence of a Role for Plant Proteases in the Degradation of Herbage

Proteins in the Rumen of Grazing Cattle. *Journal of Dairy Science* 82, 2651–2658.
doi:10.3168/jds.S0022-0302(99)75522-0

Chapter V:

Highly promiscuous oxidases discovered by functional exploration of the bovine rumen microbiome

Ufarté L.^{1,2,3}, Potocki-Veronese G.^{1,2,3}, Cecchini D.^{1,2,3}, Rizzo A.^{1,2,3},
Morgavi D.⁴, Cathala B.⁵, Moreau C.⁵, Cleret M.^{1,2,3}, Robe P.⁶, Klopp
C.⁷, Bozonnet S.^{1,2,3}, Laville E.^{1,2,3}

¹Université de Toulouse, INSA, UPS, INP, LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

²INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France

³CNRS, UMR5504, F-31400 Toulouse, France

⁴INRA, UMR 1213 Herbivores, 63122 Saint-Genès Champanelle, France

⁵UR1268 Biopolymères Interactions Assemblages, INRA, F-44316 Nantes, France

⁶LibraGen S.A., F-31400 Toulouse, France

⁷Plateforme Bio-informatique Toulouse Genopole, UBIA INRA, BP 52627, F-31326 Castanet-Tolosan Cedex, France

To be submitted (2016)

As shown in the previous chapter, enzymes naturally present in the ruminal ecosystem have the ability to degrade various structures of synthetic molecules, in particular pollutant ones.

To further explore the high functional diversity of the rumen ecosystem, and to increase the potential of the activity-based functional metagenomic approach for enzyme discovery, we decided to look for red-ox enzymes acting on polycyclic compounds. The bacterial diversity of such enzymes, especially those acting on lignin, a major component of plant cell wall, is at the moment largely underexplored. In particular, the existence and potential of bacterial ligninases in the rumen has never been studied to our knowledge, lignin degradation being attributed to fungal red-ox enzymes.

In addition to being key tools for biorefineries, the potential of such enzymes for bioremediation is also important, since they may be able to degrade different polyaromatic substrates issued, for example, from textile industries.

The following paper, which will be submitted later on, first describes the methodological developments we had to do in order to mine large metagenomic sequence space for oxidoreductases. This approach, applied to the bovine ruminal microbiome, allowed us to isolate original promiscuous red-ox enzymes, capable of acting on lignin derivatives and on synthetic dyes.

In addition of the team members, this part of the work was performed in collaboration with D. Morgavi (INRA Theix, rumen content sampling), P. Robe (Libragen SA, library construction), C. Klopp (INRA Toulouse, read assembly), B. Cathala, C. Moreau and N. Beury (INRA Nantes, construction of lignin films).

Abstract

The bovine rumen hosts a diverse microbiota, which is highly specialized in the degradation of lignocelluloses. Ruminal bacteria, in particular, are very well equipped to deconstruct plant cell wall polysaccharides. Nevertheless, their potential role in the breakdown of the lignin network has never been investigated. In the present study, we used functional metagenomics to discover novel bacterial red-ox enzymes acting on polyaromatic compounds. A novel methodology was developed to explore the potential of uncultured microbes for degradation of lignin derivatives, namely Kraft lignin and liginosulfonate. From a fosmid library covering 0.7 Gb of metagenomic DNA, three hit clones were discovered, producing enzymes that are able to oxidize a large diversity of polyaromatic compounds, without requiring addition of copper, manganese, or mediators. These highly promiscuous red-ox enzymes are thus highly interesting both for plant biomass biorefinery, and for dye bioremediation. They were issued from uncultured Clostridia, and belong to complex catabolic pathways involving proteins of different functional types, including hemicellulases, likely to work in synergy for substrate degradation.

1. Introduction

With the carbohydrates, the aromatic compounds are the most widely distributed class of organic compounds in natural environments. In nature, aromatic compounds are mainly present into lignin, aromatic amino acids and fossil hydrocarbons. There are polymers of ring shaped benzene molecules and because of inertness of C-H and C-C bonds, they are among the most stable and persistent molecules. Cleaning up contaminated soils, wastewater, groundwater and marine environments from these aromatic compounds is thus a challenging task. In addition, lignin being one of the major components of plant biomass, understanding how microbes degrade it is indispensable to green the industrial processes of biorefinery and for lignin conversion into biosourced synthons. Microbial degradation represents the major route for bioremediation of aromatic contaminated sites and lignin degradation. Fungi have long been known for their ability to initiate lignin depolymerisation (Kirk and Farrell, 1987). This capability has extended their use to a series of biotechnological applications, all of them related to the degradation of structurally diverse aromatic compounds (Field et al., 1993). In natural environments, bacteria are in contact with these compounds and have evolved to develop a wide diversity of aerobic and anaerobic catabolic strategies to degrade the huge range of organic compounds present in the ecosystems they colonize. Like fungi, many bacterial strains degrade aromatic compounds including lignin (Vicuña, 1988; Zimmermann, 1990; Masai et al., 2007).

However, compared to fungi, bacteria are much less well characterized with respect to their lignin metabolism. Even if the bacterial systems are reported to be less oxidatively powerful compared to fungal ones, Brown and Chang (2014) suggested that bacteria may be a rich source to recover new accessory enzymes that could act synergistically with the major oxidative enzymes to activate and uncap various sites of the substrates. This statement is based on the fact that bacteria metabolism interacts with lignin as a substrate for use as an aromatic carbon source rather than carry out its rapid degradation to access the sugar-based cellulose and hemicellulose substrates as is believed to be the case in many fungal systems.

According to environmental conditions, aromatic compounds can be degraded by different ways depending on the presence or absence of oxygen, given that many bacteria are not strict aerobics or anaerobics. In aerobic conditions, oxygen is the final electron acceptor and also a co-substrate for two key processes. Phenolic hydroxyl groups are introduced by oxygenases to activate the aromatic ring. The subsequent cleavage of aromatic carbon-carbon bonds also requires oxygen. The phenolic compounds generated by ring hydroxylation of aromatic hydrocarbons are then direct precursors for the oxygenolytic ring cleavage. These reactions are done by monooxygenases and dioxygenases (Dagley, 1971). In aerobic conditions, another class of enzymes, the laccases (*p*-diphenol: dioxygen oxidoreductases) are copper-containing oxidase enzymes acting on phenols and similar molecules by removing hydrogen from the OH group of the C molecules involved in the cycle and directly reducing dioxygen to water. Because of steric hindrance, enzymes might not come directly into contact with the polymers. Instead, small organic compounds or metals that can also be oxidized and activated by laccases, mediate the radical-catalysed depolymerization. Moreover, peroxydase act on aromatic compounds through the formation of radicals which participate in reactions where polymers and oligomers are produced. These last two classes of enzymes have been explored specifically for their involvement in the degradation of lignin, particularly heme peroxidases.

In addition to Mn- and Cu-containing oxidases, bacteria are also reported to produce enzymes that could break down lignin through hydroxylation or demethylation, such as cytochrome P450s, or non-heme iron enzymes (Brown and Chang, 2014). Other classes of proteins were found to be involved in aromatic compound metabolization. Strachan et al. (2014) identified in bacteria from coal beds, six functional classes of genes mediating lignin transformation clustered on the DNA strand. The genes were predicted to encode oxidoreductase activity (associated with lignin transformation), co-substrate generation (hydrogen peroxide formation), protein secretion (secretion apparatus or signal peptide), small molecule transport (multidrug efflux superfamily), motility (methyl accepting chemotaxis proteins), and signal transduction (PAS domain-containing sensors). Nevertheless, in bacteria, ligninase encoding genes do not seem to be associated in a same

genomic loci to genes encoding carbohydrate active enzymes active on plant cell wall polysaccharides.

In contrast, in anaerobic conditions, the aromatic ring is destabilized via a reductive attack and the ring is subsequently opened hydrolytically by a series of conventional hydration and dehydrogenation reactions (Harwood et al., 1998; Carmona and Díaz, 2005). The bacteria use various strategies based on reductive reactions including carboxylation of phenolic compounds, reductive elimination of ring substituents like hydroxyl or amino groups, oxidation of methyl substituents to carboxyl groups, O-demethylation reactions and shortening of aliphatic side chains (Harwood et al., 1998). The mineralization of aromatic compounds by facultative or obligate anaerobic bacteria is coupled to anaerobic respiration using locally available electron donors and acceptors like nitrate, sulfate, iron(III), manganese(II), selenite, ferredoxin, NAD(P)H. Aromatic compounds can participate in anaerobic metabolism by serving either as electron acceptors or electron donors (Carmona et al., 2009). The complete degradation of aromatic compounds in anaerobic bacteria requires the complementary activity of diverse enzymes (reductases, dehydrogenases, methyltransferases, decarboxylases, hydratases, phosphatases, hydrolases) whose genes are often grouped in clusters (Harwood et al., 1998; Carmona and Díaz, 2005). If lignin can be degraded only at a slow rate by anaerobic bacteria, they play a can major role in the mineralization of large amounts of lignin-derived low-molecular-weight compounds. Masai et al. (2007) indeed described catabolic pathways of lignin-derived aromatic compounds by the proteobacteria *Sphingomonas paucimobilis* of which more than 20 genes involved in the catabolism of lignin have been characterized (Li et al., 2009). Colberg and Young also confirmed that the intermediate metabolic products produced by aerobic degradation of lignin, were partially degraded by anaerobic bacteria (Colberg and Young, 1982).

The greatest advances in the identification of the genetic and molecular bases of aromatic compound catabolism have been made in recent decades thanks to the development of next generation sequencing and other 'omic' technologies. Most knowledge has been obtained by functional genomics of aromatic compound-degrading microorganisms or model microbial communities (Pieper et al., 2004). However, the quasi-totality of the microorganisms constituting the complex ecosystems that are the most powerful in breaking down aromatic compounds remains uncharacterised, due to our inability to isolate and culture them.

Functional metagenomics, which consists of assigning functions to proteins encoded by all genomes of a microbial community with no isolation and cultivation step, is a highly efficient way to boost the discovery of novel biocatalysts from the huge diversity of uncultured microbes. Diverse polluted environments, like, for example, activated sludge collected from wastewater treatment facilities belonging to the petroleum, coke and pharmaceutical industries (Suenaga et al., 2007; Sharma et al., 2012; Silva et al., 2013), or

contaminated soils (Ono et al., 2007; Brennerova et al., 2009; Lu et al., 2011) have been screened to retrieve new oxygenases for the degradation of aromatic compounds with potential applications in the development of bioremediation processes (for a review see Ufarté et al., 2015). But examples of ecosystems screened to find enzymes involved in the degradation of lignin or lignin-derived products are very scarce. A bacterial laccase acting on guaiacol, a product of lignin combustion, was first discovered by activity-based screening of mangrove soil metagenome (Ye et al., 2010). And more recently, a pseudo-laccase owing to the requirement of exogenous Cu(II) for oxidase activity was retrieved from a coal bed metagenome, screened for lignin catabolic activities (Strachan et al., 2014). This pioneering work, which included development of a very original screening method for polyaromatic degrading enzymes, highlighted the promiscuity of many biocatalysts towards this kind of substrates, and the metagenome flexibility in loci encoding polyaromatic degrading pathways. Lastly, let's mention the activity-based metagenomic study performed by Ferrer et al., who discovered from the bovine rumen metagenome a polyphenol oxidase with laccase activity, which was the first functionally characterised member of this new enzyme family (Ferrer et al., 2005; Beloqui et al., 2006). If this study proved that redox enzymes acting on polyaromatic compounds can be retrieved from uncultured ruminal bacteria, this enzyme and its homologs were nevertheless never tested on lignin or its derivatives. In addition, such enzymes are probably very rare in this ecosystem, since no laccase sequence could be found in the cow rumen metagenome (Ausec et al., 2011). The structural and functional diversity of bacterial ligninases is thus today still underexplored. This is especially the case for those acting in anaerobic or microaerobic conditions, which were never searched for by using activity-based approaches (Brown and Chang, 2014). The major bottleneck to their discovery and characterization is indeed the lack of experimental screens of red-ox enzymes allowing the exploration of sufficiently large sequence space to identify such rare enzymes.

In this study, we used an activity-based metagenomic approach to identify novel red-ox enzymes and metabolic pathways involved in aromatic compound catabolism, including lignin derivatives, by ruminal bacteria. In rumen, lignin is indeed highly present, under the form of dietary plant cell wall constituents. It is known to be partly degraded by anaerobic fungi. However the eventual role of ruminal bacteria in the metabolization of lignin-derived compounds remains to be elucidated (Ruiz-Dueñas and Martínez, 2009).

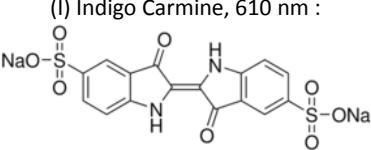
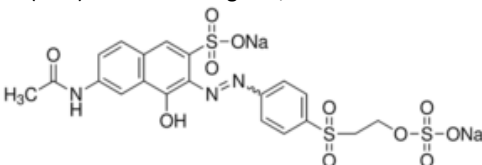
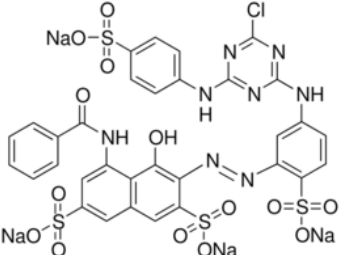
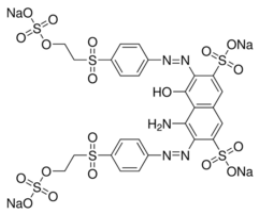
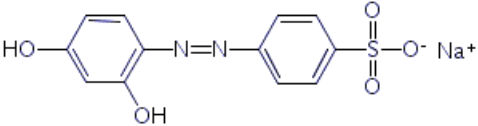
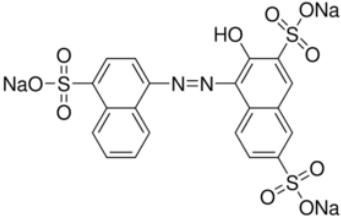
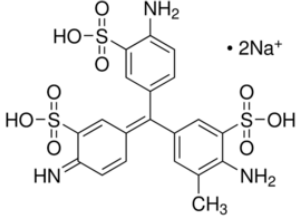
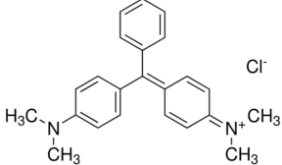
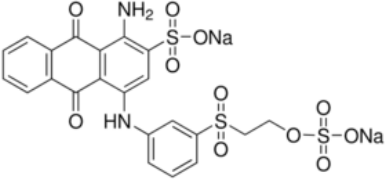
2. Material and methods:

2.1. Chemicals

2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS), 1-hydroxybenzotriazole (1-HBT), 2,2,6,6-Tetramethyl-1-piperidinyloxy, free radical (TEMPO), 3',5'-Dimethoxy-4'-hydroxyacetophenone (acetosyringone), 3,5-Dimethoxy-4-

hydroxybenzaldehyde (syngaldehyde) were purchased from Sigma-Aldrich (France). The dyes: Acid fuchsin (AF), amaranth (A), and tropaeolin O (TO) were purchased from Fisher Scientific (France) and the reactive orange 16 (RO16), indigo carmine (IC), malachite green (MG), cibracon brilliant red 3BA (CBR3BA), reactive black 5 (RB5), remazol brilliant blue R (RBBR) were purchased from Sigma-Aldrich (France). For dyes structure, see Table 9.

Table 9: Structures of the tested dyes (type I, indigo dye; AZO, azo dye; A, anthraquinonic dye; T, triarylmethane dye), and their maximal wavelength of absorbance.

<p>(I) Indigo Carmine, 610 nm :</p> 	<p>(AZO) Reactive Orange 16, 385 nm-495 nm :</p> 
<p>(AZO) Cibracon Brilliant Red 3BA, 525 nm :</p> 	<p>(AZO) Reactive Black 5, 600 nm :</p> 
<p>(AZO) Tropaeolin O, 390 nm :</p> 	<p>(AZO) Amaranth, 520 nm :</p> 
<p>(T) Acid fuchsin, 545 nm :</p> 	<p>(T) Malachite Green, 425 nm-620 nm</p> 
<p>(A) Remazol Brilliant Blue R, 595 nm :</p> 	

The MF resin (Madurit 75%, MW 112) was kindly supplied from INEOS Melamines GmbH (Frankfurt, Germany) and lignosulfonate comes from Tembec (Montréal, Canada). Xyloglucan (XG) from *Tamarindus indica* was provided by Megazyme (Bray, County Wicklow, Ireland) ($M_w = 202$ kD; viscosity: 6.5 dL/g; sugar composition: xylose = 38%, glucose = 42%, galactose = 16%, arabinose = 4%). Poly(allylamine hydrochloride) (PAH; $M_w=120,000-200,000$ g mol⁻¹) was provided by PolySciences (Germany).

2.2. Metagenomic DNA sampling and library construction

The two non-producing Holstein dairy cows used in this study were reared according to the national standards fixed by the legislation on animal care (Certificate of Authorization to Experiment on Living Animals, No. 004495, Ministry of Agricultures, France). In particular, the cows were fed a standard dairy cow ration composed of corn silage (64% DM), hay (6% DM) and concentrate (30% DM). They were fed *ad libitum* once a day in the morning. In order to enrich the rumen microbiome in lignolytic activities, before sampling the two cows received during 7 weeks a diet supplemented in wheat straw (a highly lignified plant material). The diet consisted of 80% wheat straw, 12% concentrate and 8% beetroot molasses. Rumen contents were taken from various parts of the rumen and manually homogenized. Both samples were mixed before metagenomics DNA extraction.

Metagenomic DNA was extracted, and fragments ranging in size from 30 to 40 kb DNA were isolated as previously described (Tasse et al., 2010) and cloned into pCC1FOS fosmid (Epicentre Technologies) as recommended by the manufacturer.. EPI100 *E. Coli* cells were then transfected to obtain a 20,000 clone library from the rumen samples. Recombinant clones were transferred to 384-well microtiter plates containing Luria Bertani medium, supplemented with 12.5 mg/L chloramphenicol and 8% glycerol. After 22 hours of growth at 37°C without any shaking, the plates were stored at -80°C.

All other culture media mentioned in this study contained 12.5 mg/L chloramphenicol.

2.3. Metagenomic library screening

The 19,968 clones of the library were gridded on 22 cm x 22 cm trays containing solid agar medium, using an automated microplate gridder (K2, KBiosystem, Basildon, UK). Agar minimum solid medium containing 1% filter sterilized lignin alkali (w/v) as sole carbon source was used. Minimal medium was composed of salts (3.6 g/L Na₂HPO₄, H₂O; 0.62 g/L KH₂PO₄; 0.11 g/L NaCl; 0.42 g/L NH₄Cl), 2 mM MgSO₄, 0.03 mM CaCl₂, other salts (15 mg/L Na₂EDTA, 2H₂O; 4.5 mg/L ZnSO₄, 7H₂O; 0.3 mg/L CoCl₂, 6H₂O; 1 mg/L MnCl₂, 4H₂O; 1 mg/L H₃BO₃; 0.4 mg/L Na₂MoO₄, 2H₂O; 3 mg/L FeSO₄, 7H₂O; 0.3 mg/L CuSO₄, 5H₂O), leucine 0.04 g/L and thiamine hypochloride 0.1 g/L. The assay plates were incubated at 37°C between 1 and 3 weeks, depending on the time necessary to visualize the growth of hit clones.

2.4. Functional characterization of the hit clones

The hit clones isolated from primary screening for their growth on lignin alkali as sole carbon source were then assayed to determine the optimal conditions of ABTS oxidation and were further examined for their activities on different aromatic compounds: (1) sulfonated lignin films, (2) mediators which are model substrates usually used to assess oxidoreductase activities and, (3) dyes. Due to the ability of *E. coli* to degrade some aromatic compounds (Diaz et al., 2001; Solís et al., 2012), an *E. coli* EPI100 clone containing the pCC1FOS fosmid without metagenomic DNA fragment was used as controle.

Each experiment was carried out at least twice and means were reported in figures and tables. Results were presented as percentages of the highest value taken as a reference.

2.4.1. Determination of optimal reaction conditions

The clones were grown at 37°C in 20 mL LB medium, with orbital shaking at 120 r.p.m.. After 16 h, cells were harvested by centrifugation for 5 min at 5,000 r.p.m., re-suspended and concentrated in activity buffer to obtain a final OD at 600 nm of 80. Cell lysis was done using sonication. Cell debris were centrifuged at 13,000 r.p.m. for 10 min and the cytoplasmic extracts were filtered with a 0.20 µm Minisart RC4 syringe filter. Enzymatic reactions were carried out in 96 wells microtiter plate assays using ABTS as reductive chromogenic substrate. Each well contained 20 µl of cytoplasmic extract and 5 mM ABTS. Optimal pH was determined at 30°C in 100 mM sodium citrate buffer for pH 4.0, 4.5, 5.0, 5.5 and 6.0. At optimal pH, optimal CuSO₄ concentration was determined at 30°C with 0, 0.1, 1 or 10 mM CuSO₄. Optimal temperature was then determined at optimal pH and CuSO₄ concentration by quantifying ABTS oxidation at 22, 30, 37, 40, 50 and 60°C, during 30 minutes and 17h, in order to assess enzyme stability. Peroxidase activity was also tested at optimal pH, temperature and CuSO₄ concentration, by adding 3% (v/v) of H₂O₂ to the reaction medium.

Enzymatic activity was determined by monitoring absorbance at 420 nm during 30 min, and at endpoint after 17 hours incubation (long-term reaction) using a microplate spectrophotometer (BioTek™ Eon™ Microplate Spectrophotometers, Colmar, France). The activity of ABTS oxidation was expressed in µmol/min/Lculture, using an extinction coefficient value of ε_{ABTS, 420 nm} = 36,000 M⁻¹.cm⁻¹.

2.4.2. Enzymatic activity on mediators

Optimal conditions determined after 17 hours of reaction on ABTS were used to measure in microtiter plates the activity of the three clones on different mediators. Final concentrations of substrates in 100 mM sodium citrate buffer were the following: 18 mM of 1-HBT, 2 mM of acetosyringone, TEMPO, and syringaldehyde. For easier comparison

between substrates, activity was expressed, after spectrophotometric observation at 408 nm, 528 nm, 300 nm and 370 nm respectively, as UOD/min/Lculture.

2.4.3. Degradation of semi-relective layers of sulfonated lignin

An anchoring solution of 4 g/L poly(allylamine hydrochloride) (PAH) was deposited on the silicium wafer, allowed to adsorb for 5 minutes and the support was spin-coated at 3,500 r.p.m. for 1 minute with an acceleration of 1,400 r.p.m./s². The mix resin MF (10% stock solution) along with lignosulfonate (100 g/L lignin stock solution, 75 g/L final) or Tamarind xyloglucan (20 g/L stock solution, 10 g/L final) was deposited on the wafer and spin-coated at 1,500 r.p.m. for 3 minutes with an acceleration of 100 r.p.m.. The films were put in an oven for 35 minutes at 135°C, then left to drop back to an ambient temperature. The films were washed in ultra-pure water for 2 hours to shake off unfixed residues. Two microliters of cellular extracts were deposited on the film, and incubated at optimal temperature for 3 hours in a saturated humidity atmosphere. Films were then washed with ultra-pure water and dried before use. Two microliters of cellular extracts were deposited on the films, and incubated at optimal temperature for 3 hours in a saturated humidity atmosphere. Films were then washed with ultra-pure water and dried before observation.

2.4.4. Dye discolouration assays

2.4.4.1. Selective growth on solid media with dyes as sole carbon source

The dye decolorizing activity of clones was first screened on agar plates containing agar with minimal medium (composition above), 1 mM CuSO₄, and 70 ppm of dyes as sole carbon source (except agar itself) in sodium citrate buffer (100 mM final concentration) at the optimal pHs determined for the 3 clones. The plates were inoculated with the hit clones as well as the *E. coli* Epi100 control and incubated at 37°C during 5 days.

2.4.4.2. Discolouration halos on solid media

On the plates containing agar (15 g/L), 10 mM CuSO₄, 70ppm dyes, buffered at optimal pHs in sodium citrate buffer (100 mM final concentration) were deposited 30 µl of cellular extracts, and incubated at 30°C, a temperature with all the clones active, and without potential problems with the medium, until a discolouration halo formed around positive extracts (1 to 2 days).

2.4.4.3. Discolouration of liquid reaction media

Optimal conditions determined for long-term reactions were used to quantify dye discolouration in liquid media. The experiment was carried out during 72 hours, in 96 well microtiter plates containing a final concentration of 70 ppm of dye dissolved of 100 mM

sodium citrate buffered at the optimal pHs, and 20 µl of cellular extract, for a final volume of 200 µL. The reaction media were centrifuged at 3,700 rpm for 10 minutes. An absorbance spectrum of each supernatant was registered using a microplate spectrophotometer (BioTek™ Eon™ Microplate Spectrophotometers, Colmar, France) at initial and final reaction times. The dye discoloration yield was calculated as follows by using the decrease of the absorbance at the wavelegenth of maximal absorbance determined for each dyebetween initial and final reaction times : discolouration (%) = $[(A_0 - A_{72}) / A_{72}] \times 100$, where, A_0 : initial absorbance of the dye/cellular extract at initial reaction time, A_{72} : absorbance of the dye/cellular extract after 72 hours of incubation (Phugare et al., 2011).

2.5. Metagenomic sequence Analysis

Fosmid DNA of the clone hits was extracted with the NucleoBond Xtra Midi kit from Macherey-Nagel (France) following supplier recommendation. Fosmids were then sequenced using the MiSeq System (Illumina), at the Genomic Platform GeT-PlaGe (Auzeville, France). Read assembly was performed using Masurca (<http://www.genome.umd.edu/masurca.html>). The contigs were cleaned from the vector pCC1FOS sequence using Crossmatch (<http://bozeman.mbt.washington.edu/phredphrapconsed.html>). Open reading frames (ORF) of at least 20 amino acids were predicted using MetaGene (Noguchi et al., 2006). Functions of the ORFs were inferred and manually annotated, based on BLASTX analysis against the NCBI non redundant and environmental databases (e-value < 10^{-8} , identity > 35%, and query length coverage $\geq 50\%$). ORFs were assigned to clusters of orthologous groups of proteins (COGs) using RPS-BLAST analysis against the COG database (e-values $\leq 10^{-8}$). A comparison was performed using BLASTP analysis against the Laccase and multicopper oxidase engineering database (LccED) (e-value = 1, identity $\geq 20\%$, and query length coverage $\geq 20\%$) (Sirim et al., 2011). The protein signatures was detected using the InterProScan software (Jones et al., 2014).

Contig taxonomic assignment was carried out using MEGAN v5.10.6 (Huson et al., 2011), based on BLASTX analysis against the non-redundant NCBI data base (min score = 35, min support = 1).

3. Results & discussion

3.1. Metagenomic library screening

The library consisted of 19,968 *Escherichia coli* fosmid clones, covering in total 0.7Gb of metagenomic DNA from rumen bacteria, each clone comprising a 30-40 kb DNA insert. The library was first screened for the ability to metabolize a depolymerised product of native lignin, namely lignin-alkali, used as sole carbon source for metagenomic clone growth. Lignin

alkali, or Kraft lignin, is the major by-product of alkaline sulfide treatment of lignocelluloses in the pulp and paper industry. Such minimal medium with Kraft lignin has already been used for the isolation of strains able to degrade this substrate (for example, Raj et al., 2007). But this kind of screen was never tested before to screen metagenomic clones, which contains each only a small fraction of the genome issued from the native bacterium, limiting the substrate harvesting and metabolic potential of the recombinant clones. However, here, this functional assay allowed the detection of 3 clones (of which one presented an orange phenotype) able to grow in mineral medium with Kraft lignin as a sole carbon source. Hit frequency was 0.015%. This is a value comparable to the hit yield obtained for oxidases screening (even not specific for the degradation of lignin-related products) in the rumen ecosystem (0.007 % in Beloqui et al., 2006), even being much lower than those obtained from environments contaminated with polyaromatic compounds, such as activated sludge (0.09% in Suenaga et al., 2007) and oil contaminated waters (0.2% in Vasconcellos et al., 2010 and 3% in Silva et al., 2013).

3.2. Red-ox activity characterisation

3.2.1. Determination of optimal reaction conditions

The enzymatic characterization of the metagenomic clones allowed to assess their enzyme stability as well as their versatility and efficiency towards structurally different types of substrates, i.e. mediators and dyes. All the assays were performed on cell extracts, which contained both *E. coli* and the recombinant enzymes, plus the molecules from the cell metabolism, like ions and cofactors. Therefore, the observed activities were interpreted as the result of synergic activities of these enzymes.

To characterize substrate specificity of the three clones, optimal conditions of activity were determined by monitoring the oxidation of ABTS at various pH and temperatures (Table 10).

Table 10: Optimal conditions for ABTS oxidation by the hit metagenomic clones.

Optimal Conditions	Clone 1	Clone 2	Clone 3
pH	4.5	5.0	4.5
CuSO ₄ concentration	10 mM	10 mM	10 mM
Temperature	60°C	50°C	60°C
Temperature for long-term reactions	50°C	30°C	50°C

All clones displayed optimal activity at acidic pH (optimal pH being 4.5 for clones 1 and 3, and 5.0 for clone 2) (**Figure 24**). The oxidative activity was totally lost for a pH value higher than 6.0 (clones 1 and 3) and even 5.5 for the clone 2.

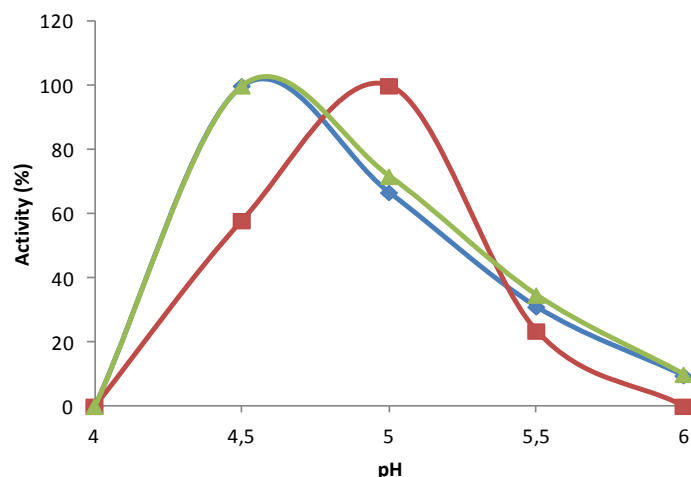


Figure 24: Determination of optimal pH of ABTS oxidation by the metagenomic clones at 30°C. Data represent means from duplicates, and are expressed in percent of the highest value for each clone. Blue: clone 1; red: clone 2; green: clone 3.

The optimal temperature for oxidative activity was determined at pH 4.5 for the clones 1, 3 and at the pH 5 for the clone 2 (Figure 25). The clones displayed enzymes functioning optimally at 60°C (clones 1 and 3) and 50°C (clone 2). However, after 30 minutes at the optimal temperatures, only 12.5, 10.8 and 31.0 % of activity remained respectively for clones 1, 2 and 3, indicating a moderate enzymatic extract thermal stability. Because the functional characterisation of clones on mediators and on dyes needed to be performed with longer incubation times, the optimal temperature was determined according to the stability of the extract for 17 hours of incubation. These temperatures were 50°C for clones 1 and 3, and 30°C for clone 2 (Figure 26). In order to keep the extract as unaltered as possible, it was decided to forgo adding protease inhibitors, which means that the decrease in stability might come from the degradation of the enzyme by the proteases present in the extract. Further analyses are underway to purify and characterise the enzyme of interest.

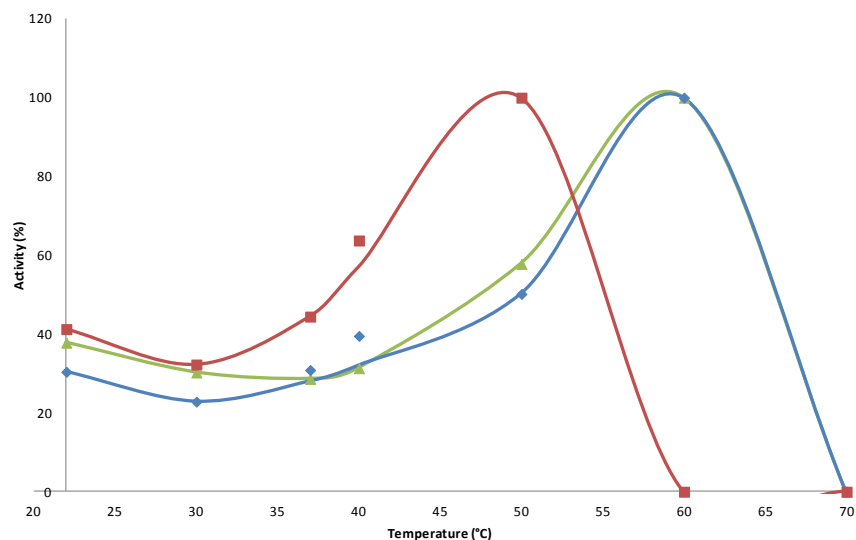


Figure 25: Determination of optimal temperature of ABTS oxidation by the metagenomic clones, at optimal pH. Activity was measured during 30 minutes. Data are expressed in percent of the highest value of each clone. Blue: clone 1; red: clone 2; green: clone 3.

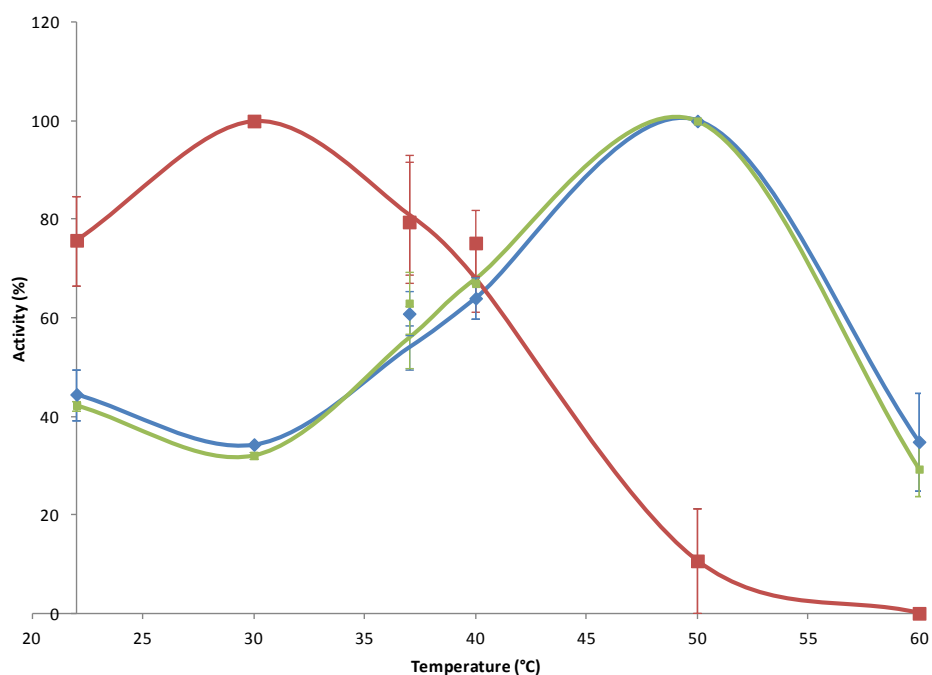


Figure 26 Determination of the optimal temperature of ABTS oxidation by the metagenomic clones at optimal pH, after 17 hours of incubation at each temperature. Data represent means from duplicates, and are expressed in percent of the highest value for each clone. Blue: clone 1; red: clone 2; green: clone 3.

To determine if oxidative enzymes might be laccases, metal ion effect on activity was assessed by adding copper to the reaction medium. Addition of copper was not required for

oxidative activity of clones 1 and 3, and had no significant effects on their reaction rate. In contrast, clone 2 had a low activity on ABTS in absence or with low quantities of Cu^{2+} . A summary of the optimal reaction conditions can be found in Table 10.

To determine if oxidative enzymes might be peroxidases, hydrogen peroxide was added to the reaction medium. No effect on the oxidative activity of the metagenomic clones was observed (data not shown).

3.2.2. Enzymatic activity on model substrates

Using optimal reaction conditions (Table 10), substrate specificity was characterized by measuring activity of the cell extracts on four others mediators of known structures and redox potential: TEMPO, HBT, acetosyringone and syringaldehyde. Syringaldehyde and acetosyringone are phenolic compounds, among the main products from the degradation of syringyl-rich lignins. They are characterized by the presence of two methoxy substituents, in ortho positions of the phenol, which lowers their redox potential. They have more stable radicals since the substitutions impose steric hindrance for the polymerization via radical fusion. The TEMPO molecule is a stable radical characterized by a nitroxyl group that benefits from the steric protection provided by the four methyl groups adjacent to the nitroxyl group N-O_2 . The methyl groups prevent a double bond occurring between either carbons adjacent to nitrogen. The stability of the radical is also indicated by the weakness of the O-H bond in the hydrogenated derivative TEMPO-H. ABTS contains two sulfonate groups that can be deprotonated. The HBT substrate is characterized by N-OH group of which enzymatic oxidation is mediated by the formation of highly active nitroxyl radical $>\text{N-O}^\bullet$ owing to the removal of an electron followed by release of a proton (Camarero et al., 2005; Morozova et al., 2007; Tavares et al., 2008; Torres-Duarte et al., 2009; Pardo et al., 2013).

Clone 1 was active on all the substrates and displayed the highest activity compared to the other two clones (Figure 27). The clone 2 was poorly active on syringaldehyde, acetosyringone and HBT, while the clone 2 was not clearly active on HBT. Globally the three clones had a high flexibility towards structurally different substrates, with a highest efficiency for clone 1 and 3.

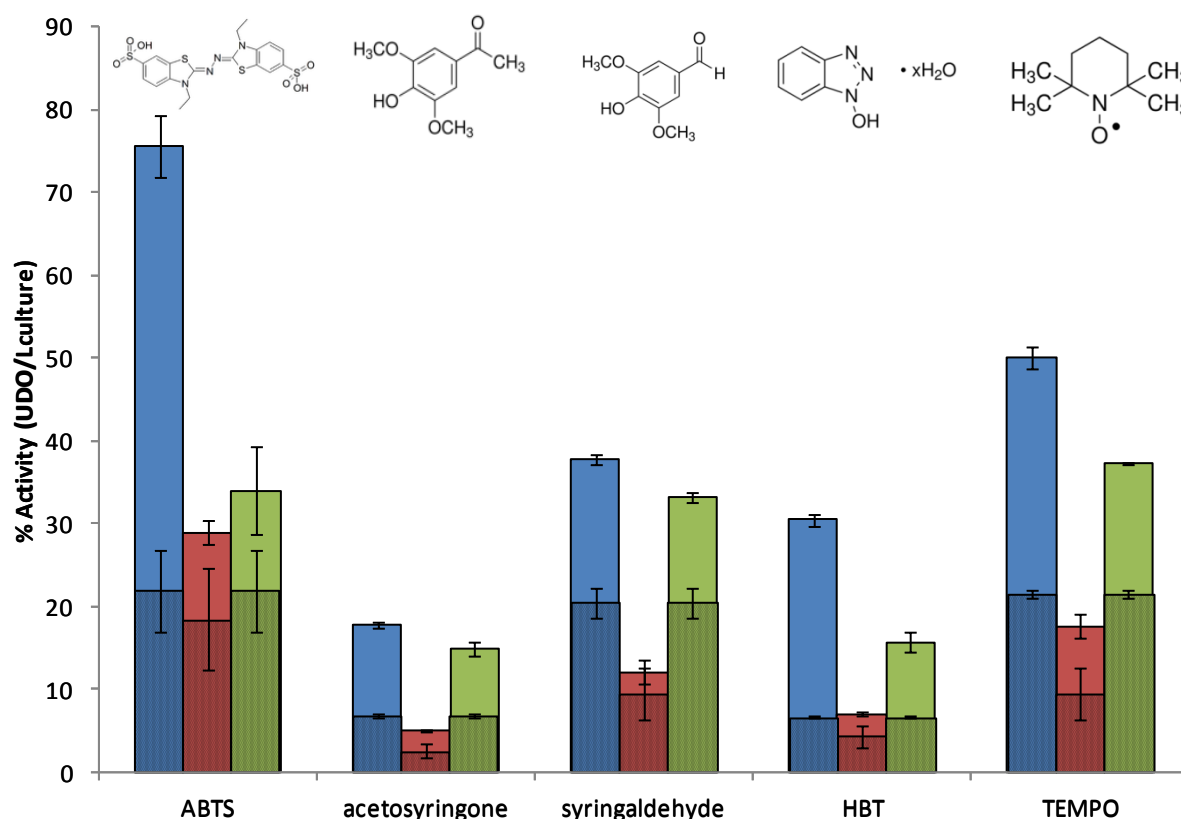


Figure 27: Effect of the mediator type on oxidative activity of the metagenomic clones after 17 hours of incubation, at optimal pH and temperature. Data are means from duplicates, and are expressed in percent of the highest value for each substrate. The part of the activity due to the *E. coli* enzymes is shown as hatching. Blue: clone 1; red: clone 2; green: clone 3.

3.2.3. Lignin-derivative depolymerization

The results of the primary screening indicated the presence of ruminal bacterial oxidoreductases, which may be involved in lignin degradation. Nevertheless, at this stage, we did not evidence their ability to break down polymeric lignin. We thus developed a depolymerization screening strategy, using coloured semi-reflective films of polymeric sulfonated lignin, which is a by-product of wood chemical pulping (Lebo et al., 2001). Like butterfly wings, the principle is based on modulation of the colour of the semi-reflective nano-layers of polymer, which depends on the film thickness. The incident light reaches an air-film interface, where part of it is reflected back while the rest is transmitted into the film. The second reflection occurs at the film-substrate interface. The net reflected light intensity depends on the combination of reflected light waves from both interfaces. This principle was previously exploited for detection of polysaccharide degrading enzymatic activities (Cerclier et al., 2011). Nevertheless, the construction of such semi-reflective nano-layers of soluble lignin derivatives, or of polysaccharide/lignin mixes mimicking simplified plant cell wall polymer assemblies, were never attempted. In the present study, cell extracts of the three hit clones were tested for their ability to degrade two types of films: one was composed of a

mix of resin and xyloglucan, the other contained mix of resin and sulfonated lignin. Because the resin used to polymerise the substrate onto the layers contains aromatic molecules likely to be attacked by oxidative enzymes, we used the xyloglycan layer as a control to indicate that the resin was not attacked. The results are presented into Figure 28.











	10 g/L Xyloglucan – 1% of resin	75 g/L sulfonated lignin – 1% of resin
Buffer pH 4.5		
Negative control		
Clone 1		
Clone 2		
Clone 3		

Figure 28 : Degradation of semi-reflective layers of xyloglucan and sulfonated lignin by the hit metagenomics clones.

When cell extracts were deposited on the sulfonated lignin/resin film, its colour slightly changed, for the 3 hit clones and the *E. coli* control, indicating that the films were a bit attacked. Clone 2 was the most effective to degrade the sulfonated lignin/resin layer. The control clone had a low effect, which may be due to the effect of the *E. coli* oxidative activity observed on ABTS. The control xyloglucan/resin layer was also at least slightly degraded by all clones, including the *E. coli* control. This indicates that the *E. coli* enzymes are responsible for the slight destructure of both the xyloglucan/resin and the sulfonated lignin/resin layers, through their action on the resin. The deposit mark was indeed not so visible when only reaction buffer was deposited on both films. In addition, the xyloglucan/resin layer was completely degraded by clone 1, indicating that it probably produces a xyloglucanase activity (which will be further discussed in the sequence analysis section).

This new screening approach thus allowed to demonstrate that the enzymes produced by clone 2 are able to depolymerize sulfonated lignin. Once this method will be optimized and automatized for production in micro-plate format with stable sulfonated lignin layers of homogeneous thickness, it will allow to screen large libraries for lignin-depolymerization activities, at a throughput of hundreds thousands assays per week. This should allow to massively increase the rate of discovery and engineering of microbial ligninases, issued from cultivated and not cultivated bacteria and fungi.

3.2.4. Application for dye elimination

Since the metagenomic clones displayed highly flexible specificity towards aromatic substrates, they were tested for their ability to degrade a panel of nine polycyclic aromatic dyes of diverse structures (Table 9), in order to evaluate the potential of the produced enzymes for industrial dye elimination. Five type of AZO dyes, one indigo, one anthraquinonic and two triphenylmethane dyes were tested. They all possess aromatic rings that may be fused, and diverse types of functional groups linked to the aromatic ring (-OH, -CH₃, -NCH₃, -SO₃H, -SO₃Na and -NH₂), which are reported to favour the extent of dye mineralization (Spadaro et al., 1992).

Three types of experiments were performed, i.e., assessment of the ability of metagenomic clones to metabolize the targeted dyes (by following their growth on minimal medium with the dye as sole carbon source), and to discolour dyes in solid and liquid media.

Only clone 3 was able to grow on selective medium with Tropaeolin O as sole carbon source and presented a degradation halo (Figure 29). Tropaeolin O being the less structurally complex dye molecule of our dye list, we suppose that the complementarity between *E. coli* and recombinant enzymes was sufficient to metabolize the substrate. On rich solid medium, clone 1 formed a 1 cm diameter discolouration halo for each dye after 2 days of incubation, while at the same time, clones 2 and 3 formed halos only on azo-dyes. Less visible small white halos were observed after 9 days on MG and RBBR dyes. The control clone also formed small and pale halos on three substrates: TO, CBR3BA and MG. However, these halos were always less visible than those obtained with other clones.

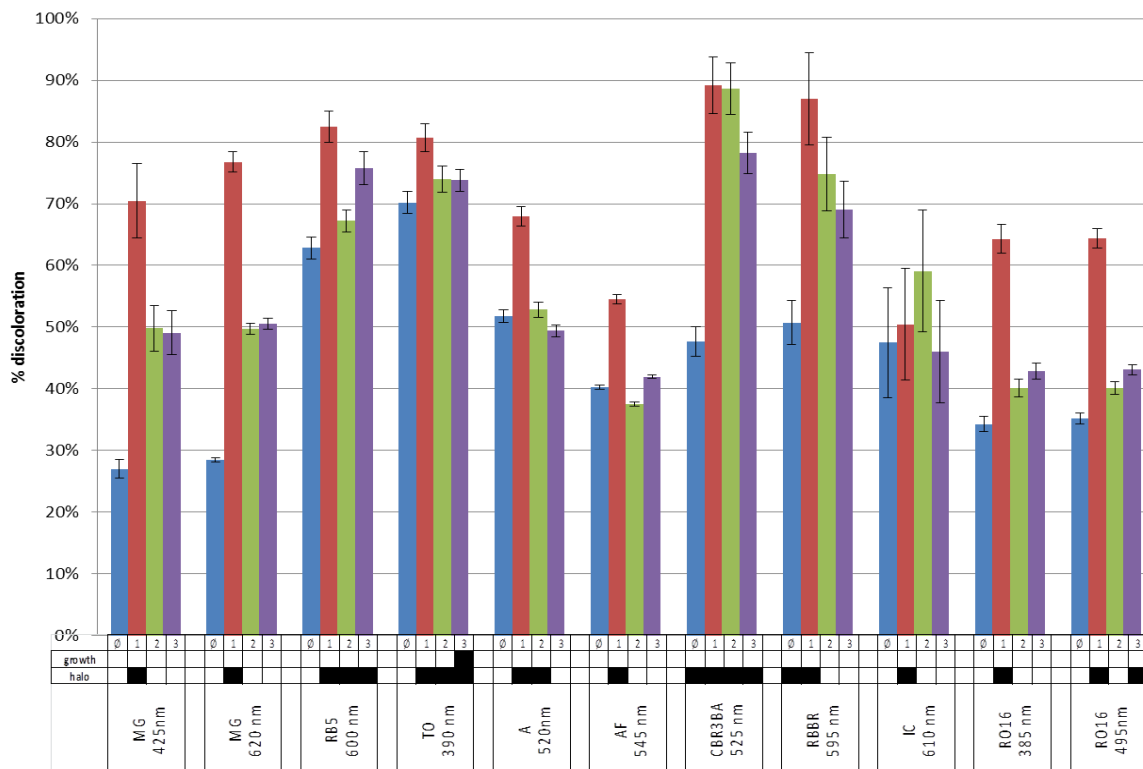


Figure 29 : Dye degradation by the hit clones, in solid and liquid media: quantification of the yields of liquid medium discoloration after reaction; Growth: dark square corresponds to the clone capable of growing and forming halos on solid minimum medium supplemented with dye; halo: dark cases correspond to the clones capable of forming halos on solid rich medium supplemented with dye. IC: Indigo Carmine, TO: Tropaeolin O, A: Amaranth, RO16: Reactive Orange 16, CBR3BA: Cibracon Brillant Red 3BA, RB5: Reactive Black 5, RBBR: Remazol Brillant Blue R, AF: Acid Fuschin, MG: Malachite Green

Cytoplasmic extracts of the three clones were then incubated during 72 hours in liquid media containing the dyes (Figure 29). After 72 hours of incubation in liquid medium, the formation of more or less significant solid precipitates at the bottom of wells, associated with a discolouration of the supernatant, was observed, even in wells corresponding to the control clone. After centrifugation, the spectral analysis of supernatants was performed before and after enzymatic reactions (Figure 30). On spectra collected at initial reaction times were determined the wavelength of maximal absorbance (λ_{max}) of each dye (Figure 29). The spectra from RO and MG dyes presented two λ_{max} values, both being reported in Figure 29. On the spectra corresponding to the final reaction time, we observed in any case a decrease of absorbance at λ_{max} (both values in the case of RO16 and MG dyes), which was more or less important according to the clones/dye couple. The differences of absorbance between initial and final reaction times were used to calculate the yield of dye discolouration, due to the removal of dyes from the aqueous phase by precipitation of the reaction products. Clone 1 was able to discolour all the dye solutions. Clone 2 discoloured most of the solutions except IC and AF, and clone 3 did not discolour IC, AF and A solutions.

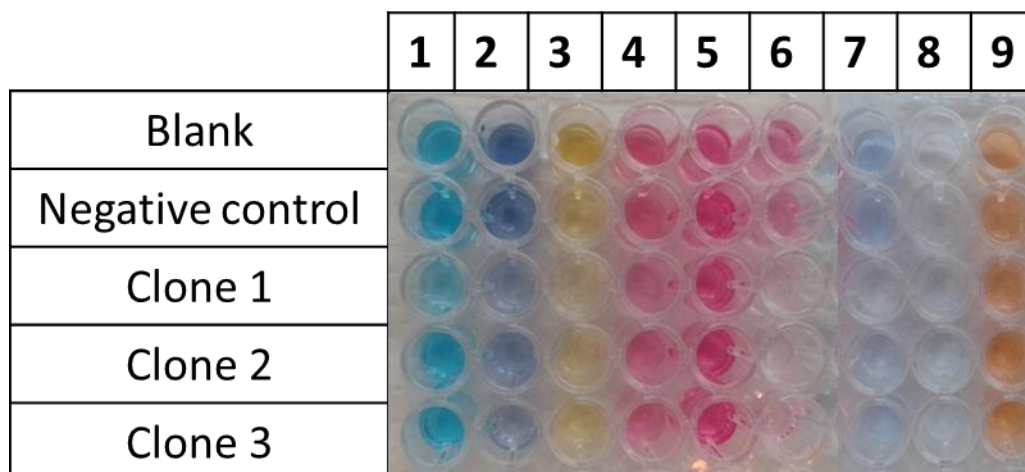


Figure 30: Dye discoloration by the hit metagenomic clones. The pictures show the supernatants of the reaction media after 72h of reaction with enzymatic extracts. 1: Malachite Green; 2: Reactive Black 5; 3: Tropaeolin O; 4: Amaranth; 5: Acid Fuchsin; 6: Cibracon Brilliant Red 3BA; 7: Remazol Brilliant Blue R; 8: Indigo Carmin; 9: Reactive Orange 16

Quantification of discolouration is very frequently performed in the studies dedicated to the research and assessment of biocatalysts for dye bioremediation. Most of the time, microbial consortia or individual strains are cultured in a medium containing the dye in solution and the disappearance of the colour is directly measured from these media using UV–Vis spectroscopy. In such conditions the authors did not mention precipitate formation. In many cases they observed a decrease of the absorbance at the dye λ_{max} and a concomitant increase of absorbance at a new λ_{max} corresponding to soluble degradation products which may also be toxic, even more so than the original dye. When the objective is to evaluate the potential of discolouration of a simple enzyme in a reaction mixture containing the dye in solution, the authors, like us, often report the formation of solid dye precipitates associated to the disappearance of the colour from the reaction medium. Several investigators have shown earlier that the aromatic compounds were either degraded or precipitated by the action of peroxidases/polyphenol oxidases. Some of them attributed the precipitation to phenoxy radicals formation followed by their spontaneous polymerization (Cooper and Nicell, 1996; Durán and Esposito, 2000; Mielgo et al., 2001; Mohan et al., 2005). It was the case in the study of Khan and Husain (2007), who monitored dye discolouration by potato and brinjal polyphenol oxidases. Dye treatment resulted in the formation of insoluble precipitates that the authors attributed to the formation of quinone-derivatives, which mediate the aggregation of aromatic pollutants. These precipitates can be easily removed from the reaction mixture by simple centrifugation, sedimentation or filtration.

Laccases also decolorize azo dyes through a highly nonspecific free radical mechanism forming phenolic compounds. Their relatively low substrate specificity is associated to the use of intermediate substrates (i.e. chemical mediators) assisting in the oxidation of different substrates by facilitating electron transfer from O₂ to laccase substrate (Forootanfar et al., 2012; Si et al., 2013). They have the advantage of not requiring H₂O₂ for oxidation reaction, which is an expensive co-substrate (and a potential pollutant) considered to be responsible for dye precipitation, possibly involving free-radical formation followed by polymerization of many aromatic compounds (Bhunja et al., 2001). In many studies involving laccases, no mention was made of such by-products except in the study of Zille et al. (2005), which reported the production by polymerization of a large amount of coupled products leading to a darkening of the solution. Despite their interest for such depollution processes, laccases present bottlenecks for biorefinery of plant lignocelluloses, since they are inhibited by copper chelation by lignin, and also because of their double ability to depolymerize and repolymerize lignin, blocking the access of cellulose and hemicellulose degrading enzymes to their substrates (Ruiz-Dueñas and Martínez, 2009).

In our assay conditions, the enzymes discovered from the present study do not need any mediator, nor addition of copper or H₂O₂ to be active beyond the usual quantity in the culture medium. This is thus of particular interest for their industrial use, both for biorefinery and bioremediation.

3.3. Sequence analysis

In order to identify the genes encoding the proteins responsible for oxidative activities, the three metagenomic DNA inserts were individually sequenced. The DNA from clones 1, 2 and 3 were each assembled into a single contig, their respective sizes being 40.6, 22.5 and 35 kb. The high sequencing depth (100X) allowed accurate gene prediction. The number of predicted genes was 28, 11 and 31 for clones 1, 2 and 3, respectively.

3.3.1. Functional annotation

Results of the BLASTP comparison with the NCBI_NR and Swissprot protein databases and of protein signature detection using InterProscan are given on Table S1. Mining our metagenomic sequences for laccase encoding genes by comparison with the LccED database did not give significant results. The genes responsible for clone growth using lignin alkali as sole carbon source thus do not belong to known laccase families. This result is in accordance with the fact that copper is not required for activity, except for the clone 2, whose activity increased with copper concentration.

Functional annotations allowed the designation of at least one gene for each clone that might be responsible for the detected activities.

In contig 1, ORF 15, annotated as D-3-phosphoglycerate dehydrogenase (PGDH), was the most probable target (Table S1). The best BLASTP hit with a proven activity was a distant D-lactate dehydrogenase from *Lactobacillus pentosus* (from the Firmicute phylum) (Taguchi and Ohta, 1991), which shares 29% identity on 65 % of the sequence length. Dehydrogenases are oxidoreductases catalysing a redox reaction in which a -OH group acts as a hydrogen or electron donor and reduces NAD^+ or NADP. The results of InterProScan confirms this annotation, showing a D-isomer specific 2-hydroxyacid dehydrogenase catalytic and NAD binding domain, along with an ACT domain which can be, for example, found in 3-phosphoglycerate dehydrogenases as a C-terminal regulatory domain. We found no report on dehydrogenase involvement in dye discolouration. However, some dehydrogenases, particularly cellobiose dehydrogenases (EC 1.1.99.18) belonging to auxiliary redox enzymes of the CAZy database, have been proposed to be involved in cellulose, hemicellulose and lignin degradation (Levasseur et al., 2013). In the present contig, the putative PGDH coding gene (ORF 15) is flanked with a gene encoding a putative phosphoserine aminotransferase (ORF 14), which are both involved in the L-serine biosynthesis pathway, previously described in Dey et al., (2005). In this pathway, the PGDH substrate, 3-D-phosphoglycerate, is an intermediate product of glycolysis, released through the Embden-Meyerhof pathway. These two genes are surrounded by other genes likely to constitute a functional cluster (ORFs 2, 3, 4, 8, 9, 10, 11, 12, 13) involved in the transport and degradation of polysaccharides from the plant cell wall (GH10 and GH94). Within the nearby genomic environment of PGDH, the gene coding for a methyltransferase (ORF 13) might be involved in oxidative processes through methyl group removal. Indeed, like the dehydrogenase activities, methyltransferase activities are reported to be involved in some of the lignin derivatives degradation steps in fungi and bacteria (Jeffers et al., 1997; Masai et al., 2007). Finally, CAZy mining in the clone 1 sequence confirmed the presence of genes encoding hemicellulases belonging to the GH10 and GH94 families, which, individually or synergistically, would be responsible for the degradation of the xyloglucan films.

In contig 2, the gene most probably responsible for the detected activities was ORF 3, in the sequence of which a cupin 2 domain was detected using InterProScan. The two known motifs were present: 43-G(ASVGL)H(T)H(QGDC)E(MLLVLS)G-62 and 79-G(QAHYC)P(Q)G(HT)H(TVM)N-94. No proximity of our cupin sequence to a characterised enzyme has been found with an e-value higher than 0.001 (Table S1). The cupin structure is described as a six-stranded β -barrel fold at the centre of which is a metal ion, most commonly ferrous iron (Fe(II)) but other forms are known, including copper or nickel. The cupin family is one of most functionally diverse of any described protein family. The largest subset of the cupin superfamily is the 2-oxyglutarate Fe^{2+} dependent dioxygenases. They use a single catalytic iron molecule to incorporate either one or both atoms of oxygen into a substrate. The metal ion is presumed to act as an electron conduit for single electron

transfer from the metal-bound substrate anion to O₂, resulting in activation of both substrates to radical species. During the reaction, a Fe(IV)-oxo (ferryl) intermediate is formed. This powerful oxidant is then used to carry out various reactions, including hydroxylation, halogenation, and demethylation (Dunwell et al., 2004). In hydroxylation reaction, an hydrogen atom is abstracted from the substrate, yielding a substrate radical and a ferrous hydroxide (Fe(III)-OH). This radical then couples to the hydroxide ligand, producing the hydroxylated product and the Fe(II) resting state of the enzyme. Ferrous hydroxide itself is practically white, but even traces of oxygen impart it with a greenish colour that can vary to reddish brown. This property may explain the orange phenotype of clone 2. The oxidative activities of this clone were enhanced in presence of copper, we thus may hypothesize that this cupin-folded protein used copper ions instead of/or in addition to iron. In its review, Fetzner (2012) stated that some cupins are cambialistic enzymes, capable of using several divalent metal ions as cofactors.

Cupins from the subfamily dioxygenase have shown ability to degrade aromatic substrates (Dunwell et al., 2004; Fetzner, 2012). However, we did not find any biochemical evidence of cupin involvement in lignin degradation and dye discolouration, except the report of Schluter et al.(2007). These authors indeed isolated a plasmid from bacteria residing in the activated sludge compartment of a wastewater treatment plant coding for a cupin 2 conserved barrel protein with an unknown function and a triphenylmethane reductase gene, which was shown to mediate discolouration of the triphenylmethane dyes.

In clone 2 sequence, the cupin encoding gene is flanked by a chelatase coding one (ORF 4). Chelatase function is insertion of divalent metal ions into tetrapyrroles to generate hemes, like cobalamin. This gene probably belongs to a cluster that continues, beyond the seven genes of the opposite strand, by five other genes (ORF 13 to 18). These last five include three genes (methionine synthase; methyltransferase, amidohydrolase) that are known to be involved in reactions using cobalamin and methylcobalamin as a cofactor in a variety of transmethylation and rearrangement reactions for amino acid transport and metabolism, especially methionine (Rodionov, 2003). The L-methionine is an essential amino acid and is required for a number of important cellular functions, including the aromatic compound methylation by methyltransferase (Jeffers et al., 1997; Rodionov, 2004; Masai et al., 2007). Here, we suggest that the ORFs 15, 16 and 17 might be involved in the degradation of Kraft lignin, sulfonated lignin and dyes through transmethylation reactions complementary to the dioxygenation ones catalysed by the cupin protein.

Contig 3 contains a cluster of genes with an organization previously described in catabolism of phenolic compounds in few facultative anaerobe bacteria (Carmona and Díaz, 2005). In these bacteria, degradation is due to the phenolic substrate reduction to a non-aromatic compound, carried out by a multimeric reductase. In contig 3, the reductase

encoding genes correspond to ORFs 6, 7, 8 and 9, annotated as the $\delta\alpha\beta\gamma$ subunits of a 2-oxoglutarate ferredoxin oxidoreductase. The β and δ subunits were in particular annotated as 4Fe4S ferredoxin binding sites, and the α and γ ones as the reductase module. The best BLASTP hit of these subunits with a proven activity is a distant 2-oxoisovalerate oxidoreductase (VOR, Heider et al., 1996; Tersteegen et al., 1997) (Table S1) and the predicted domains obtained with InterProScan imply the same with an added 4Fe4S ferredoxin iron-sulfur binding domain. The complex belongs to iron dependent metabolism where, in anoxic condition, it allows bacteria to gain energy through reduction/oxidation of iron minerals. This complex acts on aldehyde or oxo group of donors with ferredoxin as acceptor and a coenzyme A. The reaction that generates reduced ferredoxin was shown to accomplish the reduction of aromatic rings (Boll et al., 2002; Dorner and Boll, 2002). In oxic conditions, 2-oxoglutarate ferredoxin oxidoreductase is responsible for the oxidative decarboxylation of 2-oxo acids such as pyruvate and 2-oxoglutarate, to their acyl coenzyme A derivatives. According to the COGs database, two components of the oxidoreductase were annotated as encoding proteins involved in energy metabolism and electron transport. On the same strand in the contig 3, the oxidoreductase encoding gene was close to two genes encoding putative transporters (ORFs 2 and 3) specific to redox compounds i.e. riboflavin and iron. These genes, probably involved in the redox activity of clone 3, are located adjacent, in the opposite direction, to a cluster of 9 genes (ORFs 10 to 19), including the ferredoxin:NAD⁺ oxidoreductase (Rnf complex) corresponding to ORFs 14-19. A structurally equivalent complex was originally described in *Rhodobacter capsulatus* (Jouanneau et al., 1998), as the first enzyme in the mitochondrial respiratory chain. It is suggested to be a membrane bound electron/ion transport complex, which might have roles in nitrogen fixation. The complex contains a FeS centre, flavin- and NADH-binding sites. The enzyme catalyses the reversible electron transfer between NADH and ferredoxin using flavin directly as a redox cofactor (Biegel et al., 2011).

2-oxoglutarate ferredoxin oxidoreductase and ferredoxin:NAD⁺ oxidoreductase are both electron-generating enzymes for ring reduction (Ebenau-Jehle et al., 2003; Barragan et al., 2004). Both clusters of the contig might be involved in redox activities of the metagenomic clone observed on lignin and dyes. In addition, interestingly, the Rnf complex is concomitant to 4 genes (ORFs 11-13 and 25) related to cytochrome C, the terminal enzyme of the respiratory chain. In particular, ORF25 encode a putative cytochrome C peroxidase, a water-soluble heme-containing enzyme that reduces hydrogen peroxide to water, taking reducing equivalents from cytochrome C, though small molecules can also serve as substrate (Welinder, 1992). The cytochrome C association to Rnf complex and its co-transcription has previously been described in two methanogenic archaea *Methanosarcina acetivorans* and *Methanococcoides burtonii* (Biegel et al., 2011). Finally, another point that should be noted is that in the same metagenomic locus, 3 genes encode CAZymes, of which one is a putative plant cell wall polysaccharide degrading enzyme (CBM22-GH10).

3.3.2. Taxonomic assignation and sequence prevalence in the bovine ruminal microbiome

Metagenomic inserts of the three clones were impossible to assign accurately from a taxonomical point of view, as their sequences were too distant from any available sequenced genome. The Megan analysis using low stringent criteria revealed that the sequences were probably issued from Firmicute bacteria. Respectively, 6/28, 5/31 and 9/18 ORFs from clone 1, 3 and 2 sequences were indeed assigned to the Firmicute phylum. Let's note that the large majority of ORFs of the sequences of clone 1 and 3 (22 and 23, respectively) were assigned to bacteria from environmental samples without more taxonomic information, except that they were similar to sequences retrieved in cow rumen. Six out of 28 ORFs of contig1 and 17 out of 31 ORFs of contig3 were indeed similar to segments of the fosmid sequences from the ruminal metagenome of two Jersey cows evidenced by functional screening of polysaccharide degrading activity in the study of Wang et al. (2013). This result is explained by the strong enrichment of databases through massive metagenome sequencing projects of the last decades and the lack of knowledge of organisms that colonize these ecosystems. This observation indicate that clones 1 and 3 sequences probably come from prevalent ruminal bacteria composing the core ruminal microbiota evidenced in previous studies (Jami and Mizrahi, 2012; Wu et al., 2012). Otherwise, it is interesting to note that the gene clusters of contigs 1 and 2 were retrieved both from screening of polysaccharide degrading enzymes (Wang et al. (2013)) and in the present study, of red-ox enzymes acting on lignin derivatives. This constitutes the first evidence that bacterial ligninases and CAZymes are sometimes encoded by the same genomic loci, dedicated to degradation of plant cell wall.

4. Conclusion

In this study, functional metagenomics was used to discover new red-ox enzymes and metabolic pathways from the bovine rumen bacteriome, active on various aromatic substrates derived from chemical treatment of lignocelluloses in the pulp and paper industries, and from textile dyeing. None of these enzymes require addition of metals or mediators to the reaction media, conferring to them a particular interest compared to the laccases and peroxydases that are mostly used for biorefinery or dye bioremediation. Sequence analysis revealed the presence of several red-ox enzymes of different functional and structural families, which probably work in synergy to degrade and metabolize the targeted substrates. It also suggests that the identified red-ox enzymes produced by clones 1 and 3 would not require oxygen supplying their electron acceptors/ donors would rather be ions or NAD(P), for instance, that are present in the bacterial cytoplasm. Microbial processes

using recombinant bacteria producing the entire pathways discovered here would thus be more appropriate than enzyme-based processes, which would require supplementation in cofactors. Conversely, the cupin enzyme identified in clone 2 would require oxygen. This sequence could thus be issued from an aerobic or micro-aerobic bacteria, issued from soil or from dietary compounds, rather than from the native rumen itself. In addition, two of the genomic loci discovered here harbour genes encoding both red-ox enzymes capable of acting on lignin derivatives and hemicellulases, suggesting, contrarily to what was deduced until now from functional genomics studies that these bacterial enzymes could act synergistically to break down the plant cell wall network. Nevertheless, the exact identification of functions of the different enzymes encoded on these loci will require transcriptomic analysis and rational truncature of the fosmid inserts, and in structural characterization of the products released from a simple model substrate. Finally, their potential for biorefineries will have to be estimated by testing these clones on native lignin matrixes, and their mechanism on this kind of substrate elucidated.

Acknowledgments

We cordially thank Nadège Beury and Emilie Amblard for their technical assistance. This research was funded by the French Ministry of Education and Research (Ministère de l'Enseignement supérieur et de la Recherche), the INRA metaprogramme M2E (project Metascreen), and the French National Agency of Research (project REFLEX, grant number ANR 2011 NANO 007 03). The high throughput screening work was carried out at the Laboratory for BioSystems & Process Engineering (Toulouse, France) with the equipments of the ICEO facility, dedicated to the screening and the discovery of new and original enzymes. ICEO is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund and the Institut National de la Recherche Agronomique (INRA).

Artwork

Figure 24: Determination of optimal pH of ABTS oxidation by the metagenomic clones.at 30°C.....	192
Figure 25: Determination of optimal temperature of ABTS oxidation by the metagenomic clones, at optimal pH.	193
Figure 26 Determination of the optimal temperature of ABTS oxidation by the metagenomic clones at optimal pH, after 17 hours of incubation at each temperature..	193
Figure 27: Effect of the mediator type on oxidative activity of the metagenomic clones after 17 hours of incubation, at optimal pH and temperature.	195
Figure 28: Degradation of semi-reflective layers of xyloglucan and sulfonated lignin by the hit metagenomics clones.	196

Figure 29: Dye degradation by the hit clones, in solid and liquid media: quantification of the yields of liquid medium discoloration after reaction; 198
Figure 30: Dye discoloration by the hit metagenomic clones..... 199

Table content

Table 9: Structures of the tested dyes and their maximal wavelength of absorbance. 186
Table 10: Optimal conditions for ABTS oxidation by the hit metagenomic clones. 191

Supplementary data

Table S1: Functional and taxonomical assignation of the hit metagenomic clones. Column 1: Clone number and Megan-based taxonomic assignment by BLASTX analysis against the non redundant NCBI database. Contigs were assigned to a class, genus or species only if at least 50% of the ORFs were assigned to the same organism. Column 5: ORFs were inferred and manually annotated, based on BLASTX analysis against the NCBI non redundant and environmental database (E-value < 10^{-8} , identity > 35%, query length coverage $\geq 50\%$), and BLASTP analysis against the CAZy data base. In bold, proteins most probably involved in aromatic metabolisation Column 6: ORF assignment to Cluster of Orthologous Groups using RPS-BLAST analysis against the COG database (E-values $\leq 10^{-8}$).

Highly promiscuous oxidases discovered by functional exploration of the bovine rumen microbiome.

Clone number and taxonomic annotation	Contig size(bp)	ORF	Strand	ORF fonctional assignement	GOGs assignment	Best BLASTP characterized homolog (% coverage, e. value, % identity). E-value stringincy : > 0.001
clone 1 10N15 6/28 ORFS assigned to Firmicute	40,622 bp	1	+	Phenylalanyl-tRNA synthetase beta chain	J	
		2	+	N-Acetyl-D-glucosamine ABC transport system, sugar-binding protein	G	
		3	+	binding-protein-dependent transport systems inner membrane component	G	
		4	+	binding-protein-dependent transport systems inner membrane component	G	
		5	+	hypothetical protein	R	
		6	+	NHL repeat containing protein		
		7	+	hypothetical protein		
		8	+	sugar ABC transporter permease	G	
		9	+	putative sugar uptake ABC transporter permease protein		
		10	+	sugar ABC transporter substratebinding protein	G	
		11	+	Endo-1,4-beta-xylanase CBM22-GH10	G	
		12	+	Chitobiose phosphorylase GH94	G	
		13	+	methyltransferase YaeB family		
		14	+	Phosphoserine aminotransferase		
		15	+	D-3-phosphoglycerate dehydrogenase	H/E	D-lactate dehydrogenase / D-LDH / D-specific 2-hydroxyacid dehydrogenase [Lactobacillus pentosus] 65% ; 4e-23 ; 29%
		16	-	DNA-directed RNA polymerase beta' subunit		
		17	-	DNA-directed RNA polymerase beta subunit	F	
		18	+	DNA internalization-related competence protein ComEC/Rec2/		
		19	+	Firmicutes ribosomal L7Ae family protein		
		20	+	SSU ribosomal protein S12p (S23e)		
		21	+	SSU ribosomal protein S7p (S5e)		
		22	+	Translation elongation factor G		
		23	+	Translation elongation factor Tu		
		24	+	transcription antitermination protein nusG	J	
		25	+	Preprotein translocase subunit SecE		
		26	+	Transcription antitermination protein NusG		
		27	+	LSU ribosomal protein L11p (L12e)	K	
		28	+	LSU ribosomal protein L1p (L10Ae)	J	

Highly promiscuous oxidases discovered by functional exploration of the bovine rumen microbiome.

clone 2 31D12 9/18 ORFs assigned to Firmicute	23,332 bp	1	-	ABC transporter ATPbinding protein	E	
		2	+	hypothetical protein	na	
		3	+	cupin 2 conserved barrel domain protein	G/R	
		4	+	MG(2+) CHELATASE FAMILY PROTEIN / ComM-related protein	O	
		5	+	Basic prolinerich protein precursor	na	
		6	-	GNAT family acetyltransferase	J	
		7	-	prolipoprotein diacylglyceryl transferase	na	
		8	-	biosynthesis protein related to N-acetylglucosamine-1-phosphodiester alpha-N-ace	G/R	
		9	-	Ribosomal RNA small subunit methyltransferase C	J	
		10	-	Histone acetyltransferase HPA2 and related acetyltransferases	J/O	
		11	-	Beta-galactosidase GH2_6	G	
		12	-	Ribulose-phosphate 3-epimerase	G	
		13	+	multidrug ABC transporter ATPase	V	
		14	+	multidrug ABC transporter ATPase	V	
		15	+	vitamin B12 dependent methionine synthase activation domain protein	E	
		16	+	5-methyltetrahydrofolate--homocysteine methyltransferase	E	
		17	+	amidohydrolase	Q	
		18	+	Putative Holliday junction resolvase YggF	K	
		1	+	Endo-1,4-beta-glucanase CE2 fragt N-term		
		2	+	Substrate-specific component RibU of riboflavin ECF transporter		
		3	+	ferrous iron transporter A		
		4	-	aminoacyltRNA synthetase (ligase)		
		5	-	Phosphoglycolate phosphatase		
		6	+	2-oxoglutarate oxidoreductase, delta subunit, putative		2-oxoisovalerate oxidoreductase, subunit gamma [Methanothermobacter marburgensis str. Marburg] / ketoisovalerate oxidoreductase subunit vorC [Methanothermobacter marburgensis] / Ketoisovalerate oxidoreductase subunit VorC; VOR; / 2-oxoisovalerate ferredoxin reductase subunit gamma / 2-oxoisovalerate oxidoreductase gamma chain / 2-oxoisovalerate oxidoreductase, subunit gamma [Methanothermobacter marburgensis str. Marburg] 88% 2e-04 37%

Highly promiscuous oxidases discovered by functional exploration of the bovine rumen microbiome.

clone 3 32D12 5/31 ORFS assigned to Firmicute	35,084 bp	7	+	2-oxoglutarate oxidoreductase, alpha subunit	C	Ketoisovalerate oxidoreductase subunit VorB; VOR / 2-oxoisovalerate ferredoxin reductase subunit beta / 2-oxoisovalerate oxidoreductase beta chain [Methanothermobacter thermautotrophicus str. Delta H] 97% 3e-115 50%
		8	+	2-oxoglutarate oxidoreductase, beta subunit	C	Ketoisovalerate oxidoreductase subunit VorA; VOR / 2-oxoisovalerate oxidoreductase alpha chain / 2-oxoisovalerate-ferredoxin oxidoreductase subunit alpha [Methanothermobacter marburgensis str. Marburg] 95% 3e-84 54%
		9	+	2-oxoglutarate oxidoreductase, gamma subunit		Ketoisovalerate oxidoreductase subunit VorA; VOR / 2-oxoisovalerate oxidoreductase alpha chain / 2-oxoisovalerate-ferredoxin oxidoreductase subunit alpha [Methanothermobacter marburgensis str. Marburg] 97% 1e-31 42%
		10	-	acylCoA thioesterase (acyl hydrolase)	E	
		11	-	Cytochrome c-type biogenesis protein CcsA/ResC		
		12	-	Cytochrome c-type biogenesis protein Ccs1/ResB		
		13	-	Fumarate reductase flavoprotein subunit (EC 1.3.99.1):dehydrogenase		
		14	-	Electron transport complex protein RnfG		
		15	-	Electron transport complex protein RnfB		
		16	-	Electron transport complex protein RnfA		
		17	-	Electron transport complex protein RnfE		
		18	-	Electron transport complex protein RnfD	C	
		19	-	Electron transport complex protein RnfC	C	
		20	-	Sensory box/GGDEF family protein	C	
		21	+	Putative surface protein bspAlake		
		22	-	Lysophospholipase		
		23	+	hypothetical protein		
		24	+	Malate permease		
		25	-	cytochrome-c peroxidase		
		26	-	hypothetical protein		
		27	-	Amylopullulanase GH13		
		28	-	Beta-1,3-glucoyltransferase GT2		
		29	-	Endo-1,4-beta-xylanase CBM22-GH10		
		30	+	hypothetical protein		
		31	-	Collagen-binding A protein-like		

References

- Ausec, L., Zakrzewski, M., Goesmann, A., Schlüter, A., Mandic-Mulec, I., 2011. Bioinformatic analysis reveals high diversity of bacterial genes for laccase-like enzymes. *PLoS One* 6, e25724. doi:10.1371/journal.pone.0025724
- Barragan, M.J.L., Carmona, M., Zamarro, M.T., Thiele, B., Boll, M., Fuchs, G., Garcia, J.L., Diaz, E., 2004. The bzd Gene Cluster, Coding for Anaerobic Benzoate Catabolism, in *Azoarcus* sp. Strain CIB. *J. Bacteriol.* 186, 5762–5774. doi:10.1128/JB.186.17.5762-5774.2004
- Beloqui, A., Pita, M., Polaina, J., Martinez-Arias, A., Golyshina, O.V., Zumarraga, M., Yakimov, M.M., Garcia-Arellano, H., Alcalde, M., Fernandez, V.M., Elborough, K., Andreu, J.M., Ballesteros, A., Plou, F.J., Timmis, K.N., Ferrer, M., Golyshin, P.N., 2006. Novel Polyphenol Oxidase Mined from a Metagenome Expression Library of Bovine Rumen: BIOCHEMICAL PROPERTIES, STRUCTURAL ANALYSIS, AND PHYLOGENETIC RELATIONSHIPS. *J. Biol. Chem.* 281, 22933–22942. doi:10.1074/jbc.M600577200
- Bhunja, A., Durani, S., Wangikar, P.P., 2001. Horseradish peroxidase catalyzed degradation of industrially important dyes. *Biotechnol. Bioeng.* 72, 562–567. doi:10.1002/1097-0290(20010305)72:5<562::AID-BIT1020>3.0.CO;2-S
- Biegel, E., Schmidt, S., González, J.M., Müller, V., 2011. Biochemistry, evolution and physiological function of the Rnf complex, a novel iron-motive electron transport complex in prokaryotes. *Cell. Mol. Life Sci.* 68, 613–634. doi:10.1007/s00018-010-0555-8
- Boll, M., Fuchs, G., Heider, J., 2002. Anaerobic oxidation of aromatic compounds and hydrocarbons. *Curr. Opin. Chem. Biol.* 6, 604–611. doi:10.1016/S1367-5931(02)00375-7
- Brennerova, M.V., Josefiova, J., Brenner, V., Pieper, D.H., Junca, H., 2009. Metagenomics reveals diversity and abundance of meta -cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environ. Microbiol.* 11, 2216–2227. doi:10.1111/j.1462-2920.2009.01943.x
- Brown, M.E., Chang, M.C., 2014. Exploring bacterial lignin degradation. *Curr. Opin. Chem. Biol.* 19, 1–7. doi:10.1016/j.cbpa.2013.11.015
- Camarero, S., Ibarra, D., Martinez, M.J., Martinez, A.T., 2005. Lignin-Derived Compounds as Efficient Laccase Mediators for Decolorization of Different Types of Recalcitrant Dyes. *Appl. Environ. Microbiol.* 71, 1775–1784. doi:10.1128/AEM.71.4.1775-1784.2005
- Carmona, M., Díaz, E., 2005. Iron-reducing bacteria unravel novel strategies for the anaerobic catabolism of aromatic compounds: Anaerobic benzoate degradation. *Mol. Microbiol.* 58, 1210–1215. doi:10.1111/j.1365-2958.2005.04937.x
- Carmona, M., Zamarro, M.T., Blázquez, B., Durante-Rodríguez, G., Juárez, J.F., Valderrama, J.A., Barragán, M.J.L., García, J.L., Díaz, E., 2009. Anaerobic catabolism of aromatic

- compounds: a genetic and genomic view. *Microbiol. Mol. Biol. Rev.* MMBR 73, 71–133. doi:10.1128/MMBR.00021-08
- Cerclier, C., Guyomard-Lack, A., Moreau, C., Cousin, F., Beury, N., Bonnin, E., Jean, B., Cathala, B., 2011. Coloured Semi-reflective Thin Films for Biomass-hydrolyzing Enzyme Detection. *Adv. Mater.* n/a–n/a. doi:10.1002/adma.201101971
- Colberg, P.J., Young, L.Y., 1982. Biodegradation of lignin-derived molecules under anaerobic conditions. *Can. J. Microbiol.* 28, 886–889. doi:10.1139/m82-132
- Cooper, V.A., Nicell, J.A., 1996. Removal of phenols from a foundry wastewater using horseradish peroxidase. *Water Res.* 30, 954–964. doi:10.1016/0043-1354(95)00237-5
- Dagley, S., 1971. Catabolism of aromatic compounds by micro-organisms., in: *Advances in Microbial Physiology* APL. Academic Press, pp. 1–46.
- Diaz, E., Ferrandez, A., Prieto, M.A., Garcia, J.L., 2001. Biodegradation of Aromatic Compounds by *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 65, 523–569. doi:10.1128/MMBR.65.4.523-569.2001
- Dorner, E., Boll, M., 2002. Properties of 2-Oxoglutarate:Ferredoxin Oxidoreductase from *Thauera aromatica* and Its Role in Enzymatic Reduction of the Aromatic Ring. *J. Bacteriol.* 184, 3975–3983. doi:10.1128/JB.184.14.3975-3983.2002
- Dunwell, J.M., Purvis, A., Khuri, S., 2004. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 65, 7–17. doi:10.1016/j.phytochem.2003.08.016
- Durán, N., Esposito, E., 2000. Potential applications of oxidative enzymes and phenoloxidase-like compounds in wastewater and soil treatment: a review. *Appl. Catal. B Environ.* 28, 83–99. doi:10.1016/S0926-3373(00)00168-5
- Ebenau-Jehle, C., Boll, M., Fuchs, G., 2003. 2-Oxoglutarate:NADP⁺ Oxidoreductase in *Azoarcus evansii*: Properties and Function in Electron Transfer Reactions in Aromatic Ring Reduction. *J. Bacteriol.* 185, 6119–6129. doi:10.1128/JB.185.20.6119-6129.2003
- Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A.P.M.D., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N., Golyshin, P.N., 2005. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora: Enzymatic diversity from bovine rumen metagenome. *Environ. Microbiol.* 7, 1996–2010. doi:10.1111/j.1462-2920.2005.00920.x
- Fetzner, S., 2012. Ring-Cleaving Dioxygenases with a Cupin Fold. *Appl. Environ. Microbiol.* 78, 2505–2514. doi:10.1128/AEM.07651-11
- Field, J.A., de Jong, E., Feijoo-Costa, G., de Bont, J.A.M., 1993. Screening for ligninolytic fungi applicable to the biodegradation of xenobiotics. *Trends Biotechnol.* 11, 44–49. doi:10.1016/0167-7799(93)90121-O
- Forootanfar, H., Moezzi, A., Aghaie-Khozani, M., Mahmoudjanlou, Y., Ameri, A., Niknejad, F., Faramarzi, M.A., 2012. Synthetic dye decolorization by three sources of fungal laccase. *Iran. J. Environ. Health Sci. Eng.* 9, 27. doi:10.1186/1735-2746-9-27

- Harwood, C.S., Burchhardt, G., Herrmann, H., Fuchs, G., 1998. Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. *FEMS Microbiol. Rev.* 22, 439–458. doi:10.1111/j.1574-6976.1998.tb00380.x
- Heider, J., Mai, X., Adams, M.W., 1996. Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. *J. Bacteriol.* 178, 780–787.
- Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., Schuster, S.C., 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552–1560. doi:10.1101/gr.120618.111
- Jami, E., Mizrahi, I., 2012. Similarity of the ruminal bacteria across individual lactating cows. *Anaerobe* 18, 338–343. doi:10.1016/j.anaerobe.2012.04.003
- Jeffers, M.R., McRoberts, W.C., Harper, D.B., 1997. Identification of a phenolic 3-O-methyltransferase in the lignin-degrading fungus *Phanerochaete chrysosporium*. *Microbiology* 143, 1975–1981. doi:10.1099/00221287-143-6-1975
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi:10.1093/bioinformatics/btu031
- Jouanneau, Y., Jeong, H.S., Hugo, N., Meyer, C., Willison, J.C., 1998. Overexpression in *Escherichia coli* of the *rnf* genes from *Rhodobacter capsulatus*--characterization of two membrane-bound iron-sulfur proteins. *Eur. J. Biochem. FEBS* 251, 54–64.
- Khan, A.A., Husain, Q., 2007. Potential of plant polyphenol oxidases in the decolorization and removal of textile and non-textile dyes. *J. Environ. Sci. China* 19, 396–402. doi:10.1016/S1001-0742(07)60066-7
- Kirk, T.K., Farrell, R.L., 1987. Enzymatic “Combustion”: The Microbial Degradation of Lignin. *Annu. Rev. Microbiol.* 41, 465–501. doi:10.1146/annurev.mi.41.100187.002341
- Lebo, S.E., Gargulak, J.D., McNally, T.J., 2001. Lignin, in: John Wiley & Sons, Inc. (Ed.), *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., Henrissat, B., 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6, 41. doi:10.1186/1754-6834-6-41
- Li, J., Yuan, H., Yang, J., 2009. Bacteria and lignin degradation. *Front. Biol. China* 4, 29–38. doi:10.1007/s11515-008-0097-8
- Lu, Y., Yu, Y., Zhou, R., Sun, W., Dai, C., Wan, P., Zhang, L., Hao, D., Ren, H., 2011. Cloning and characterisation of a novel 2,4-dichlorophenol hydroxylase from a metagenomic library derived from polychlorinated biphenyl-contaminated soil. *Biotechnol. Lett.* 33, 1159–1167. doi:10.1007/s10529-011-0549-0

- Masai, E., Katayama, Y., Fukuda, M., 2007. Genetic and Biochemical Investigations on Bacterial Catabolic Pathways for Lignin-Derived Aromatic Compounds. *Biosci. Biotechnol. Biochem.* 71, 1–15. doi:10.1271/bbb.60437
- Mielgo, I., Moreira, M., Feijoo, G., Lema, J., 2001. A packed-bed fungal bioreactor for the continuous decolourisation of azo-dyes (Orange II). *J. Biotechnol.* 89, 99–106. doi:10.1016/S0168-1656(01)00319-4
- Mohan, S.V., Prasad, K.K., Rao, N.C., Sarma, P.N., 2005. Acid azo dye degradation by free and immobilized horseradish peroxidase (HRP) catalyzed process. *Chemosphere* 58, 1097–1105. doi:10.1016/j.chemosphere.2004.09.070
- Morozova, O.V., Shumakovich, G.P., Shleev, S.V., Yaropolov, Y.I., 2007. Laccase-mediator systems and their applications: A review. *Appl. Biochem. Microbiol.* 43, 523–535. doi:10.1134/S0003683807050055
- Noguchi, H., Park, J., Takagi, T., 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* 34, 5623–5630. doi:10.1093/nar/gkl723
- Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y., Tsuda, M., 2007. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl. Microbiol. Biotechnol.* 74, 501–510. doi:10.1007/s00253-006-0671-4
- Pardo, I., Chanagá, X., Vicente, A., Alcalde, M., Camarero, S., 2013. New colorimetric screening assays for the directed evolution of fungal laccases to improve the conversion of plant biomass. *BMC Biotechnol.* 13, 90. doi:10.1186/1472-6750-13-90
- Phugare, S.S., Kalyani, D.C., Surwase, S.N., Jadhav, J.P., 2011. Ecofriendly degradation, decolorization and detoxification of textile effluent by a developed bacterial consortium. *Ecotoxicol. Environ. Saf.* 74, 1288–1296. doi:10.1016/j.ecoenv.2011.03.003
- Pieper, D.H., Martins dos Santos, V.A., Golyshin, P.N., 2004. Genomic and mechanistic insights into the biodegradation of organic pollutants. *Curr. Opin. Biotechnol.* 15, 215–224. doi:10.1016/j.copbio.2004.03.008
- Raj, A., Reddy, M.M.K., Chandra, R., Purohit, H.J., Kapley, A., 2007. Biodegradation of kraft-lignin by *Bacillus* sp. isolated from sludge of pulp and paper mill. *Biodegradation* 18, 783–792. doi:10.1007/s10532-007-9107-9
- Rodionov, D.A., 2004. Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. *Nucleic Acids Res.* 32, 3340–3353. doi:10.1093/nar/gkh659
- Rodionov, D.A., 2003. Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes. *J. Biol. Chem.* 278, 41148–41159. doi:10.1074/jbc.M305837200
- Ruiz-Dueñas, F.J., Martínez, Á.T., 2009. Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.* 2, 164–177. doi:10.1111/j.1751-7915.2008.00078.x

- Schluter, A., Krahn, I., Kollin, F., Bonemann, G., Stiens, M., Szczepanowski, R., Schneiker, S., Puhler, A., 2007. IncP-1 Plasmid pGNB1 Isolated from a Bacterial Community from a Wastewater Treatment Plant Mediates Decolorization of Triphenylmethane Dyes. *Appl. Environ. Microbiol.* 73, 6345–6350. doi:10.1128/AEM.01177-07
- Sharma, N., Tanksale, H., Kapley, A., Purohit, H.J., 2012. Mining the metagenome of activated biomass of an industrial wastewater treatment plant by a novel method. *Indian J. Microbiol.* doi:10.1007/s12088-012-0263-1
- Si, J., Peng, F., Cui, B., 2013. Purification, biochemical characterization and dye decolorization capacity of an alkali-resistant and metal-tolerant laccase from *Trametes pubescens*. *Bioresour. Technol.* 128, 49–57. doi:10.1016/j.biortech.2012.10.085
- Silva, C.C., Hayden, H., Sawbridge, T., Mele, P., De Paula, S.O., Silva, L.C.F., Vidigal, P.M.P., Vicentini, R., Sousa, M.P., Torres, A.P.R., Santiago, V.M.J., Oliveira, V.M., 2013. Identification of Genes and Pathways Related to Phenol Degradation in Metagenomic Libraries from Petroleum Refinery Wastewater. *PLoS ONE* 8, e61811. doi:10.1371/journal.pone.0061811
- Sirim, D., Wagner, F., Wang, L., Schmid, R.D., Pleiss, J., 2011. The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases. *Database* 2011, bar006–bar006. doi:10.1093/database/bar006
- Solís, M., Solís, A., Pérez, H.I., Manjarrez, N., Flores, M., 2012. Microbial decolouration of azo dyes: A review. *Process Biochem.* 47, 1723–1748. doi:10.1016/j.procbio.2012.08.014
- Spadaro, J.T., Gold, M.H., Renganathan, V., 1992. Degradation of azo dyes by the lignin-degrading fungus *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 58, 2397–2401.
- Strachan, C.R., Singh, R., VanInsberghe, D., Ievdokymenko, K., Budwill, K., Mohn, W.W., Eltis, L.D., Hallam, S.J., 2014. Metagenomic scaffolds enable combinatorial lignin transformation. *Proc. Natl. Acad. Sci.* 111, 10143–10148. doi:10.1073/pnas.1401631111
- Suenaga, H., Ohnuki, T., Miyazaki, K., 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ. Microbiol.* 9, 2289–2297. doi:10.1111/j.1462-2920.2007.01342.x
- Taguchi, H., Ohta, T., 1991. D-lactate dehydrogenase is a member of the D-isomer-specific 2-hydroxyacid dehydrogenase family. Cloning, sequencing, and expression in *Escherichia coli* of the D-lactate dehydrogenase gene of *Lactobacillus plantarum*. *J. Biol. Chem.* 266, 12588–12594.
- Tavares, A.P., Cristóvão, R.O., Loureiro, J.M., Boaventura, R.A., Macedo, E.A., 2008. Optimisation of reactive textile dyes degradation by laccase-mediator system. *J. Chem. Technol. Biotechnol.* 83, 1609–1615. doi:10.1002/jctb.1952
- Tersteegen, A., Linder, D., Thauer, R.K., Hedderich, R., 1997. Structures and functions of four anabolic 2-oxoacid oxidoreductases in *Methanobacterium thermoautotrophicum*. *Eur. J. Biochem. FEBS* 244, 862–868.

- Torres-Duarte, C., Roman, R., Tinoco, R., Vazquez-Duhalt, R., 2009. Halogenated pesticide transformation by a laccase–mediator system. *Chemosphere* 77, 687–692. doi:10.1016/j.chemosphere.2009.07.039
- Ufarté, L., Laville, É., Duquesne, S., Potocki-Veronese, G., 2015. Metagenomics for the discovery of pollutant degrading enzymes. *Biotechnol. Adv.* 33, 1845–1854. doi:10.1016/j.biotechadv.2015.10.009
- Vasconcellos, S.P. de, Angolini, C.F.F., García, I.N.S., Martins Dellagnezze, B., Silva, C.C. da, Marsaioli, A.J., Neto, E.V. dos S., de Oliveira, V.M., 2010. Screening for hydrocarbon biodegraders in a metagenomic clone library derived from Brazilian petroleum reservoirs. *Org. Geochem.* 41, 675–681. doi:10.1016/j.orggeochem.2010.03.014
- Vicuña, R., 1988. Bacterial degradation of lignin. *Enzyme Microb. Technol.* 10, 646–655. doi:10.1016/0141-0229(88)90055-5
- Wang, L., Hatem, A., Catalyurek, U.V., Morrison, M., Yu, Z., 2013. Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PloS One* 8, e78507. doi:10.1371/journal.pone.0078507
- Welinder, K.G., 1992. Superfamily of plant, fungal and bacterial peroxidases. *Curr. Opin. Struct. Biol.* 2, 388–393. doi:10.1016/0959-440X(92)90230-5
- Wu, S., Baldwin, R.L., Li, W., Li, C., Connor, E.E., Li, R.W., 2012. The Bacterial Community Composition of the Bovine Rumen Detected Using Pyrosequencing of 16S rRNA Genes. *Metagenomics* 1, 1–11. doi:10.4303/mg/235571
- Ye, M., Li, G., Liang, W.Q., Liu, Y.H., 2010. Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. *Appl. Microbiol. Biotechnol.* 87, 1023–1031. doi:10.1007/s00253-010-2507-5
- Zille, A., Gornacka, B., Rehorek, A., Cavaco-Paulo, A., 2005. Degradation of Azo Dyes by *Trametes villosa* Laccase over Long Periods of Oxidative Conditions. *Appl. Environ. Microbiol.* 71, 6711–6718. doi:10.1128/AEM.71.11.6711-6718.2005
- Zimmermann, W., 1990. Degradation of lignin by bacteria. *J. Biotechnol.* 13, 119–130. doi:10.1016/0168-1656(90)90098-V

Chapter VI:

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

Ufarté L.^{1,2,3}, Laville E.^{1,2,3}, Cecchini D.^{1,2,3}, Rizzo A.^{1,2,3}, Amblard E.^{1,2,3},
Drula E.^{4,5,6,7,8}, Henrissat B.^{7,8,9}, Cleret M.^{1,2,3}, Lazuka A.^{1,2,3},
Hernandez G.^{1,2,3}, Morgavi D.¹⁰, Dumon C.^{1,2,3}, Robe P.¹¹, Klopp C.,
Bozonnet S.^{1,2,3}, Potocki-Veronese G.^{1,2,3}

¹Université de Toulouse, INSA, UPS, INP, LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

²INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France

³CNRS, UMR5504, F-31400 Toulouse, France

⁴Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique et Aix-Marseille Université 13288 Marseille cedex 9, France.

⁵INRA, UMR1163 BBF, Polytech'Marseille

⁶Marseille Université, UMR1163 BBF, Polytech'Marseille

⁷CNRS, UMR7257 Architecture et Fonction des Macromolécules Biologiques

⁸INRA, USC1408 Architecture et Fonction des Macromolécules Biologiques

⁹Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

¹⁰INRA, UMR 1213 Herbivores, 63122 Saint-Genès Champanelle, France

¹¹LibraGen S.A., F-31400 Toulouse, France

¹²Plateforme Bio-informatique Toulouse Genopole, UBIA INRA, BP 52627, F-31326 Castanet-Tolosan Cedex, France

To be submitted (2016)

To better understand the natural functioning of the ruminal ecosystem, especially catabolism of dietary fiber, and to obtain highly efficient lignocellulolytic consortia, we developed an original approach of bacteriome enrichment on wheat straw, both *in vivo* and in controlled reactors. After enrichment, the consortia were then used to construct two metagenomic libraries that were screened for glycoside-hydrolases, carbohydrate esterases and redox-enzymes that act synergistically to break down the plant cell wall network.

As such, the next paper, which will be submitted later on, gives a broader description of the functional capacity of the bacterial fraction of the cow rumen ecosystem, using activity-based metagenomics. For the first time, metagenomic clones were characterized on a real, highly lignified, industrial substrate, namely wheat straw. Their functional profiling allowed us to select some of the hit clones to design lignocellulolytic cocktails, and to study the effects of supplementation by esterases and red-ox enzymes on straw degradation efficiency. Sequence analysis revealed numerous novel CAZy module families, and metagenomic loci encoding batteries of synergistic lignocellulases, mostly issued from Firmicutes and unknown bacterial genera. Further data integration will be performed before submission of this article, in particular to discuss the taxonomic shift observed between the two enriched microbiomes and the native one, and to analyse the prevalence and abundance of the genes identified here in the bovine ruminal metagenome.

In addition of the team members, this part of the work was performed in collaboration with D. Morgavi (INRA Theix, *in vivo* enrichment and rumen content sampling), G. Hernandez-Raquet and A. Lazuka (LISBP, bioreactor enrichment), P. Robe (Libragen SA, library construction), C. Klopp (INRA Toulouse, read assembly), B. Henrissat and E. Drula (AFMB Marseilles, CAZy annotation).

Abstract

To face the structural complexity of plant cell wall, which compose their main carbon source in the bovine ruminal ecosystem, bacteria have evolved sophisticated multi-functional enzyme complexes. In order to unlock them from the vast world of uncultured bacteria, a high-throughput activity-based metagenomic approach was developed. In particular, a multi-step screening methodology was set-up to highlight metagenomic clones acting on polysaccharides and polyaromatic compounds, and to sort out those that are the most efficient to break down a highly lignified industrial substrate, wheat straw. This approach was used to explore the functional potential of two different bovine ruminal microbiomes, totalizing 1.5 Gb, issued from *in vivo* and *in vitro* enrichments on wheat straw. 172 fosmid metagenomic clones were isolated, producing various original glycoside-hydrolases, esterases and oxidoreductases. In particular, 23 enzymes, mostly issued from unknown bacterial genera, harbour still uncharacterized structural modules. These novel enzymes present high potential for numerous biotechnological processes, as they are highly efficient, alone or acting synergistically, to break down crude plant cell wall extracts.

Keywords: activity based metagenomics, bovine rumen microbiome, plant cell wall, wheat straw

1. Introduction

The bovine ruminal microbiota is a highly complex ecosystem constituted mainly by bacteria, which represent 10^{10} - 10^{11} cells per gram of rumen content, and by bacteriophages, archae, protozoa and fungi. Due to the paucity in their genome of genes coding for polysaccharide-degrading enzymes (the so-called Carbohydrate Active enZymes, or CAZymes (Lombard et al., 2014)), ruminants, like other mammals, depend on these symbiotic microorganisms within their digestive tract to break down most of dietary constituents, in particular plant cell wall (PCW). This complex polymeric assembly consists of a cellulose scaffold cross-linked with hemicelluloses, pectins, lignins and some proteins. The structural complexity of these biopolymers and their distribution, which varies for each plant type and growth stage, makes the complete degradation of PCW a complex issue.

Within the rumen, bacteria are the main actors of PCW degradation and fermentation, producing volatile fatty acids which are directly absorbed through the stomach wall to supply 60 to 80% of the energy requirements of the animal (Hall and Silver, 2009). Ruminal bacteria have evolved sophisticated machineries to bind, transport and degrade PCW and energy storage polysaccharides, like those encoded by the Bacteroidetes

Polysaccharide Utilization Loci (PULs, Terrapon and Henrissat, 2014), or the cellulosome-like structures described in cultivated Firmicutes (White et al., 2014; Ze et al., 2015).

Since only a minor fraction of the ruminal bacterial species is culturable (Wu et al., 2012), their sequence-based metagenomics has been largely used in the past decade to study their plant biomass degradation potential (Brulc et al., 2009; Hess et al., 2011; Wang et al., 2013; Patel et al., 2014). These studies revealed that the ruminal bacteriome is one of the richest and the most diverse one in terms of CAZy encoding genes. This specificity has been exploited many times to mine it for novel PCW polysaccharide degrading enzymes, by using activity-based metagenomics. This approach indeed provides a direct link between DNA and the encoded protein function, allowing one to discover novel enzyme families and/or novel functions. In total, 131 novel glycoside-hydrolases were identified by applying this approach to the ruminal bacteriome (Morgavi et al., 2013). Nevertheless, no novel GH family was still discovered from this ecosystem (André et al., 2014), and none of these enzymes was tested for efficiency of crude plant biomass degradation. However, these last two points are crucial for developing novel processes of lignocellulose bioconversion into second generation biofuels and biosourced synthons, which is still a scientific and economic challenge in the context of rarefaction of fossil resources and global warming (Menon and Rao, 2012). To be the most efficient as possible, glycoside-hydrolases and auxiliary enzymes like esterases and oxidoreductases have indeed to be used in cocktails, to take advantage of their various substrate specificities for disrupting the PCW network.

The efficiency of the bovine ruminal bacteria in degrading plant cell wall has also been exploited by engineered mixed cultures inoculated by rumen content. Such *in vitro* microbiome enrichment strategy, using lignocellulosic waste like wheat straw as sole carbon source, recently allowed obtaining stable hyper-hydrolytic mixed cultures (Lazuka et al., 2015). The microbial diversity, together with cellulase, cellobiohydrolase and xylanase activities, were significantly affected during enrichment. Nevertheless, the enzymes that are finally responsible for lignocellulose degradation in this enriched microbiota have still not been identified. Several metaomic approaches could be used to access them. Metatranscriptomics and metaproteomics would indeed be highly informative to identify the most expressed functional genes and the most produced proteins. However, these methods will not provide any experimental proof of their function.

In the study described here, we thus used activity-based metagenomics in order to identify key enzymes of lignocellulose degradation by two hyper-lignocellulolytic bacterial consortia issued from the bovine rumen. These microbiomes were obtained after *in vitro* and *in vivo* enrichments on wheat straw, an agricultural waste which tends to be used in European biorefineries to produce second generation biofuels (Menon and Rao, 2012). A multi-step high-throughput screening strategy was first set up to meet the three following challenges: i) accessing not only glycoside-hydrolases, but also esterases and

oxidoreductases, that are all necessary to completely break down PCW; ii) quantifying their efficiency to degrade wheat straw; iii) designing synergistic enzyme cocktails to increase the performances of crude plant biomass degradation. The taxonomic and functional profiles of the two enriched microbiomes has then been compared, in the light of the biochemical and metagenomic data today available regarding these semi-natural and artificial ruminal ecosystems.

2. Results and discussion

2.1. Primary high-throughput screening of the metagenomic libraries

The '*in vivo* enriched' (IVVE) rumen metagenomic library was constructed from a mix of the ruminal contents of two cows fed with a diet enriched in wheat straw. It consisted of 19,968 *Escherichia coli* fosmid clones, each clone comprising a 30-40 kb DNA insert. The '*in vitro* enriched' (IVTE) library was obtained from the ruminal contents of the same cows before dietary enrichment, but was enriched *in vitro* in a bioreactor, following the method described by Lazuka et al. (2015). This library consisted of 20,352 *Escherichia coli* fosmid clones.

Each library thus covers around 700 Mbp of metagenomic DNA, which corresponds to the equivalent of more than one hundred bacterial genomes. After primary screening, 172 hits were isolated on both libraries, some of them being active on more than one substrate.

Chromogenic substrates were used for CAZy mining. Endo-acting hemicellulases were screened by using AZCL-xylan and AZO-Carob Galactomannan. From the IVVE library, 30 hits (yield 0.15%) were active on the first, while 9 hits (yield 0.04%) were found on the latter. Very surprisingly at this stage, the results were completely different for the IVTE library, since 4 (yield 0.02%) and 53 hits (yield 0.26%) were found, respectively, on these substrates.

Endo- β -glucanase activities, including endo-(1,4)- β -D-glucanase (cellulase) and endo-1,4-1,3- β -D-glucanase (lichenase) activities, were screened by using the AZO-CarboxyMethyl-Cellulose (Azo-CMC) and AZCL-Barley β -glucan substrates. Here again, the primary screening results were different with the two libraries. From the IVVE library, 8 hits (yield 0.04%) were active on AZCL-Barley β -glucan, while none was found on Azo-CMC. For the IVTE library, 19 hits (yield 0.09%) and 44 hits (yield 0.22%) were found, respectively, on these substrates. Only eight of these clones were positive on both substrates, highlighting the interest to screen for endo- β -glucanase activities by using substrates with various structures (for instance polymers constituted by β -(1-4)/(1-3)-linked D-glucose residues like in barley β -glucan, or only by β -(1-4) linked D-glucose residues like in cellulose) and chromogenic groups, to sort out enzymes depending on their specificities.

In addition, the libraries were screened for esterase activities, since carbohydrate-esterases are also key enzymes to break down hemicellulosic components. Esterases were first screened by using Tween 20 (Polyethylene glycol sorbitan monolaurate, an ester with C12 chain length), a substrate known to be easily degraded by esterases (Kulkarni et al., 2013), including carbohydrate-esterases (Ufarté et al., 2016a). After this primary screening, 26 hits (yield 0.13%) were isolated from the IVVE library, while only 3 (yield 0.01%) were found for the IVTE one (Table 11).

Table 11 : Number of hits obtained from the primary screening of both metagenomic libraries. The percentage in brackets corresponds to the hit yield

Activity	Substrate	IVVE library (rumen enrichment)	IVTE library (bioreactor enrichment)
esterase	Tween 20	26 (0.13%)	3 (0.01%)
endo-1,4- β -D-xylanase	AZCL-xylan (birchwood)	30 (0.15%)	4 (0.02%)
endo-1,4- β -D-mannanase	AZO-Carob galactomannan	8 (0.04%)	53 (0.26%)
β -glucanase, including endo-1,4- β -D-glucanase (cellulase) and endo-(1,4)-(1,3)- β -D-glucanase (lichenase) activities	AZCL-Barley β -glucan	8 (0.04%)	19 (0.09%)
endo-1,4- β -D-glucanase (cellulase)	AZO-CM-cellulose	0 (0%)	44 (0.22%)
oxidoreductase	Lignin alkali	3 (0.02%)	3 (0.01%)
oxidoreductase	ABTS	0 (0%)	2 (0.01%)

Finally, screening for bacterial oxidoreductases, which might be involved in ruminal lignin degradation, was also performed on both libraries, using lignin alkali, a degradation product of lignin, and ABTS, a chromogenic substrate usually used for characterizing oxidases acting on polyaromatics, especially laccases. ABTS-based assays were previously set up in liquid medium, allowing activity detection of either extracellular fungal oxidases (Regalado et al., 1999) or intracellular recombinant enzymes expressed in *E. coli* (Boonen et al., 2014). However, to our knowledge, no screen was previously described for screening oxidases on

solid medium using this specific substrate. Here, we developed a very-high throughput method to screen oxidase activity on solid plates. We transferred the recombinant clones from the culture medium to the screening medium, which is lethal to *E. coli* cells because of the presence of high amounts of Cu and Mn, by using a PVDF membrane. Like other solid medium assays, this method allows one to screen enzyme libraries at a throughput of 400,000 assays per week, by using a simple colony picker automate.

From the IVVE library, only 3 clones (0.02%) were found active on lignin alkali (LA), and none reacted on ABTS. For the bioreactor enrichment, 3 hits (0.01%) were found on lignin alkali, and 2 (0.01%) on ABTS (Table 11). None of these clones was found positive on the two screening media. These low oxidase hit yields are in accordance with the few data available in the literature regarding bacterial rumen oxidases acting on cyclic compounds, which are likely not abundant in this ecosystem, although poorly characterized. Our results are also consistent with the data of Lazuka et al. (2015), which show that lignin is not degraded by the rumen-wheat straw-derived consortium corresponding to the IVTE one. Finally, the anaerobic conditions of plant biomass degradation by the IVVE and IVTE consortia do not favor oxidative reactions, although all oxidoreductases do not need oxygen (Ufarté et al., 2016b), even if some digestive reactions may partially occur in aerobic conditions when ruminal content is regurgitated during rumination.

A much lower yield of hit clones active on Tween20 and AZCL-xylan has been observed, along with a higher hit yield on AZCL-galactomannan, AZCL- β -glucan and Azo-CMC, by comparing the results obtained from the IVTE library with those from the IVVE one. Duration and efficiency of substrate degradation in both ecosystems might explain these differences. In the bovine rumen, the transit time is indeed between 20 to 48h, while each step of the *in vitro* enrichment on wheat straw lasted 7 days for each of the 10 cycles that were necessary for stabilization of the ecosystem. In both ecosystems, the plant biomass that is degraded first is the readily-accessible one, like soluble hemicelluloses. The crystalline parts of semi-crystalline polymers, like mannans and cellulose, are probably further degraded only if needed, once they become accessible to some specific enzymes that are able to act on semi-crystalline polymer structures. Endo- and exo-hemicellulases and, in minor extend, cellulases acting on soluble polysaccharides are thus probably predominant in the natural ruminal ecosystem, even when the sampled cows were predominantly fed with wheat straw. This is in accordance both with the high xylanase and esterase hit yields in our screening data, and with the sequence-based metagenomic data analyzed by Wang et al. (2013), who compared the abundance of CAZY families, and their main related activities, in diverse ruminal microbiomes. According to these authors, in the cow rumen, the sequences of the major enzyme families responsible for hemicellulose (except mannans) and oligosaccharide degradation indeed represent 85% of the ruminal CAZome. In contrast, the families which mostly represent cellulases and mannanases only represent 15 %.

The fact that no hit active on Azo-CMC was found from the IVVE library is probably due to the three following reasons: i) the quite low yield of β -glucanase encoding genes in the native cow rumen microbiome: from the data of Hess et al. (2011), we calculated that GH5, GH6, GH7, GH9, GH44, GH45 encoding genes would be present at a frequency of only 1 hit every 109 Mb, a number particularly close to the hit frequency obtained after screening on AZCL- β -glucan (1 hit every 100 Mb); ii) the possible differential heterologous expression of genes encoding enzymes that might be active on this screening substrate, depending on their taxonomical origin.

As mentioned before, we obtained much more β -glucanase active clones from the IVTE library, but less xylanase and esterase active ones. Again, this might be due to the duration of substrate degradation in the bioreactor, which is much longer than *in vivo*, and to the stringency of the microbial culture conditions which is much more higher *in vitro*, with wheat straw constituting the sole carbon source. These two factors explain why a large part of the readily accessible biomass was degraded at the sampling time. Lazuka et al. (2015) indeed measured that around 60% of the initial cellulose and hemicelluloses were metabolized after 15 days incubation using the rumen-wheat straw derived consortium at the end of the last enrichment step, and that the ratio between crystalline and amorphous cellulose significantly increased. This IVTE consortium is thus particularly efficient to break down hemicelluloses and amorphous cellulose, which is consistent with the high xylanase and CMCase activities quantified at the end of enrichment. These data explain why we found much more hit clones on Azo-CMC, AZCL-Barley β -glucan (representing 9 β -glucanase hits every 100 Mb of screened metagenomics DNA) and AZO-Carob Galactomannan from the IVTE library, compared to the IVVE one, which was issued from a much more natural microbiome. This highlights strong differences in communities composition between the two enriched microbiota, those producing CMCases that can be expressed in *E.coli*, being probably not present in the same amounts.

On the contrary, the data from Lazuka et al. do not explain why we obtained low yields of xylanase and esterase hit clones from the IVTE library. The fact that the IVTE consortium produces a high xylanase activity does not necessarily mean that the corresponding metagenome is enriched in xylanase encoding genes that, in addition, are compatible with heterologous expression in *E. coli*. The high xylanase activity detected in the IVTE consortium might indeed be due i) to a high abundance of only some xylanase encoding genes (which would not be properly expressed in *E. coli*), because of a specific enrichment in the strains which produce them; and/or ii) to a strong expression of some xylanase encoding genes in these specific enrichment conditions; and/or iii) to a particularly high catalytic efficiency of the enzymes that are specifically expressed at the end of the enrichment step.

2.2. Discrimination screening

To determine the substrate specificities of the encoded enzymes, and to select clones that harbor multigenic clusters encoding various synergistic activities, all the clones selected after primary screening were then tested on other chromogenic hemicellulosic (AZCL-arabinoxylan, AZCL-arabinan, AZCL-xyloglucan) and cellulosic substrates (AZCL-hydroxyethyl-cellulose (AZCL-HEC), AZO- α cellulose and AZO-Avicel).

Among the 172 hits from the primary screening, 68 were active on more than one polysaccharidic substrate (Table 12, Figure 31 and Figure 32). In particular, 32 of the 34 clones active on AZCL-xylan were also active on AZCL-arabinoxylan.

Table 12 : Number of hits obtained from secondary screening.

	Functional activity	Substrate	IVVE clones	IVTE clones
esterases	esterase	<i>para</i> -nitrophenyl acetate	16	3
	esterase	<i>para</i> -nitrophenyl butyrate	20	3
	lipase	<i>para</i> -nitrophenyl palmytate	3	3
polysaccharidases	endo-1,4- β -D-arabino-xylanase	AZCL-arabinoxylan (wheat)	30	8
	endo-1,5- α -L-arabinanase	AZCL-debranched arabinan	1	1
	endo-cellulase	AZCL-xyloglucan (tamarind)	2	10
	endo-cellulase	AZCL-HE Cellulose	3	10
	endo-cellulase	AZO-Avicel	0	0
	endo-cellulase	AZO- α cellulose	0	0
oxidoreductases	oxidoreductase	ABTS	3	5

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

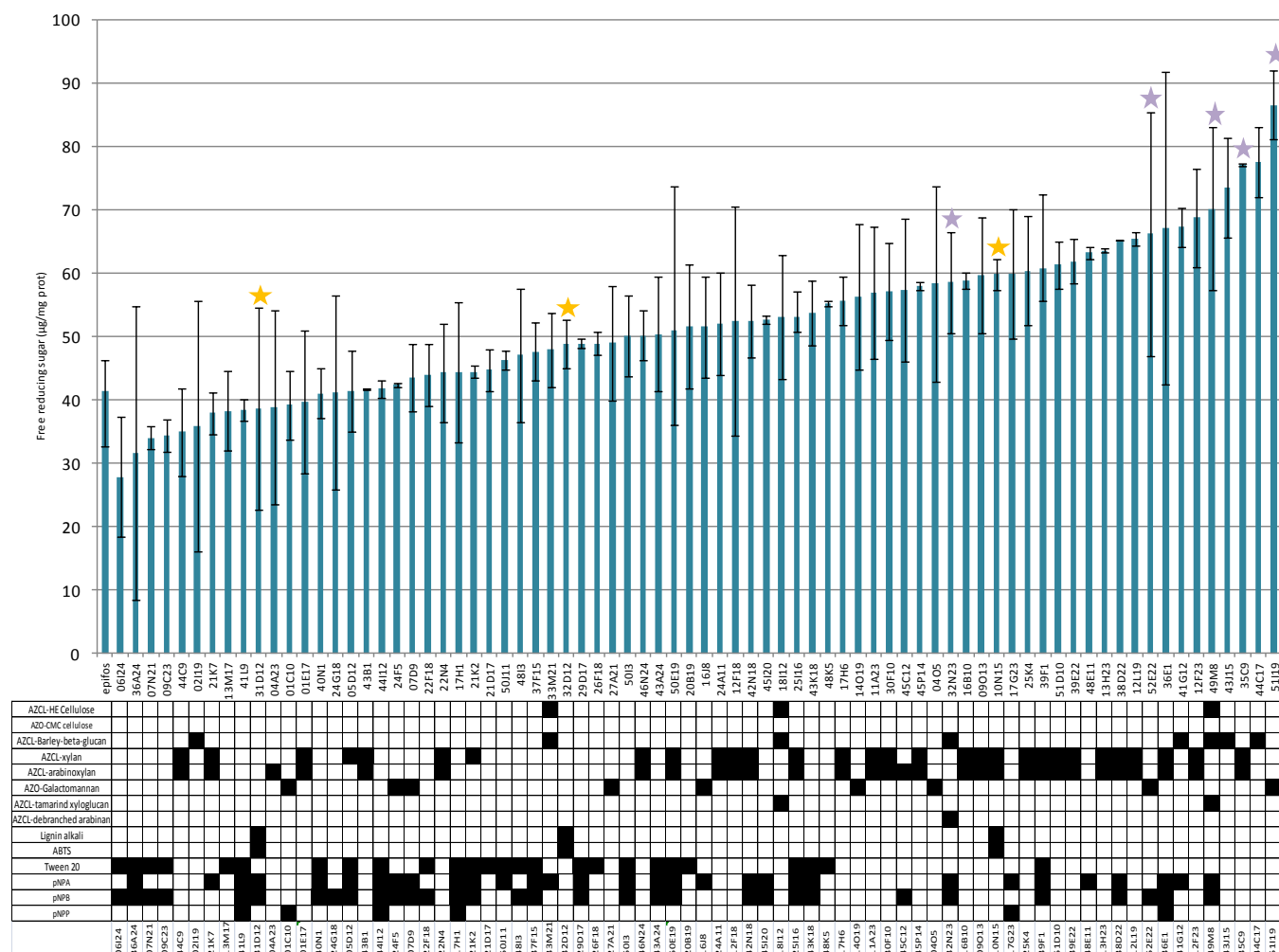


Figure 31: Activity of the IVVE hit clones on wheat straw. The stars show the clones included in the enzymatic cocktails. Purple star: clones with polysaccharidase and/or esterase activity; gold star: clones with oxidase activity. The table shows the substrates on which the clones are active.

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

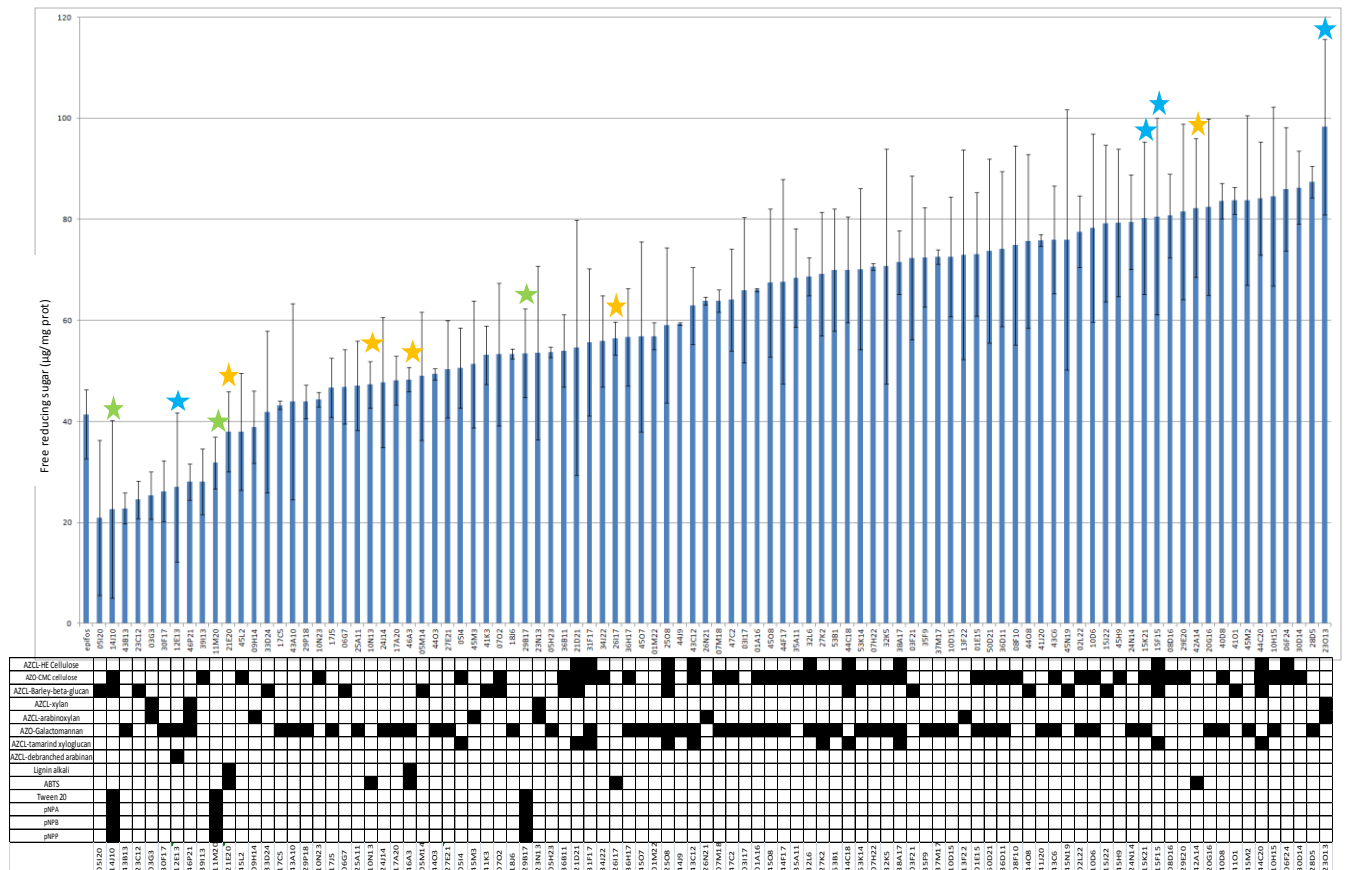


Figure 32: Activity of the IVTE hit clones on wheat straw. The stars show the clones included in the enzymatic cocktails. Purple star: clones with polysaccharidase and/or esterase activity; gold star: clones with oxidase activity. The table shows the substrates on which the clones are active.

Regarding endo-β-glucanase activities, most of those selected on Azo-CMC were also positive on AZCL-HEC, another chemically modified amorphous and soluble cellulose. But unfortunately, no clone was found active on the chromogenic substrates representing the most resistant fractions of cellulose, namely AZO-α cellulose (the insoluble amorphous alkali resistant cellulose fraction, with a very high polymerization degree) and AZO-Avicel (a highly crystalline cellulose fraction). The amorphous parts of cellulose are indeed the easiest to degrade, and this fraction was the first to be degraded by the IVTE consortium. However, α-crystalline, as well as crystalline parts of cellulose, are known to be highly resistant to most of endo-enzymes (Schwarz, 2001; Lynd et al., 2002; Wilson, 2011). To our knowledge, there is still no cellulase found active on Avicel or α-cellulose, discovered by activity-based metagenomics, whatever is the mined ecosystem.

In addition, 10 of the 44 clones obtained from the IVTE library on CMC are also active on AZCL-xyloglucan, a hemicellulosic component. At this stage, one can think that most of the endo-cellulases retrieved here thus present a high flexibility towards amorphous substrates (CMC, HEC, xyloglucan, β-(1,3)-(1,4)-glucans). The double specificity endo-

xyloglucanases / endo-cellulases, is a trait that has already been observed for many enzymes, like those belonging to family GH5, GH8 and GH44 (Ko et al., 2011; Warner et al., 2011; Bhat et al., 2013). Nevertheless, the ability of these clones to degrade both cellulosic substrates and xyloglucan might also be due to the expression of various GH encoding genes present on the same metagenomic DNA insert. This will be further analyzed in this paper, in the light of sequence analysis results.

All the primary screening hits were also tested on *para*-nitrophenyl acetate (pNPA), *para*-nitrophenyl butyrate (pNPB), *para*-nitrophenyl palmitate (pNPP). The activity levels were low (between 0.4 and 2×10^{-4} U/ml of extract), since in fosmids, the metagenomic gene expression is not under the control of strong promoters, but rather under the control of sequences randomly scattered on the metagenomic DNA insert, which are recognized as promoters by *E. coli* (Tauzin et al. 2015., In prep.). However, these activity levels allowed us to discriminate clones in function of their substrate specificity. 47 clones were active only on pNPA and/or pNPB, confirming the presence of esterases, some of them being carbohydrate-esterases involved in hemicellulose breakdown, as explained below. Only 9 clones were also active on pNPP, a substrate specific to lipases, acting on long fatty acid chains, concomitantly or not with pNPA and/or pNPB. These enzymes might be involved in the still not well known microbial lipolysis mechanisms in the cow rumen.

Regarding clones unearthed by the oxidoreductase screening, discrimination screening was performed using concentrated enzymatic extracts on the chromogenic polycyclic compound, ABTS. Unsurprisingly, again the measured activities were low (between 0.1 and 0.8 U/L of culture). However, activity displayed by these 8 clones was at least 3 times higher than that of the negative control (an *E. coli* clone containing an empty fosmid), which is itself able to slightly break down ABTS, thanks to any of the *E. coli* oxidoreductase activities. Among the 8 hit clones, seven showed activity on ABTS without any copper added to the reaction medium, indicating that they are not laccases. The last one, issued from the IVTE library, did not require any copper, but H₂O₂. This clone thus probably harbors a peroxidase (data not shown).

It is difficult to know what the role played by these oxidoreductases is in their natural habitat, and thus, what is their physiologic substrate. While cows fed predominantly on wheat straw might be better equipped to degrade lignin, it is known that fungi have a specific activity towards it along with an ability to produce esterases to hydrolyse ester linkages between lignin and hemicelluloses or cellulose (Goel et al., 2015). Nevertheless, it is possible that the presence, abundance and functional importance of bacterial ligninases in the rumen would be underestimated, as it was the case until recently in various microbial ecosystems (Ausec et al., 2011; Strachan et al., 2014). It is also possible that lignin degradation in the rumen has been overestimated. Further biochemical and sequence analyses are now necessary to estimate the applicative potential of the oxidases retrieved here, and to investigate their role *in vivo*.

2.3. Synergic activities of PCW degradation

All the hit clones were tested for their ability to degrade wheat straw. Screening of polysaccharide hydrolysis performances was performed in micro-plates by quantifying the reducing sugars released after incubation of clone cytoplasmic extracts with a fixed amount of milled straw. To render it compatible with the low levels of activity produced by fosmid clones, we had to adapt the method previously described by Song et al. (2010) who used plasmid clones expressing heterologous enzymes under the control of strong inducible promoters (Figure 31 and Figure 32). Because the DNS method was not sufficiently sensitive, we thus used the Nelson-Somogyi method (Nelson, 1944), which allows the quantification of reducing sugars down to 250 mg/L, which is 10 times more sensitive. In addition, in order to eliminate bias due to differences of cell growth and protein production between the metagenomic clones, we normalized the data with the amount of total proteins in the cellular extracts.

Clear differences between IVVE and IVTE clones were observed. Indeed, 23 % of the IVVE clones presenting activity on any of the primary or secondary screening polysaccharide substrates produced at least 1.5 times more reducing sugars than the negative control. This value reached 52 % for the IVTE clones, highlighting the efficiency of bioreactor enrichment for mining crude plant biomass degrading enzymes. In addition, surprisingly the most efficient clones on wheat straw (which produced up to 100 µg of reducing sugars per mg of total proteins, a value corresponding to 110 µg of reducing sugars per ml of culture) were not necessarily those producing complementary activities of PCW degradation, like for instance cellulases and xylanases. Nevertheless, since we did not screen for exo-acting CAZymes, at this stage, it is difficult to know if this result is due to a particularly high expression and/or activity of the endo-acting enzymes produced by the most efficient clones. It might also be due to the presence of several glycosidases which could massively contribute to the release of reducing sugars, by cleaving hemicellulosic side chains. This will be discussed below, in the light of the sequence annotation data.

The primary and secondary screening data, as well as those of wheat straw degradation efficiency by individual clones were used to select clones to be mixed in PCW degrading cocktails. The potential synergistic effects of the various activities produced by the metagenomic clones for crude plant biomass breakdown were tested on wheat straw, by using the same method as the one described above. Several clones were chosen to be mixed in enzymatic cocktails, according to 2 criteria: i) the production of several complementary activities required for PCW degradation, due to multi-genic clusters encoding a battery of CAZymes, and potentially also of oxidases; ii) the clone efficiency of straw degradation, which had to correspond at least to 1.5 times more released reducing sugars than with the negative control, at least for the 'polysaccharidases', except when only one clone shows a specific activity (Figure 31 and Figure 32).

From the IVVE clones, 5 clones were selected to constitute a ‘polysaccharidases-esterases’ mix, since the several of the most efficient clones on wheat straw displayed both activities. This ‘polysaccharidases-esterases’ mix is constituted by enzymes acting on xylan, arabinoxylan, debranched arabinan, barley β -glucan, xyloglucan, galactomannan, HEC and pNP-substrates. A second IVVE mix was created, named ‘polysaccharidases-esterases-oxidases’, containing the 5 clones of the ‘polysaccharidases-esterases’, supplemented with the three clones harbouring an oxidoreductase activity.

Four different mixes were created from the IVTE clones. The ‘polysaccharidases’ one was constituted by 4 clones, producing enzymes active on xylan, arabinoxylan, barley β -glucan, xyloglucan, galactomannan, debranched arabinan, CMC, and/or HEC. The ‘polysaccharidases-esterases’ mix was supplemented by 3 clones with esterase activities. A third mix, named ‘polysaccharidases-oxidases’, was created by adding to the ‘polysaccharidases’ mix the 5 clones with oxidoreductase activity. The fourth and last IVTE mix, named ‘polysaccharidases-esterases-oxidases’, contains the 12 different clones (Table 13).

Table 13: List of the clones included in the lignocellulase cocktails, their functional properties and CAZome, deduced from sequence analysis. CE* corresponds to a sequence that is very distantly related to a CE family

Mix	Library	Clones	Activity on wheat straw (μ g free reduced carbohydrate/mg protein)	CAZy content	Activities
Polysaccharidases-esterases	IVVE	51I19	87 \pm 5	No known family detected	Azo-Galactomannan, wheat straw
		49M8	71 \pm 13	CE9 CBM22-CE1 DOC-GH5_4	AZCL-HE cellulose, AZCL-Barley-beta-glucan, AZCL-xyloglucanpNPA, pNPB, wheat straw
		32N23	59 \pm 8	Sequencing underway	AZCL-Barley-beta-glucan, AZCL-debranched arabinan, pNPA, pNPB, wheat straw
		35C9	78 \pm 1	GH5_4 CBM22-GH10 GH5_4	AZCL-xylan, AZCL-arabinoxylan, wheat straw

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

		52E22	67±19	GH130_2 CBM35-GH26 GT5 GH3	Azo-Galactomannan, pNPB, wheat straw
Polysaccharidases-esterases-oxidases (clones added to the polysaccharidases-esterases mix)	IVVE	10N15	60±3	CBM22-GH10 GH94	AZCL-xylan, AZCL-arabinoxylan, lignin alkali, wheat straw
		31D12	39±16	GH2_6	Lignin alkali, pNPA, pNPB, pNPP, ABTS
		32D12	49±4	GH13 GT2 CBM22-GH10 CE2	Lignin alkali, wheat straw
Polysaccharidases	IVTE	12E13	27±14	GH51 GH43_4 GH26	AZO-galactomannan, AZCL-debranched arabinan
		15K21	80±15	GH5_7 GH5_4 GH26 GH130_1 GH130_1	AZO-CM cellulose, AZO Galactomannan, wheat straw
		23O13	99±17	GH5 GH53 GH11-X238	AZCL-xylan, AZCL-arabinoxylan, wheat straw
		15F15	81±19	GH5_4	AZCL-Barley beta glucan, AZO-CM cellulose, AZCL Tamarind xyloglucan, AZCL-HE cellulose, wheat straw
Polysaccharidases-esterases (clones added to the polysaccharidases mix)	IVTE	11M20	32±5	No known family detected	Tween 20, pNPA, pNPB, pNPP
		29B17	54±9	Sequencing underway	
		14J10	23±17	GH5_37 CE*_CE1	AZCL-Barley beta glucan, AZO-CM cellulose, Tween 20

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

Polysaccharidases-oxidases (clones added to the polysaccharidases mix)	IVTE	42A14	82±14	No known family detected	ABTS, wheat straw
		46A3	48±2	GH5_4 GH26 GH130_1 GH23-CBM50	lignin alkali, AZO-galactomannan, wheat straw
		21E20	38±8	GH2GT19	lignin alkali
		10N13	47±5	No known family detected	ABTS, wheat straw
		26I17	56±3	No known family detected	ABTS, wheat straw
Polysaccharidases-esterases-oxidases (clones added to the polysaccharidases-esterases mix)	IVTE	42A14	82±14	No known family detected	ABTS, wheat straw
		46A3	48±2	GH5_4 GH26 GH130_1 GH23-CBM50	lignin alkali, AZO-galactomannan, wheat straw
		21E20	38±8	GH2 GT19	lignin alkali
		10N13	47±5	No known family detected	ABTS, wheat straw
		26I17	56±3	No known family detected	ABTS, wheat straw

First, a synergistic effect of the various activities present in the cocktails was observed, since the efficiency of straw degradation was between 1.2 and 2.5 times higher than the theoretical value corresponding to the mean of the degradation efficiencies of all the clones constituting the cocktails (Figure 33). The highest synergistic effect was observed for the IVTE 'polysaccharidases-esterases' mix, of which the straw degradation efficiency is 2.5 folds higher than the theoretical value. The IVTE esterases (which are classified in carbohydrate-esterase families, as further explained) clearly positively affect straw degradation efficiency, by facilitating glycoside-hydrolase access to hemicellulose main chains.

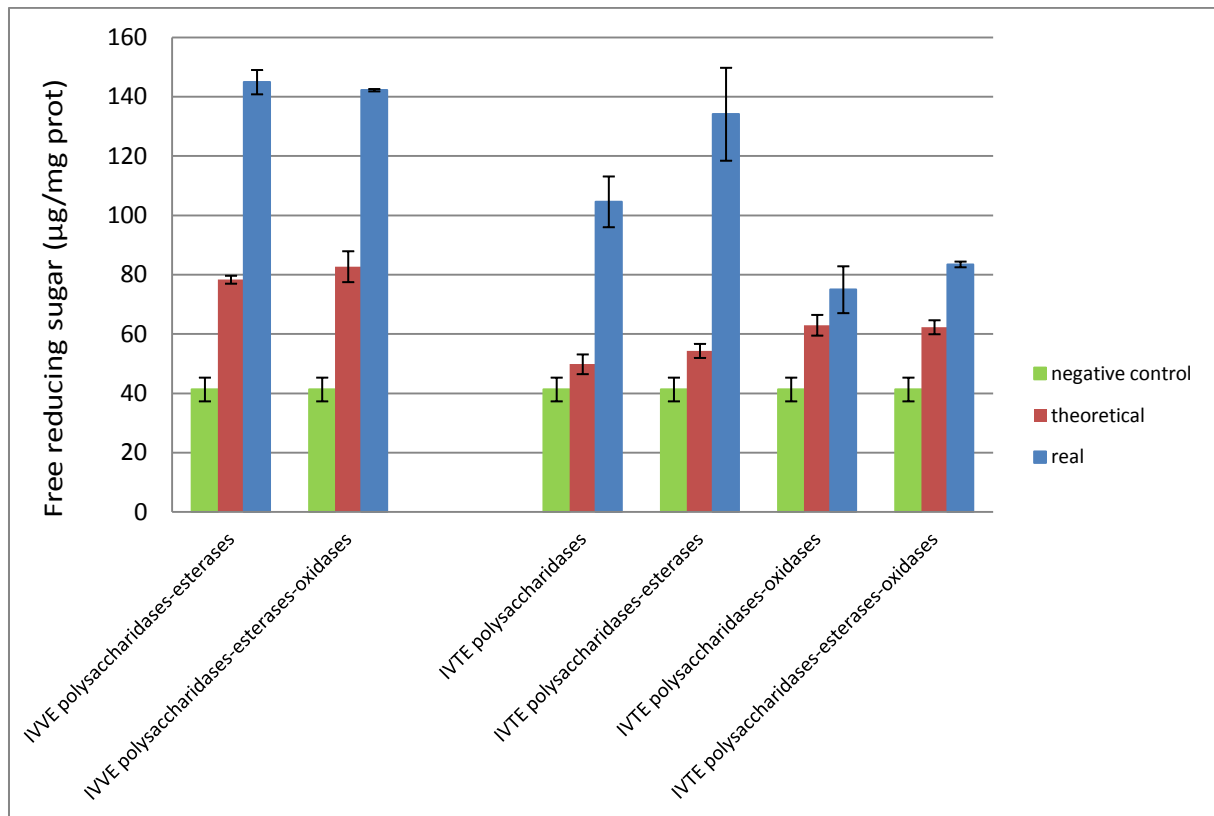


Figure 33: Activity of the different lignocellulolytic mixes on wheat straw.

Effect of oxidoreductases was different. The efficiency of the two IVVE cocktails was indeed equivalent, indicating that addition of IVVE oxidoreductases to the ‘polysaccharidases-esterases’ mix did not impact to the efficiency of straw degradation. On the other hand, the effect of the IVTE oxidoreductases on the IVTE ‘polysaccharidases’ and ‘polysaccharidases-esterases’ mixes was rather negative, indicating that at least one of these particular oxidases might have released lignin-degradation products that inhibit CAZymes, or reorganized lignin, even through depolymerization/repolymerization, limiting accessibility to the polysaccharidic network (Jørgensen et al., 2007).

These results highlight the interest of such a screening strategy to select metagenomic clones which, alone or acting in synergy, are the most efficient to break down crude plant biomass. This is crucial, since activity-based metagenomics, which is frequently used to mine microbial ecosystems for novel CAZymes, often generates hundreds of hit clones per study, which, until now, were difficult to sort out based on non-quantitative assays performed on chromogenic substrates, which are far from mimicking the complexity of real PCW matrixes.

2.4. Sequencing and data analysis

Screening results, particularly ability to break down several PCW constituents, as well as efficiency of straw degradation allowed us to select 32 IVVE and 27 IVTE clones for

sequencing of the metagenomic DNA inserts (Table S1). In total, 2 Mb of metagenomic DNA was sequenced, with an average coverage of 40 x. One or two contigs sizing more than 30 kb were obtained for each clone. No sequence redundancy was observed between IVVE and IVTE clones. Nevertheless, several cases of total and partial sequence redundancy were observed, between clones from the same library, especially the IVTE one. Two loci, sizing between 9 and 35 kb, were indeed found in 3 (clones 15F15, 44C18 and 44C20) and 7 IVTE clones (clones 15K21, 20G16, 28D5, 38A17, 44D8, 44F17 and 46A3), respectively. This phenomenon might be due to a lowest microbial diversity in the IVTE sample, compared to the IVVE one. The Simpson index of the IVTE sample was indeed of only 11.7 (Lazuka et al., 2015), against 36.8 in the non-enriched ruminal microbiota of the sampled cows. The determination of the IVVE Simpson index will allow us soon to confirm this hypothesis. Nevertheless, the high sequence redundancy between IVTE clones might also be due to the difficulties encountered to discriminate them based on the screening results. Most of the clones harboring these redundant sequences indeed present the same activity profile.

In contrast, only 2 IVVE clones (44C17 and 43J15) harbor identical sequences, 3 share highly similar sequences (> 98% identity and > 70% contig coverage) (clones 12F23, 32D12 and 38D22), and 3 others (clones 50E19, 45P14 and 25I16) contain sequences presenting redundancies at their extremities, allowing one to artificially contigate them to obtain a 52 kb metagenomic sequence probably issued from one unique species.

2.4.1. Taxonomic assignation

Firstly, BLASTn analysis against the non redundant database of the NCBI revealed that 2 IVVE clone sequences (50E19 and 45P14) present high homology (> 80 % sequence identity >50% coverage) with metagenomic contigs obtained from ruminal samples issued from Chinese Holstein cattle (Cheng et al., 2012) and 5 others (16B10, 25I16, 42N18, 45P14 and 50E19) from Jersey cows (Wang et al., 2013). These sequences, which belong to xylanase active clones, are thus issued from highly prevalent xylanolytic ruminal bacteria, that have been assigned to the Firmicute phylum (Cheng et al., 2012; Wang et al., 2013).

Surprisingly, significant homology (> 99 % sequence identity, >90% coverage) was also found between IVTE contigs and genomic loci of the human bacterium *Enterococcus faecium* T110 (clone 45H9), or human gut metagenomic sequences (clone 10D6, corresponding to the sequence GU942953.1, of a metagenomic clone active on AZCL- β -glucan and Yeast β -glucan) assigned to the *Bacteroides* genus (Tasse et al., 2010), indicating that the function-targeted strategy allowed us to discover major fibrolytic pathways of gut microbiomes.

Secondly, Megan sequence analysis revealed that the most probable common ancestors to which belong the sequences of the IVVE library were mostly bacteria from the

Firmicute phylum (19 of the 32 clones of the IVVE library). Sequences were too distant from those of sequenced genomes, to annotate them below the phylum rank, except for two clones which were assigned to Clostridiales. However, in many cases, a few of the ORFs from these Firmicutes were assigned to *Faecalibacterium prausnitzii*, which is one of the most prevalent species of the rumen community (Ross et al., 2012). Regarding the other clones of this library, the assignments matched to Bacteria from environmental samples, without more taxonomic information, apart from their similarity to sequences retrieved from cow ruminal metagenome (Wang et al., 2013) and other environments. It means that our sequences, while similar to known sequences from different ecosystems do not belong to known organisms.

The assignment of the 27 IVTE clones showed a radical shift in bacterial population. Most of the clones (17 out of 27) were assigned to the Bacteroidetes phylum. In this population, it was possible to assign the sequences more precisely. Some clones were indeed assigned to the *Prevotella* or *Bacteroides* genus and even at the species rank to *Bacteroides graminisolvens* and *Bacteroides uniformis*, which are common anaerobic commensal and dominant strains of the rumen microbiome, known for their ability to degrade plant cell wall. Only two clones were assigned to the Firmicute phylum and one clone was assigned at species level to *Paenirhodobacter enshiensis* from the Proteobacteria phylum. In addition, two clones were assigned to bacteria whose sequences have similarities with bacterial sequences from environmental samples. Moreover, among the clones with sequences assembled in two contigs, six present a very different assignments for each contig: one being related to Firmicutes and the other to Bacteroidetes (clones 14J10, 15F15, 41J20, 44C18, 44C20), or one assigned to Clostridiales and the other to *Enterococcus faecium* (Lactobacillales) (clone 45H9). We did not observe this feature into the IVVE rumen library where only two clone sequences were assembled in two contigs, assigned to the same taxonomic group. The assignment of the two contigs to different taxonomic groups might indicate the mix of two fosmid in the same clone, each containing a metagenomic insert from different organisms. However, this hypothesis seems unlikely since fosmidic DNA extraction was performed from a culture of a single colony, and since such phenomenon was never observed for the hundreds of metagenomic clones sequenced by ourselves and others, which were obtained by using the same cloning strategy. In addition, this phenomenon cannot be due to miscontigation between DNA from different clones, since DNA samples were individually labeled for sequencing and read assembly. Another hypothesis that appears more likely, since among both contigs, one was very short, would be recombination events between DNAs from distant organisms, through horizontal gene transfers. However, we did not find insertion elements like transposases in the outcome of functional annotation, contrarily to what we previously observed in the human gut metagenome (Tasse et al., 2010; Cecchini et al., 2013).

We compared our taxonomic results to those of Lazuka et al. (2015), who, while working on the same IVTE metagenomic sample, studied the microbial diversity evolution before and after *in-vitro* enrichment by using 16S rDNA sequencing. In agreement with the previous findings of Ross et al. (2012), they found that Bacteroidetes, Firmicutes and to a lesser extent Proteobacteria, were the predominant phyla of the bovine ruminal community, even after artificial enrichment (these phyla represent 85 % of the total bacterial community in the IVTE sample). This is globally in accordance with our results since our sequences were assigned to the most represented phyla, Firmicutes and Bacteroidetes only one clone being assigned to Proteobacteria. Regarding the relative distribution of the main classes into the IVTE community, Lazuka et al. described an equal repartition of bacteria between Clostridia, Bacteroidia (a class of Bacteroidetes) and an unclassified class of Bacteroidetes. Our results did not reflect these findings since we found that most of our IVTE library sequences belonged to Bacteroidia.

As for the IVVE community, the equivalent microbial diversity analysis is still not available. In this library, we found that annotated sequences belonged exclusively to Firmicutes or to non-assigned bacteria. The absence of Bacteroidia in assignation of the IVVE clone hits was unexpected since bacteria from this class are known for their involvement in plant cell wall degradation. It is also well established and demonstrated in previous studies (Tasse et al., 2010) that *E. coli*, as an expression host, is capable of expressing genes from a great diversity of bacterial classes including Clostridia, Bacteroidia, Lactobaccillia... The coming results of 16S rDNA analysis of the IVVE community will help us, hopefully, to clarify this point.

2.4.2. Functional annotation

ORFs were then functionally annotated by using the RAST annotation server, and the results were confronted to those of BLASTP analysis against the Uniprot/Swissprot database, and of the CAZy mining. For nearly all hit clones, we found at least one ORF coding for an enzyme likely responsible for the evidenced clone activities i.e., carbohydrate hydrolysis, hydrolysis of ester bounds or oxidoreduction. In many cases these genes were organized into functional clusters, and surrounded by other genes having complementary and accessory functions for plant cell wall harvesting, degradation and metabolisation: complementary GHs (endo- and exo-enzymes, that were found in particular in the most active clones on wheat straw, like 12F23, 28D5, 30D14, 6F24), and CEs, but also carbohydrate transporters, isomerases, kinases, epimerases, sortases (a transpeptidase that attaches surface proteins to the cell wall) and signal peptidases.

In five cases (IVTE clones 10N13, 26I17, 42A14 and the IVVE clones 51I19, 43C17 and its redundant 43J15) the metagenomic clones degraded wheat straw, though no glycoside hydrolase sequence was found in their contigs. The three IVTE clones were also active on

ABTS or lignin alkali for which we found genes likely to be responsible of the activity (oxidoreductases, dehydrogenases, methyltransferases, acetylases). The IVVE clone 51119 was not active on lignin alkali, nor on ABTS though it harbored in its sequence at least one oxidoreductase which has not been evidenced by the screening. Finally, redundant sequences of the IVVE clones 43C17 and 43J15 contain only a GH32 encoding gene, which make no sense with the lignocellulosic activities of these clones.

Regarding the absence of CAZy encoding genes (or of families that have no characterized member acting on the same substrate that the hit clone, like 43C17 and its redundant 43J15) from the contigs associated with GH activities, we suggest that a CAZy encoding gene exists into the target sequence but was too original to be identified by sequence comparison. In particular, even if sequence analysis did not reveal any known lytic polysaccharide mono-oxygenase, it is not impossible that the red-ox enzymes we isolated may act on polysaccharides and release both oxidized and non-oxidized oligosaccharides. Very original GHs may also be produced by these clones. This is the case in particular for clone 10N13, for which RAST annotation suggests that it contains a probable β -1,4-exo-glucanase (even not retrieved by BLAST analysis against the CAZy database), and for clone 51119, which is highly active on AZO-galactomannan. Rational truncature, or transposition mutagenesis of these metagenomic inserts will allow us to identify which gene is responsible for carbohydrate degradation, and thus, to identify completely novel CAZy families.

For all other clones active on polysaccharides, 132 putative CAZyme sequences were identified in total. All are original compared to biochemically characterized, since they present less than 50% identity for less than 95% coverage with the best BLAST hits of the Swissprot database. Among them, several sequences are very distantly related to known GH or CE sequences, and appear as GH* or CE*. Some of them seem to be related to an already known CAZy family, but cannot be classified in them, because specific features of enzymes from this family (like for example one of the catalytic residues) are missing in the target sequence. Five GH* (2 related to GH63, and 2 to GH39) and one CE* were found from the IVVE library, and only one GH* (related to GH9) and one CE* (related to CE1) from the IVTE library. This is coherent with the higher diversity found in IVVE. Furthermore, X modules were also annotated, corresponding to a group of sequences forming a family, without any member characterized to date. Four X modules were found in the sequences from both enrichments. Two X91, one X19 and one X103 were found in IVVE, and one X238, one X153 and two X modules, distantly related to CBM4 in IVTE. Another point, observed only for the IVVE enrichment, is the presence of unknown modules of more than 50 amino acids that could correspond, depending on their size, to small non-catalytic modules (dockerin), CBMs, or even still uncharacterized catalytic modules for the longest ones. Ten were annotated as 'unknown' in the IVVE, with 4 sizing between 50 and 100 amino acids (1 being homologous to a dockerin, and the other to the C-terminal domain of a GH2 enzyme, according to Interproscan annotation), 4 between 100 and 200 amino acids (3 being homologous to the

C-terminal domain of GHs), and 2 with more than 250 amino acids (one being annotated as a mutarotase-like domain, and the other only as an uncharacterized conserved protein). These original sequences, in particular the GH*, CE* and X catalytic modules, along with the CAZy sequences we will further identify from the clones active on polysaccharides, for which none was found by the present analysis, constitute highly interesting targets for creation and characterization of novel CAZy families. Let's note that the number of still uncharacterized CAZy modules per Mb is more than 10 times higher than that retrieved from the human gut metagenome (Tasse et al. 2010, Cecchini et al., 2013) by using the same approach, highlighting the fascinating discovery potential of the bovine rumen microbiome.

3. Materials and methods

3.1. Chemicals

Tween 20, Pefabloc, *para*-nitrophenyl acetate (*p*NPA), *para*-nitrophenyl butyrate (*p*NPB), *para*-nitrophenyl palmytate (*p*NPP), *para*-nitrophenol (*p*NP), 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS), lignin alkali, bovine serum albumin (BSA) were purchased from Sigma Aldrich (France).

AZCL-xylan (birchwood), AZCL-xyloglucan (tamarind), AZCL-HE Cellulose, AZCL-arabinoxylan (wheat), AZCL-debranched arabinan, AZCL-Barley β -glucan, AZO-CM Cellulose, AZO-Avicel, AZO- α cellulose, AZO-Carob Galactomannan were purchased from Megazyme (Ireland).

Lysonase™ Bioprocessing Reagent was purchased from Novagen.

Protein Assay Dye Reagent concentrate 5X was purchased from Bio-Rad.

3.2. Biomass preparation

Wheat straw (Apache variety), harvested in 2007 in southern France, was obtained from ARD (Pomacle, France). Subsequently, the straw was milled to powder form (0.5 mm) using a blade grinder and washed with distilled water at 4 °C (>10 volumes). Straw powder was recovered by filtration using a Büchner funnel with Whatman® No.4 filter paper (pore size: 20–25 μ m), then dried in a dry oven. The dried powder was then sterilized by autoclaving.

3.3. Metagenomic DNA sampling and library construction

The non-producing Holstein dairy cows used in this study were reared according to the national standards fixed by the legislation on animal care (Certificate of Authorization to Experiment on Living Animals, No. 004495, Ministry of Agricultures, France). In particular, before enrichment, the cows were fed a standard dairy cow ration composed of corn silage (64% DM), hay (6% DM) and concentrate (30% DM). They were fed *ad libitum* once a day in the morning. The two cows received during 7 weeks a diet of 80% wheat straw, 12% concentrate and 8% beetroot molasses before sampling of their rumen content. For

sampling, the whole rumen content was collected from two individuals fitted with rumen cannula. Rumen contents were taken from various parts of the rumen and manually homogenized. Subsamples (30 g) were immediately frozen in liquid nitrogen before storage at -80°C . For constructing the IVTE library, the two cows were sampled before following a particular diet. This sample was used to inoculate the bioreactors fed only by wheat straw, as described by Lazuka et al., (2015). For IVVE, the same two cows were fed for seven weeks with wheat straw enriched food before sampling and library construction.

The metagenomic libraries were constructed in *Escherichia coli* strain by Libragen from the mix of rumen bacterial DNAs extracted from the bioreactor (IVTE library) or directly from both cows (IVVE library). Briefly, total metagenomic DNA was extracted and fragmented. Fragments sizing between 30 and 40 kb were isolated as previously described (Tasse et al., 2010) and cloned into pCC1FOS fosmid (Epicentre Technologies). EPI100 *E. coli* cells were then transfected to obtain a library of 20,352 and 19,968 clones from the rumen samples of the bioreactor and rumen enrichment respectively. Recombinant clones were transferred to 53 or 52 384-well microtiter plates containing Luria Bertani (LB) medium, supplemented with 12.5 mg/L chloramphenicol and 8% glycerol. After 22 hours of growth at 37°C without shaking, the plates were long-time stored at -80°C .

3.4. Primary high-throughput screening of the metagenomic libraries

The resulting libraries were gridded on 22 cm x 22 cm trays containing solid agar medium, using an automated microplate gridded (K2, KBiosystem, Basildon, UK), which allows to perform 140,000 assays in 2 days. Three major activities were screened:

3.4.1. Esterase/lipase activity

Esterase/lipase screening medium was constituted by solid LB medium supplemented with 12.5 mg/L chloramphenicol and 1% (w/v) final concentration of Tween 20. The assay plates were incubated during 3 days at 37°C until apparition of a powdery halo, signifying the presence of a positive clone.

3.4.2. Carbohydrate degrading activity

Polysaccharide degrading activity screening media were LB media, supplemented with 12.5 mg/L chloramphenicol and with chromogenic substrates (AZCL-xylan, AZCL-Barley β -glucan, AZO-CM Cellulose, AZO-Carob Galactomannan) at final concentration of 0.2% (w/v). After incubation of plates at 37°C , for plates containing AZCL-substrates, positive clones were visually detected after 1 to 10 days of incubation by the presence of a blue halo resulting from the production of colored oligosaccharides of which color diffused around the bacterial colonies. For AZO-polysaccharides, clear halos were observed around the positive clones.

3.4.3. Oxidoreductase activity

Oxidoreductase activity screening was done on two different substrates.

The first medium was a minimal medium supplemented with 12.5 mg/L chloramphenicol, 15 g/L agar and 1% (w/v) filter sterilized lignin alkali as sole carbon source (except agar itself). Minimal medium was composed of a final concentration of salts (3.6 g/L Na₂HPO₄, H₂O; 0.62 g/L KH₂PO₄; 0.11 g/L NaCl; 0.42 g/L NH₄Cl), 2 mM MgSO₄, 0.03 mM CaCl₂, other salts (15 mg/L Na₂EDTA, 2H₂O; 4.5 mg/L ZnSO₄, 7H₂O; 0.3 mg/L CoCl₂, 6H₂O; 1 mg/L MnCl₂, 4H₂O; 1 mg/L H₃BO₃; 0.4 mg/L Na₂MoO₄, 2H₂O; 3 mg/L FeSO₄, 7H₂O; 0.3 mg/L CuSO₄, 5H₂O), 0.04 g/L leucine and 0.1 g/L thiamine hypochloride. The assay plates were incubated at 37°C until visualization of the growth of hit clones.

The other screening method used ABTS as the substrate. A first layer was composed of solid LB medium with 12.5 mg/L chloramphenicol, on which was deposited a durapore PVDF membrane of 0.22 μM (Millipore) that was preliminary autoclaved. The clones were gridded and grown on the membrane. After growth, the membrane was moved to another medium, containing 15 g/L agar, 12.5 mg/L chloramphenicol, 0.1M sodium citrate pH 4.5, 1 mM CuSO₄ 5H₂O, 1 mM MnCl₂ 4H₂O and 1mM ABTS. Positive hits became green or grey, along with the presence of halos sometimes.

Each positive clone was picked, and streaked on Petri dishes containing solid LB medium supplemented with 12.5 mg/L chloramphenicol. Three colonies were picked, grown in liquid LB medium and stored in microplates containing LB medium with 8% of glycerol, to be used for the validation process. Omnytrays containing media which allowed the discovery of positive clones during primary screening were made. Gridding was made manually using the microplates containing the three copies of each positive hit. Only the most active clone was kept for further analyses.

3.5. Discrimination screening

3.5.1. Esterase/lipase activity

Validated hit clones obtained after primary screening for the esterase activity were grown in microplates containing 200μL of LB medium for pre-culture, at 37°C overnight. These pre-cultures were used to sow deep well micro-plates containing 1.6 mL of LB liquid medium supplemented with 12.5 mg/L chloramphenicol, which were subsequently incubated overnight at 37°C. Cells were pelleted by centrifugation for 30 minutes at 4°C and 1,760 x g. Each pellet was re-suspended in 250 μL of lysis buffer (100 mM HEPES pH 7.5, 1 mM pefabloc and 666.7 μL/L lysonase), and incubated for one hour at 37°C and 600 rpm (shaking throw 25 mm) in a Multitron Pro shaker (INFORS HT) with a freeze/unfreeze cycle at -80°C in order to perform cell lysis. Cellular extracts were isolated by centrifugation for 30 minutes at 4°C and 1,760 x g, and stored at 4°C for less than 24 h before use.

Esterase/lipase activity of the cytoplasmic extracts was estimated by monitoring the hydrolysis of *p*NP esters into the corresponding acid and *p*-nitrophenol. Assays were performed in 96-well microtiter plates on *p*NPA, *p*NPB and *p*NPP at 1 mM final concentration. *p*NP was used for standard curve. For *p*NPA and *p*NPB, 195 μ L of cellular extract were mixed with 5 μ L of substrate dissolved at 40 mM in 2-methyl-2-butanol (2M2B). For *p*NPP, 175 μ L of cellular extract were mixed with 25 μ L of substrate dissolved at 8 mM in isopropanol. Enzymatic activity was determined by following the absorbance increase at 405 nm during 30 min at 30°C, in a microplate spectrophotometer (BioTek™ Eon™ Microplate Spectrophotometers, Colmar, France).

3.5.2. Polysaccharide degrading activity

Validated hit clones obtained after primary screening for the carbohydrate degrading activity were further characterized using chromogenic substrates in solid medium (AZO-Avicel, AZO HE cellulose, AZO- α cellulose, AZCL arabinoxylan, AZCL-xyloglucan, AZCL debranched arabinan) at final concentration of 0.2% (w/v). After incubation of plates at 37°C, for plates containing AZCL-substrates, positive clones were visually detected after 1 to 10 days of incubation by the presence of a blue halo resulting from the production of coloured oligosaccharides of which colour diffused around the bacterial colonies. For AZO-polysaccharides, clear halos were observed around the positive clones.

Further characterization was done using wheat straw as substrate. Validated hit clones were grown in microplates containing 200 μ L of LB medium for pre-culture, at 37°C overnight. These pre-cultures were used to sow deep well micro-plates containing 600 μ L of LB liquid medium supplemented with 12.5 mg/L chloramphenicol, which were subsequently incubated overnight at 37°C. 400 μ L of each well were used to create a new deep-well, and 40 μ L of 5g/L lysozyme in PBS pH 7.0 were added to each well, and the deep-well was incubated for one hour at 37°C and 600 rpm (shaking throw 25 mm) in a Multitron Pro shaker (INFORS HT) with a freeze/unfreeze cycle at -80°C in order to perform cell lysis. The reaction mixture was composed of 250 μ L of cellular extract, 250 μ L of PBS pH 7.0 and the equivalence of 90 μ L of wheat straw. To do so, a transfer device (45 μ L capacity Multiscreen® column loader, Millipore Inc., USA) compatible with 96-well microtiter plate format was employed to deliver a fixed amount of wheat straw to all individual wells in polypropylene microtiter plates (Song et al., 2010). The deep-well was incubated at 37°C and 600 rpm (shaking throw 25 mm) in a Multitron Pro shaker (INFORS HT). One hundred μ L samples were extracted after centrifugation at 2,250 x g for 30 minutes. Free reduced sugar was quantified using the Nelson-Somogyi method, and the activity was measured in μ g of free reduced sugar / mg protein.

In the Nelson-Somogyi method (Nelson, 1944), solution A is composed of 200 g/L of anhydrous Na₂SO₄, 25 g/L anhydrous Na₂CO₃, 25g/L of potassium sodium tartrate, 20 g/L

anhydrous NaHCO_3 . All these powders must then be sprinkled on 800 mL of warm ultrapure water, before adjusting to 1 liter. The solution is then filtered and stocked at 30-35°C in order to avoid crystallization. Solution B is composed of 150 g/L of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, and 20 drops of concentrated H_2SO_4 , before adjusting to 1 liter. Solution C is a mix of 25 mL of solution A and 1 mL of solution B. Solution D is composed of 50 g/L of $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}$, 4 H_2O , dissolved in 800 mL of ultrapure water, 42 mL/L of concentrated H_2SO_4 , quickly poured, 6 g/L of Na_2HAsO_4 , 7 H_2O dissolve in a small volume of water, before adjusting to 1 liter. Solution D is then put in an opaque bottle, and heated up for 3 hours at 55°C.

The sample was put in a deep-well, to which was added 100 μL of a solution C. The deep-well was then heated up to 95°C in a water bath for 15 minutes. After cooling down the deep-well, 100 μL of solution D was added. After shaking for 10 minutes, 1 mL of ultrapure water was added. Take 250 μL to put in a microplate and read the OD 600 nm. The standard used was xylose, at concentration 0 to 250 $\mu\text{g}/\text{mL}$.

Quantification of total proteins present in the samples was carried out using the Bradford method (Kruger, 1994). The samples were diluted by five, and 20 μL of these dilutions were mixed with 200 μL of Bradford reagent. After 20 minutes, the OD was measured at 595 nm. The standard used was BSA.

Enzymatic cocktails were prepared by mixing 250 μL of enzymatic extract from each selected clone. The enzymatic reaction medium contained 250 μL of PBS pH 7.0 and 250 μL of the mix, along with the equivalent of 90 μL of wheat straw, as for activity measurement of individual clones. The theoretical activity of the enzymatic cocktails was calculated as follows: mean of the sum of activities of all the clones included in the mix (μg free sugar carbohydrates/ml of reaction) divided by the mean of the amount of proteins in the mix (mg proteins / mL reaction).

The real synergy activity was calculated as follows: activity of the mix (μg free sugar carbohydrates/ml of reaction) divided by the mean of the amount of total proteins in the mix (mg proteins / mL reaction).

3.5.3. Oxidoreductase activity

The clones were grown at 37°C in 20 mL LB medium, with orbital shaking at 120 r.p.m.. After 16 h, cells were harvested by centrifuging for 5 min at 5,000 r.p.m., and re-suspended in activity buffer to obtain a final $\text{OD}_{600\text{nm}}$ of 80. Cell lysis was done using sonication. Cell debris were centrifuged at 13,000 r.p.m. for 10 min and the cytoplasmic extracts were filtered with a 0.20 μm Minisart RC4 syringe filter. Enzymatic reactions were carried out at 30°C, by adding 5mM ABTS as a colored substrate, in 96 wells microtiter plate assays. Each well contained 20 μL of concentrated cytoplasmic extract, 5 mM ABTS, 0, 10 mM

CuSO₄, 0 or 3% H₂O₂ and 100 mM sodium citrate buffer pH 4.5. Enzymatic activity was determined by following absorbance increase at 420 nm during 30 min, measured using a microplate spectrophotometer (BioTek™ Eon™ Microplate Spectrophotometers, Colmar, France). ABTS oxidizing activity was expressed as μmol/min/L_{culture}, using an ABTS extinction coefficient value of $\epsilon_{\text{ABTS}, 420\text{nm}} = 36,000 \text{ M}^{-1} \cdot \text{cm}^{-1}$.

3.6. Sequencing and data analysis

Fosmid DNA of the positive clones was extracted with the NucleoBond Xtra Midi kit from Macherey-Nagel (France) following supplier recommendations. Fosmids were then sequenced with the MiSeq technology, at the Genotoul platform (<http://get.genotoul.fr/>). Read assembly was performed using Masurca (<http://www.genome.umd.edu/masurca.html>). The contigs were cleaned from the vector pCC1FOS sequence using Crossmatch (<http://bozeman.mbt.washington.edu/phredphrapconsed.html>). ORF detection and functional annotation was performed using the RAST software (Aziz et al., 2008).

CAZyme encoding genes were identified by BLAST analysis of the predicted ORFs against the sequences included in the CAZy database (<http://www.cazy.org>) using a cut-off E-value of $7 \cdot 10^{-6}$ followed by visual inspection and alignment with known CAZy families.

The annotation of ORFs corresponding to target functions (CAZymes, esterases, oxidoreductases) were manually verified, based on BLASTP analysis against the NCBI UniProtKBSwiss-Prot database.

The protein signatures were detected using the InterProScan software (Jones et al., 2014).

The most probable taxonomical origin of the contigs was retrieved using MEGAN v3.2.1 (Huson et al., 2007), based on ORF BLASTX analysis against the non-redundant NCBI protein database (E-value $\leq 10^{-8}$, identity $\geq 90\%$, query length coverage $\geq 50\%$). Contigs were assigned to a phylum, order, class, genus or species only if at least 50% of their ORFs were assigned to the same level.

4. Conclusion:

In the present work, activity-based metagenomics was used to explore the functional potential of the bacterial consortia issued from the bovine ruminal ecosystem enriched *in vivo* and *in vitro* on wheat straw. A multi-step high-throughput screening strategy was developed to sort out metagenomic clones that are able to break down polysaccharidic and

polyaromatic PCW components, and to select those that are the most efficient to degrade crude plant biomass, like wheat straw particles. This generic approach, which is applicable to any natural or artificial microbiome, allowed us to highlight novel multigenic clusters involved in PCW degradation, issued from highly prevalent uncultivated ruminal bacteria that are specialized in the breakdown of hemicelluloses, cellulose, and maybe also lignin. The structure-function relationships of the most original glycoside-hydrolases, carbohydrate esterases and oxidases encoded by these loci, in particular those containing still uncharacterized structural modules, will be further studied in order to unravel their specific role in PCW catabolism.

Acknowledgments

This research was funded by the French Ministry of Education and Research (Ministère de l'Enseignement supérieur et de la Recherche), the INRA metaprogramme M2E (project Metascreen). The high throughput screening work was carried out at the Laboratory for BioSystems & Process Engineering (Toulouse, France) with the equipments of the ICEO facility, dedicated to the screening and the discovery of new and original enzymes. ICEO is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund and the Institut National de la Recherche Agronomique (INRA).

Artwork

Figure 31: Activity of the IVVE hit clones on wheat straw.	227
Figure 32: Activity of the IVTE hit clones on wheat straw..	228
Figure 33: Activity of the different lignocellulolytic mixes on wheat straw.	234

Table content

Table 11: Number of hits obtained from the primary screening of both metagenomic libraries.	223
Table 12: Number of hits obtained from secondary screening.	226
Table 13: List of the clones included in the lignocellulase cocktails, their functional properties and CAZome, deduced from sequence analysis.	231

Supplementary data

Table S1: Functional and taxonomical assignation of the sequenced metagenomic clones. In the 1st column, contig number in red are the ones used in cocktails; in the 3rd column, in dark pink, are the taxonomic problems between two contigs of the same clone; in the 4th column, substrates in red are the ones without ORFs corresponding to the activity; in the 6th column,

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

ORFs in orange are the one identified for the activity; putative clusters are highlighted in grey and brown.

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

clone ID	library	contig size Taxonomic assignment	Activities	ORFs	functional annotation (RAST)	Begins	Stop	Strand	CAZy annotation	Redundancy between CAZy sequences	Best BLASTP hit against Uniprot/Swissprot: (% cov, e value, % id) accession number	annotation of unknown domains of interest (Interproscan)	Putative function of the multigenic cluster
1C10	IVVE	1 (32,580 bp) <i>Feocalibacterium prausnitzii</i> (Firmicutes; Clostridia; Clostridiales; Ruminococcaceae)	AZO- galactomannan, pNPP	1	hypothetical protein	53	1762	-					pectin degradation (Common structural features of the side chains include polymeric (1,4)beta- linked D-galactosyl and (1,5)alpha- linked L-arabinosyl residues. Willats et al 2001)
				2	N-Acetyl-D-glucosamine ABC transport system, permease protein 2	1822	2826	-					
				3	binding-protein-dependent transport systems inner membrane component	2847	3773	-					
				4	hypothetical protein	3770	6361	-					
				5	FIG01031734: hypothetical protein	6376	7008	-					
				6	NHL repeat containing protein	7005	8486	-			RecName: Full=E3 ubiquitin- protein ligase TRIM71; AltName: Full=Protein lin-41 homolog; AltName: Full=Tripartite motif- containing protein 71 (18%, 7e-09, 35%) E7FAM5.1		
				7	binding-protein-dependent transport systems inner membrane component	8498	9487	-					
				8	binding-protein-dependent transport systems inner membrane component	9484	10416	-					
				9	N-Acetyl-D-glucosamine ABC transport system, sugar-binding protein	10418	13402	-					
				10	Mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78)	13439	14911	-	CBM35-GH26	14019 (11): 100% cov, 99% id	RecName: Full=Mannan endo-1,4-beta-mannosidase; AltName: Full=Endo-(1,4)- beta-mannanase (96%, 3e-79, 35%) P49425.3		
				11	Polygalacturonase// <i>rhamno</i>	15231	16637	-	GH28	14019 (12): 100% cov, 98% id			
				12	Putative isomerase// <i>glc et mann</i>	16653	18083	-	GH63_dist/GH*_ GH63	14019 (13): 100% cov, 99% id	65-404 six-hairpin glycosidase-like		
				13	ABC transporter, permease protein	18323	19276	+					
				14	ABC transporter, permease protein	19293	20225	+					
				15	Transcriptional regulator, AraC family// <i>xylosidase</i>	20240	21145	+	GH*_GH39		30-107 RmlC-like cupin domain // 199-297 DNA binding HTH domain, AraC- type		
				16	ABC transporter, substrate-binding protein	21207	22712	+					
				17	ABC transporter, substrate-binding protein	22829	24376	+					
				18	Transcriptional regulator, AraC family	24380	26647	+					
				19	putative glycosyl hydrolase, family 43// <i>Xyl et Ara</i>	26657	27499	+	GH43_20	14019 (20): 100% cov, 99% id			
				20	Rhamnulokinase (EC 2.7.1.5)	27546	28826	+					
				21	L-rhamnose isomerase (EC 5.3.1.14)	28823	30067	+					
				22	Rhamnulose-1-phosphate aldolase (EC 4.1.2.19)	30075	30911	+					
				23	Alpha-galactosidase (EC 3.2.1.22)// <i>glc</i>	30908	32305	+	GH4_3	1C10 (24): 100% cov, 100% id			
				24	hypothetical protein	32406	32570	-					
												heteroxylans (small amounts of oligo- and poly- saccharides containing alpha- linked galactose are present in many pasture plants. (Bailey 1962))	

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

12F23	IVVE	1 (25,530 bp) Bacterium (ES)	AZCL-xylan, AZCL-arabinoxylan, wheat straw	3	Cytochrome c-type biogenesis protein Ccs1/ResB	729	1757	-									
				4	Fumarate reductase flavoprotein subunit (EC 1.3.99.1)	1800	4724	-									
				5	Electron transport complex protein RnfG	4762	5940	-									
				6	Electron transport complex protein RnfB	5953	6747	-									
				7	Electron transport complex protein RnfA	6762	7340	-									
				8	Electron transport complex protein RnfE	7337	8095	-									
				9	Electron transport complex protein RnfD	8097	9638	-									
				10	Electron transport complex protein RnfC	9635	10984	-									
				11	Sensory box/GGDEF family protein	11020	12921	-									
				12	hypothetical protein	13193	14566	+									
				13	Lysophospholipase (EC 3.1.1.5); Monoglyceride lipase (EC 3.1.1.23); putative	14617	15501	-									
				14	Malate permease	15679	16608	+									
				15	conserved hypothetical protein 374	16611	17753	-									
				16	Neopullulanase (EC 3.2.1.135) // amylopectine	17814	19889	-	GH13	32D12 (33): 100% cov, 99% id; 38D22 (27): 100% cov, 99% id; 51D10 (3): 100% cov, 95% id							
				17	Beta-1,3-glucosyltransferase	19925	20746	-	GT2	32D12 (14): 100% cov, 99% id; 38D22 (28): 100% cov, 99% id							
				18	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	20895	22409	-	missing signal peptide; CBM22-GH10	38D22 (29): 100% cov, 99% id; 32D12 (35): 100% cov, 99% id	RecName: Full=Endo-1,4-beta-xylanase A; Short=Xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase A; AltName: Full=Endoxylanase; Flags: Precursor (98%, 7e-72, 32%) Q60042.1						
				19	hypothetical protein	22395	22508	+									
				20	Alcohol dehydrogenase (EC 1.1.1.1)	22612	25479	-									
				13M17	IVVE	1 (30,963 bp) Firmicute	Tween 20	1	hypothetical protein	41	742	+					
								2	ATPase	739	2247	+					
3	Beta-glucosidase (EC 3.2.1.21)	2332	3642					+	GH1_13	41L9 (9): 100% cov, 97% id							
4	hypothetical protein	3657	4520					+									
5	benzyl alcohol dehydrogenase	4623	5717					-									
6	hypothetical protein	5959	6864					-									
7	hypothetical protein	6885	7265					-									
8	hypothetical protein	7262	7705					-									
9	cobalamin synthesis protein, P47K	7686	8750					-									
10	Aromatic hydrocarbon utilization transcriptional regulator CatR (LysR family)	9122	10009					+									
11	Xanthine dehydrogenase iron-sulfur subunit (EC 1.17.1.4) / Xanthine dehydrogenase, molybdenum binding subunit (EC 1.17.1.4)	10124	12949					+									
12	Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	13025	15127					+									
13	hypothetical protein	15148	15324					+									
14	2-Hydroxy-6-Oxo-6-Phenylhexa-2,4-Dienoate hydrolase	15396	16169					-									
15	Esterase/lipase	16191	17096					-			RecName: Full=Monoterpene epsilon-lactone hydrolase (75%, 2e-06, 21%) Q9EX73.1						
16	hypothetical protein	17241	18578					-									
17	hypothetical protein	18863	22315					+									
18	transcriptional regulator, LysR family	22353	23258					-									
19	hypothetical protein	23433	24071					+									
20	Deblocking aminopeptidase (EC 3.4.11.-)	24523	25221					+									
21	amidohydrolase	25218	26507					+									
22	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)	26461	28059					+									
23	Dipeptide transport system permease protein DppB (TC 3.A.1.5.2)	28062	29012					+									
24	Dipeptide transport system permease protein DppC (TC 3.A.1.5.2)	29005	29868					+									
25	Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)	29879	30898					+									

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

16810	IVVE	1 (34,263 bp) Bacterium (ES)	AZCL-xylan, AZCL-arabinoxylan, wheat straw	1	binding-protein-dependent transport systems inner membrane component	32	898	+						
				2	hypothetical protein	913	2445	+						
				3	hypothetical protein	2414	3106	+						
				4	hypothetical protein	3122	5695	+						
				5	binding-protein-dependent transport systems inner membrane component	5692	6672	+						
				6	Alpha-glucoside transport system permease protein AglG	6665	7675	+						
				7	hypothetical protein	7684	7803	+						
				8	ABC transporter substrate-binding protein - sugar transport	7831	9273	+						
				9	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	9412	10968	+	CBM22-GH10			RecName: Full=Endo-1,4-beta-xylanase A; Short=Xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase A; Flags: Precursor (94%, 2e-66, 32%) Q60037.1		hemicellulose or cellulose degradation
				10	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	11167	12330	+	GH5_4		RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=EGCCA; AltName: Full=Endo-1,4-beta-glucanase A; Flags: Precursor (90%, 1e-99, 45%) P17901.1			
				11	Chitobiose phosphorylase (EC 2.4.1.-) //cellobiose phosphorylases	12435	14822	+	GH94-X91		279-782 Six-hairpin glycosidase-like:			
				12	COG1720: Uncharacterized conserved protein	14948	15610	+						
				13	Phosphoserine aminotransferase (EC 2.6.1.52)	15812	16888	+						
				14	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	17068	18243	+						
				15	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)	18368	21937	-						
				16	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)	21958	25824	-						
				17	DNA internalization-related competence protein ComEC/Rec2	26073	27059	+						
				18	Firmicutes ribosomal L7Ae family protein	27231	27476	+						
				19	SSU ribosomal protein S12p (S23e)	27541	27915	+						
				20	SSU ribosomal protein S7p (S5e)	28126	28596	+						
				21	Translation elongation factor G	28696	30762	+						
				22	Translation elongation factor Tu	30832	32046	+						
				23	LSU ribosomal protein L33p	32068	32229	+						
				24	hypothetical protein	32257	32547	+						
				25	Transcription antitermination protein NusG	32565	33101	+						
				26	LSU ribosomal protein L11p (L12e)	33248	33673	+						

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

17H1	IVVE	1 (39,405 bp) Firmicute	Tween 20, pNPA, pNPB, pNPP	1	hypothetical protein	530	694	+										
				2	hypothetical protein	691	849	+										
				3	tRNA proofreading protein STM4549	868	1347	-										
				4	hypothetical protein	1589	1816	+										
				5	ABC transporter, ATP-binding protein	1861	2598	+										
				6	ABC transporter permease protein	2591	4222	+										
				7	Lipopolysaccharide biosynthesis protein-like protein	4282	5937	-										
				8	Carbamoylphosphate synthase large subunit (split gene in MI)	5985	7169	-										
				9	hypothetical protein	7174	7998	-										
				10	hypothetical protein	8143	8625	-										
				11	Cystathionine beta-lyase, Bsu PatB (EC 4.4.1.8)	8693	9874	-										
				12	hypothetical protein	9920	10684	-	CE*					RecName: Full=Monoacylglycerol lipase ABHD12; AltName: Full=Abhydrolase domain- containing protein 12 (85%, 2e-06, 25%) Q08C93.1	35-232 alpha/beta hydrolase fold-5			
				13	Branched-chain amino acid transport protein AzIC	10842	11534	+										
				14	Branched-chain amino acid transport protein azID	11531	11854	+										
				15	Dehydrogenase	12102	13253	-										
				16	Transcriptional regulator, MerR family	13464	14195	-										
				17	Uridine kinase family protein YggC homolog	14283	15101	-										
				18	spermidine acetyltransferase	15098	15580	-										
				19	FIG0116832: hypothetical protein	15630	15950	-										
				20	Putative Nudix hydrolase YfcD (EC 3.6.-.-)	15980	16516	-										
				21	Mg- protoporphyrin O-methyltransferase (EC 2.1.1.11)	16589	17518	-										
				22	hypothetical protein	17825	18226	+										
				23	response regulator	18361	21465	-										
				24	Cell surface protein	21620	22837	-										
				25	hypothetical protein	22834	23241	-										
				26	hypothetical protein	23330	23842	+										
				27	Manganese-dependent inorganic pyrophosphatase (EC 3.6.1.1)	23864	25024	-										
				28	acetyltransferase, GNAT family	25034	26077	-										
				29	Aminoglycoside phosphotransferase	26108	26905	-										
				30	hypothetical protein	26954	27118	-										
				31	hypothetical protein	27382	27861	-										
				32	hypothetical protein	27868	28662	-										
				33	probable methyltransferase	29050	29622	+										
				34	Metal transporter, ZIP family	29718	30497	+										
				35	hypothetical protein	30559	30729	-										
				36	Sensory box histidine kinase/response regulator	30847	34113	-										
				37	hypothetical protein	34199	34396	-										
				38	hypothetical protein	34448	34588	-										
				39	DNA-binding response regulator, AraC family	34578	35531	-	GH*_GH39							26-123 AraC-type arabinose- binding/dimerisation domain // 180-278 DNA binding HTH, AraC-type	Xylan/heteroxylan degradation	
				40	hypothetical protein	35528	35746	-										
				41	Xylose isomerase (EC 5.3.1.5)	35705	36742	+										
				42	Ribokinase (EC 2.7.1.15)	36929	38032	+										
				43	Fructose-bisphosphate aldolase class II (EC 4.1.2.13)	38143	38988	+										
				44	hypothetical protein	39064	39270	+										

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

21K7	IVVE	1 (13,661 bp) Firmicute (F+ES)	AZCL-xylan, AZCL arabinoxytan, pNPA	1	Sodium:alanine symporter	266	1906	-					
				2	hypothetical protein	1997	2116	-					
				3	Probable transcriptional regulator	2123	3076	+					
				4	hypothetical protein	3191	4477	+					
				5	FIG01032461: hypothetical protein	4499	5236	+					
				6	FIG01032596: hypothetical protein	5278	5949	+					
				7	hypothetical protein	5974	7236	+					
				8	hypothetical protein	7229	7423	+					
				9	hypothetical protein	7447	7581	-					
				10	hypothetical protein	7614	8531	-					
				11	Alpha-galactosidase (EC 3.2.1.22)	8582	9889	-	GH4_3				
				12	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	10181	11734	+	CBM22-GH10		RecName: Full=Endo-1,4-beta-xylanase B; Short=Xylanase B; AltName: Full=1,4-beta-D-xylan xylanohydrolase B (64%, 2e-65, 40%) P26223.1		
				13	hypothetical protein	11704	13425	-					
25I16	IVVE	1 (37,436 bp) Bacterium (ES)	AZCL-xylan, AZCL arabinoxytan, Tween 20, pNPA, pNPB, wheat straw	1	Na+ driven multidrug efflux pump	13	1116	+					
				2	Transposase	1591	2886	+					
				3	no significant homology	3054	3266	-					
				4	Homoserine kinase (EC 2.7.1.39)	3337	3993	-					
				5	Rubredoxin	4058	4210	-					
				6	Uncharacterized protein MJ1614	4371	5207	+					
				7	hypothetical protein	5209	6930	+					
				8	ABC-type polysaccharide transport system, permease component	7185	8159	+					
				9	Predicted beta-xyloside ABC transporter, permease component	8175	9047	+					
				10	hypothetical protein	9106	10716	+					
				11	Transcriptional regulator, AraC family	10834	11760	-					
				12	Oxidoreductase	11931	13091	+					
				13	oxidoreductase domain protein	13093	14172	+					
				14	Gluconokinase (EC 2.7.1.12) / oxidoreductase domain	14176	15357	+					
				15	Phage repressor	15504	16256	+					
				16	hypothetical protein	16269	17660	+					
				17	hypothetical protein	17678	17944	+					
				18	putative esterase	17950	18735	-	CE1	45P14 (37), 50E19 (14): 100% cov, 100% id	RecName: Full=Acetyl esterase; AltName: Full=Acetylxylosidase (99%, 3e-68, 43%) P23553.1		
				19	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	18777	19832	-	GH10	45P14 (37), 50E19 (15): 100% cov, 100% id	RecName: Full=Endo-1,4-beta-xylanase B; Short=Xylanase B; AltName: Full=1,4-beta-D-xylan xylanohydrolase B (97%, 3e-81, 42%) P26223.1		xylan degradation
				20	Acetyl-coenzyme A synthetase (EC 6.2.1.1)	19914	22187	-					
21	DNA mismatch repair protein [MutS]	22641	24356	+									
22	hypothetical protein	24483	26360	+									
23	Ribosome protection-type tetracycline resistance related proteins, group 2	26389	29022	-									
24	hypothetical protein	29192	30460	+									
25	hypothetical protein	30465	31160	+									
26	hypothetical protein	31258	33780	+									
27	ABC transporter, ATP-binding protein	33767	34705	+									
28	hypothetical protein	34698	35615	+									
29	hypothetical protein	35636	37135	+									

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

32D12	IVVE	1 (33,736 bp) Bacterium (ES)	Lignin alkali, wheat straw	7	Endoglucanase E precursor (EC 3.2.1.4) (EgE) (Endo-1,4-beta-glucanase E) (Cellulase E)	6	830	+	fragment N-term; CE2	38D22 (1): 100% cov, 100% id		
				8	Substrate-specific component RibU of riboflavin ECF transporter	1108	1707	+				
				9	Fibronectin/fibrinogen-binding protein	1727	3466	+				
				10	hypothetical protein	3516	4067	-				
				11	Phosphoglycolate phosphatase (EC 3.1.3.18)	4113	4763	-				
				12	2-oxoglutarate oxidoreductase, delta subunit, putative (EC 1.2.7.3)	4935	5138	+			RecName: Full=Uncharacterized ferredoxin M10146 (95%, 8e-10, 39%) Q57610.1	
				13	2-oxoglutarate oxidoreductase, alpha subunit (EC 1.2.7.3)	5154	6230	+			RecName: Full=Ketoisovalerate oxidoreductase subunit VorB; Short=VOR; AltName: Full=2-oxoisovalerate ferredoxin reductase subunit beta; AltName: Full=2-oxoisovalerate oxidoreductase beta chain (98%, 1e-118, 50%) P80908.2	
				14	2-oxoglutarate oxidoreductase, beta subunit (EC 1.2.7.3)	6233	6970	+			RecName: Full=Ketoisovalerate oxidoreductase subunit VorA; Short=VOR; AltName: Full=2-oxoisovalerate oxidoreductase alpha chain; AltName: Full=2-oxoisovalerate-ferredoxin oxidoreductase subunit alpha (95%, 3e-86, 54%) P80907.2	
				15	2-oxoglutarate oxidoreductase, gamma subunit (EC 1.2.7.3)	6974	7510	+			RecName: Full=Ketoisovalerate oxidoreductase subunit VorA; Short=VOR; AltName: Full=2-oxoisovalerate oxidoreductase alpha chain; AltName: Full=2-oxoisovalerate-ferredoxin oxidoreductase subunit alpha (97%, 1e-33, 42%) P80907.2	
				16	putative xylanase	7591	8808	-				
				17	Cytochrome c-type biogenesis protein CcsA/ResC	8825	9640	-				
				18	Cytochrome c-type biogenesis protein Ccs1/ResB	9649	10677	-				
				19	Fumarate reductase flavoprotein subunit (EC 1.3.99.1)	10720	13644	-				
				20	hypothetical protein	13682	14839	-				
				21	Electron transport complex protein RnfB	14852	15646	-				
				22	Electron transport complex protein RnfA	15661	16239	-				
				23	Electron transport complex protein RnfE	16236	16994	-				
				24	Electron transport complex protein RnfD	16996	18537	-				
				25	Electron transport complex protein RnfC	18534	19859	-				
				26	Sensory box/GGDEF family protein	19919	21820	-				
				27	hypothetical protein	22092	23465	+				
				28	Lysophospholipase (EC 3.1.1.5); Monoglyceride lipase (EC 3.1.1.23); putative	23516	24439	-				
				29	hypothetical protein	24419	24556	+				
				30	Maltase permease	24579	25508	+				
				31	hypothetical protein	25511	26416	-				
				32	hypothetical protein	26413	26652	-				
				33	Neopullulanase (EC 3.2.1.135)	26713	28788	-	GH13	38D22 (27): 100% cov, 100% id; 12F23 (16): 100% cov, 99% id; 51D10 (3): 100% cov, 95% id		
				34	Beta-1,3-glucosyltransferase	28824	29645	-	GT2	38D22 (27): 100% cov, 100% id; 12F23 (17): 100% cov, 99% id		
				35	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	29793	31307	-	CBM22-GH10	38D22 (29): 100% cov, 99% id; 12F23 (18): 100% cov, 99% id	RecName: Full=Endo-1,4-beta-xylanase A; Short=Xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase A; AltName: Full=Endoxylanase; Flags: Precursor (98%, 3e-70, 32%) Q60042.1	
				36	hypothetical protein	31293	31406	+				
				37	Collagen-binding A protein-like	31510	33693	-				

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

33M21	IVVE	1 (35,197 bp) Firmicute	AZCL-Barley beta glucan, AZCL-HE cellulose, pNPA, wheat straw	11	Adenosylcobinamide-phosphate guanylyltransferase (EC 2.7.7.62)	81	617	+							
				12	Cobalamin synthase	614	1366	+							
				13	hypothetical protein	1363	1731	+							
				14	hypothetical protein	1728	2309	+							
				15	Cob(I)alamin adenosyltransferase (EC 2.5.1.17)	2311	2832	+							
				16	hypothetical protein	2863	3234	+							
				17	FIGO22979: MoxR-like ATPases	3262	4206	+							
				18	hypothetical protein	4203	5270	+							
				19	hypothetical protein	5286	6506	+							
				20	Phosphoglycerate mutase family 5	6752	7450	-							
				21	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	7472	8401	-	missing signal peptide; GH5_2				RecName: Full=Endoglucanase 4; AltName: Full=Cellulase 4; AltName: Full=Endo-1,4-beta glucanase 4; AltName: Full=Endoglucanase IV; Short=EG-IV (93%, 5e-101, 50%) Q07940.1		
				22	hypothetical protein	8426	9832	-							
				23	Beta-galactosidase (EC 3.2.1.23)	9895	11832	-	GH42						
				24	hypothetical protein	11991	13547	-							
				25	hypothetical protein	13653	14348	-							
				26	hypothetical protein	14368	14733	-							
				27	hypothetical protein	14891	15124	+							
				28	hypothetical protein	15127	15333	+							
				29	hypothetical protein	15330	15710	+							
				30	hypothetical protein	15715	16011	+							
				31	hypothetical protein	16061	16192	+							
				32	hypothetical protein	16255	16437	+							
				33	hypothetical protein	17045	17299	+							
				34	hypothetical protein	17292	17651	+							
				35	Protein export cytoplasm protein SecA ATPase RNA helicase (TC 3.A.5.1.1)	17664	18479	+							
				36	hypothetical protein	18472	19017	+							
				37	hypothetical protein	19087	19560	+							
				38	hypothetical protein	19572	20216	+							
				39	hypothetical protein	20254	21120	+							
				40	Single-stranded DNA-binding protein	21135	21536	+							
				41	hypothetical protein	21551	21937	+							
				42	hypothetical protein	21934	22338	+							
				43	hypothetical protein	22340	22600	+							
				44	hypothetical protein	22593	23132	+							
				45	hypothetical protein	23117	23332	+							
				46	hypothetical protein	23311	23517	+							
				47	phage Gp37Gp68 family protein	23486	24337	-							
				48	hypothetical protein	24342	24551	+							
				49	hypothetical protein	24548	24823	+							
				50	hypothetical protein	25045	25233	+							
				51	hypothetical protein	25223	25417	+							
				52	hypothetical protein	25513	26031	+							
				53	hypothetical protein	26167	26781	+							
				54	hypothetical protein	26778	27296	+							
				55	Phage terminase, large subunit	27299	29158	+							
				56	hypothetical protein	29181	29423	+							
				57	hypothetical protein	29440	31092	+							
				58	Prophage Clp protease-like protein	31085	32218	+							
				59	hypothetical protein	32245	32901	+							
				60	Putative capsid protein of prophage	32937	33974	+							
				61	hypothetical protein	33995	34345	+							
				62	hypothetical protein	34342	34500	+							
				63	hypothetical protein	34500	34844	+							

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

38D22	IVVE	1 (33,736 bp) Bacterium (ES)	AZCL-xylan, AZCL-arabinoxylan, pNPA, pNPB, wheat straw	1	Endoglucanase E precursor (EC 3.2.1.4) (EgE) (Endo-1,4-beta-glucanase E) (Cellulase E)	6	830	+	fragment N-term; CE2	32D12 (7): 100% cov, 100% id	none			
				2	Substrate-specific component RibU of riboflavin ECF transporter	1108	1707	+						
				3	Fibronectin/fibrinogen-binding protein	1727	3466	+						
				4	hypothetical protein	3516	4067	-						
				5	Phosphoglycolate phosphatase (EC 3.1.3.18)	4113	4763	-						
				6	2-oxoglutarate oxidoreductase, delta subunit, putative (EC 1.2.7.3)	4935	5138	+						
				7	2-oxoglutarate oxidoreductase, alpha subunit (EC 1.2.7.3)	5154	6230	+						
				8	2-oxoglutarate oxidoreductase, beta subunit (EC 1.2.7.3)	6233	6970	+						
				9	2-oxoglutarate oxidoreductase, gamma subunit (EC 1.2.7.3)	6974	7510	+						
				10	putative xylanase	7591	8808	-						
				11	Cytochrome c-type biogenesis protein CcsA/ResC	8825	9640	-						
				12	Cytochrome c-type biogenesis protein Ccs1/ResB	9649	10677	-						
				13	Fumarate reductase flavoprotein subunit (EC 1.3.99.1)	10720	13644	-						
				14	hypothetical protein	13682	14839	-						
				15	Electron transport complex protein RnfB	14852	15646	-						
				16	Electron transport complex protein RnfA	15661	16239	-						
				17	Electron transport complex protein RnfE	16236	16994	-						
				18	Electron transport complex protein RnfD	16996	18537	-						
				19	Electron transport complex protein RnfC	18534	19883	-						
				20	Sensory box/GGDEF family protein	19919	21820	-						
				21	hypothetical protein	22092	23465	+						
				22	Lysophospholipase (EC 3.1.1.5); Monoglyceride lipase (EC 3.1.1.23); putative	23516	24400	-						
				23	hypothetical protein	24380	24556	+						
				24	Malate permease	24546	25508	+						
				25	hypothetical protein	25511	26416	-						
				26	hypothetical protein	26413	26652	-						
				27	Neopullulanase (EC 3.2.1.135)	26713	28788	-	GH13	32D12 (33): 100% cov, 100% id; 12F23 (16): 100% cov, 99% id; 51D10 (3): 100% cov, 95% id				
				28	Beta-1,3-glucosyltransferase	28824	29645	-	GT2	32D12 (14): 100% cov, 100% id; 12F23(17): 100% cov, 99% id				
				29	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	29793	31307	-	missing signal peptide; CBM22-GH10	32D12 (35): 100% cov, 99% id; 12F23 (18): 100% cov, 99% id	RecName: Full=Endo-1,4-beta-xylanase A; Short=Xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase A; AltName: Full=Endoxylanase; Flags: Precursor (98%, 2e-69, 32%) Q60042.1			
				30	hypothetical protein	31293	31406	+						
				31	Collagen-binding A protein-like	31510	33693	-						

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

39F1	IVVE	1 (24,784 bp) Bacterium (ES+F)	AZCL-xylan, AZCL-arabinoxylan, Tween 20, pNPA, pNPB	1	Phosphopentomutase (EC 5.4.2.7)	376	2358	-															
				2	Segregation and condensation protein B	2444	2995	-															
				3	Segregation and condensation protein A	2982	3737	-															
				4	FIG004556: membrane metalloprotease	3742	4497	-															
				5	hypothetical protein	4565	4702	+															
				6	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	4728	6275	-		CBM22-GH10													
				7	ABC transporter substrate-binding protein - sugar transport	6468	7937	-															
				8	binding-protein-dependent transport systems inner membrane component	8098	9108	-															
				9	binding-protein-dependent transport systems inner membrane component	9101	10087	-															
				10	hypothetical protein	10084	12660	-															
				11	hypothetical protein	12675	13379	-															
				12	FIG01032054: hypothetical protein	13348	14895	-															
				13	binding-protein-dependent transport systems inner membrane component	14945	16171	-															
				14	binding-protein-dependent transport systems inner membrane component	16173	17120	-															
				15	N-Acetyl-D-glucosamine ABC transport system, sugar-binding protein	17143	20034	-															
				16	Pyrroline-5-carboxylate reductase (EC 1.5.1.2)	20443	21246	-															
				17	hypothetical protein	21417	22265	+															
				18	hypothetical protein	22452	22622	-															
				19	COG0613, Predicted metal-dependent phosphoesterases (PHP family)	22711	23421	-															
				20	FIG176548: DRTGG domain-containing protein	23423	23770	-															
41L9	IVVE	1 (44,123 bp) Clostridiales (Firmicute)	Tween 20, pNPA, pNPB, pNPP	1	Formate dehydrogenase-O, major subunit (EC 1.2.1.2)	1589	3988	-															
				2	Xanthine dehydrogenase, molybdenum binding subunit (EC 1.17.1.4)	4006	6312	-															
				3	Xanthine dehydrogenase iron-sulfur subunit (EC 1.17.1.4)	6314	6829	-															
				4	hypothetical protein	6858	7397	-															
				5	Methyltransferase	7592	8473	+															
				6	Cell surface protein	8542	9768	+															
				7	FIG00262634: hypothetical protein	9765	11237	+															
				8	FIG01077095: hypothetical protein	11234	12742	+															
				9	Beta-glucosidase (EC 3.2.1.21)	12827	14137	+		GH1				13M17 (3): 100% cov, 97% id									
				10	Uncharacterized conserved protein	14152	15015	+															
				11	benzyl alcohol dehydrogenase	15102	16196	-															
				12	CTP:molybdopterin cytidyltransferase	16427	17356	-															
				13	hypothetical protein	17353	17733	-															
				14	hypothetical protein	17730	18143	-															
				15	cobalamin synthesis protein, P47K	18154	19230	-															
				16	Aromatic hydrocarbon utilization transcriptional regulator CatR (LysR family)	19589	20476	-															
				17	Xanthine dehydrogenase iron-sulfur subunit (EC 1.17.1.4) / Xanthine dehydrogenase, molybdenum binding subunit (EC 1.17.1.4)	20591	23416	+															
				18	Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	23423	25594	+															
				19	hypothetical protein	25626	25802	-															
				20	2-Hydroxy-6-Oxo-6-Phenylhexa-2,4-Dienoate hydrolase	25892	26662	-															
				21	Esterase/lipase	26684	27589	-															
				22	Phenylacetate-coenzyme A ligase (EC 6.2.1.30)	27734	29071	-															
				23	hypothetical protein	29392	32844	+															
				24	transcriptional regulator, LysR family	32883	33788	-															
				25	hypothetical protein	33963	34403	+															
				26	HNH endonuclease	34743	36230	+															
				27	hypothetical protein	36507	37688	-															
				28	Mobile element protein	38041	39429	+															
				29	hypothetical protein	39401	39532	-															
				30	hypothetical protein	39668	40918	+															
				31	Mobile element protein	41085	42311	-															
				32	Mobile element protein	42521	42760	+															
				33	hypothetical protein	42781	43176	+															

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

5013	IVVE	1 (34,579 bp) Firmicute (ES)	Tween 20, pNPA, pNPB, wheat straw	1	Pili retraction protein pIT	30	596	+										
				2	N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28)	694	2115	+										
				3	Mobile element protein	2289	3752	+										
				4	tmRNA-binding protein SmpB	4299	4772	-										
				5	Gamma-glutamyl phosphate reductase (EC 1.2.1.41)	5008	6282	+										
				6	hypothetical protein	6414	6620	-										
				7	hypothetical protein	6624	6911	-										
				8	hypothetical protein	7199	7354	-										
				9	hypothetical protein	7528	7779	+										
				10	putative rRNA methylase	8278	9120	+										
				11	hypothetical protein	9194	11386	-										
				12	hypothetical protein	11922	12767	-										
				13	hypothetical protein	12786	13688	-										
				14	Cys-tRNA(Pro) deacylase YbaK	13772	14257	-										
				15	hypothetical protein	14351	15724	-										
				16	N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25)	15853	17001	-	CE9									
				17	glycoside hydrolase, family 43	17218	18744	-	GH43_C/X19					RecName: Full=Putative beta-xylosidase; AltName: Full=1,4-beta-D-xylan xylohydrolase; AltName: Full=Xylan 1,4-beta-xylosidase (69%, 2e-33, 32%) Q9WXE8.2	312-504 Concanavalin A-like lectin/glucanase domain			
				18	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	18707	21388	-	GH10-GH43_D-CBM6_2					RecName: Full=Arabinoxylan arabinofuranohydrolase; Short=AXH; AltName: Full=AXH-m2,3; Short=AXH-m23; AltName: Full=Alpha-L-arabinofuranosidase; Short=AF; Flags: Precursor (51%, 4e-128, 46%) P45796.1				
				19	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	21508	22581	-	CE1					RecName: Full=Carbohydrate acetyl esterase/feruloyl esterase; Includes: RecName: Full=Carbohydrate acetyl esterase; Includes: RecName: Full=Feruloyl esterase; AltName: Full=Ferulic acid esterase; Flags: Precursor (98%, 4e-51, 34%) D5EXZ4.1			hemicellulose degradation	
				20	putative esterase	22756	23589	-	CE1					RecName: Full=Acetyl esterase; AltName: Full=Acetylxylosidase (94%, 3e-37, 33%) P23553.1				
				21	oxidoreductase domain protein	23773	24873	-										
				22	Predicted rhamnose oligosaccharide ABC transport system, permease component	25077	25919	-										
				23	Predicted rhamnose oligosaccharide ABC transport system, permease component 2	25937	26863	-										
				24	Predicted rhamnose oligosaccharide ABC transport system, substrate-binding component	26967	28301	-										
				25	Oxidoreductase	28565	29719	-										
				26	lipoprotein	29814	31406	-										
				27	binding-protein-dependent transport systems inner membrane component	31614	32528	-										
				28	Xylose ABC transporter, permease component	32542	33486	-										
				29	Neuraminidase NanP	33599	34561	-										

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

51D10	IVVE	1 (9,817 bp) Bacterium (ES)	AZCL-xylan, AZCL arabinoxylan, wheat straw	1	hypothetical protein	709	1638	+				
				2	Integral membrane protein	1641	2786	-				
				3	Neopullulanase (EC 3.2.1.135)	2849	4924	-	GH13	32D12 (33): 100% cov, 95% id; 12F23 (16): 100% cov, 95% id; 38D22 (27): 100% cov, 95% id		
				4	Beta-1,3-glucosyltransferase	4961	5782	-	GT2		RecName: Full=Uncharacterized glycosyltransferase EpsJ (79%, 3e-39, 37%) P71059.1	
				5	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	5933	7459	-	missing signal peptide; CBM22- GH10		RecName: Full=Endo-1,4- beta-xylanase A; Short=Xylanase A; AltName: Full=1,4-beta-D-xylan xylanohydrolase A; AltName: Full=Endoxylanase; Flags: Precursor (97%, 3e-69, 32%) Q60042.1	
		6		hypothetical protein	7650	9800	-					
		7		Glycosyltransferase	270	1331	+	GT2_18-TM- UNK(57aa)			no annotation	
		8		hypothetical protein	1522	3423	+					
		9		hypothetical protein	3459	5852	+					
		51I19		IVVE	1 (33,096 bp) Firmicute	AZO- galactomannan, wheat straw	1	COG0501: Zn-dependent protease with chaperone function	217	1299	-	
2	hydrolase, putative / (ADP-ribosylation/Crystallin J1)		1363				2178	-				
3	COG2110, Macro domain, possibly ADP-ribose binding module		2265				3191	-				
4	probable secreted protein		3311				5131	-				
5	hypothetical protein / (RNA polymerase archaeal subunit P/eukaryotic subunit RPABC4)		5389				6072	-				
6	hypothetical protein		6084				6755	-				
7	HTH-type transcriptional activator tipA		6757				7197	-				
8	hypothetical protein		7312				7614	-				
9	hypothetical protein		7604				7894	-				
10	Transcriptional regulator, MerR family		7885				8298	-				
11	Fe-S oxidoreductase		8373				9482	-				
12	hypothetical protein / (Zinc finger, Sec23/Sec24-type)		9572				10579	-				
13	hypothetical protein		10723				11865	+				
14	N-acyl-L-amino acid amidohydrolase (EC 3.5.1.14)		11973				13265	-				
15	Transcriptional regulator, AraC family		13285				14184	-				
16	Prolidase (EC 3.4.13.9)		14219				15520	-				
17	N-acyl-L-amino acid amidohydrolase (EC 3.5.1.14)		15722				17002	+				
18	Short-chain fatty acids transporter		17033				18439	+				
19	N-acyl-L-amino acid amidohydrolase (EC 3.5.1.14)		18467				19747	+				
20	Transcriptional regulator		19824				23063	-				
21	potassium channel protein		23080				24573	-				
22	hypothetical protein		24616				25515	-				
23	N-acetyl-L-L-diaminopimelate deacetylase (EC 3.5.1.47)		25635				26816	+				
24	Acyl-phosphate:glycerol-3-phosphate O-acyltransferase PlsY		26882				27523	+				
25	unknown / (Bacterial Pleckstrin homology domain)		27547				27936	-				
26	hypothetical protein / (Protein of unknown function DUF3990)		28019				28549	-				
27	Probable tesA-like protease / (hydrolase GDSL [Thermosinus carboxydivorans])		28737				29291	-				
28	hypothetical protein / (yrosine/serine-protein phosphatase lphP-type)		29428				30669	+				
29	hypothetical protein		31008				32570	-				
30	hypothetical protein		32766				32990	-				

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

52E22	IVVE	1 (35,389 bp) Firmicute	AZO-galactomannan, pNPB, wheat straw	1	hypothetical protein	417	1334	-							
				2	hypothetical protein	1336	1623	-							
				3	Ribosyl nicotinamide transporter, PnuC-like	1620	2312	-							
				4	Nicotinamide-nucleotide adenyltransferase, NadR family (EC 2.7.7.1) / Ribosylnicotinamide kinase (EC 2.7.1.22)	2305	3324	-							
				5	COG2152 predicted glycoside hydrolase	3533	4558	-	GH130_2				RecName: Full=β-1,4-mannooligosaccharide phosphorylase; AltName: Full=RaMP2 (96%, 3e-120, 54%) E6UBR9.1		
				6	ABC transporter substrate-binding protein - sugar transport	4732	6102	-							
				7	hypothetical protein	6189	7922	-							
				8	N-Acetyl-D-glucosamine ABC transport system, permease protein 2	7985	8956	-							
				9	binding-protein-dependent transport systems inner membrane component	8971	9891	-							
				10	hypothetical protein	9888	12485	-							
				11	FIG01031734: hypothetical protein	12502	13137	-							
				12	NHL repeat containing protein / (gluconolactonase [Lachnodostridium phytofermentans])	13134	14609	-					RecName: Full=E3 ubiquitin-protein ligase TRIM71; AltName: Full=Protein lin-41 homolog; AltName: Full=Tripartite motif-containing protein 71 (61%, 3e-09, 23%) E7FAM5.1		mannan degradation
				13	binding-protein-dependent transport systems inner membrane component	14622	15551	-							
				14	binding-protein-dependent transport systems inner membrane component	15553	16485	-							
				15	N-Acetyl-D-glucosamine ABC transport system, sugar-binding protein	16487	19489	-							
				16	Mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78)	19515	20996	-	CBM35-GH26				RecName: Full=Mannan endo-1,4-beta-mannosidase; AltName: Full=Endo-(1,4)-beta-mannanase (91%, 6e-70, 33%) P49425.3		
				17	hypothetical protein	21391	21594	+							
				18	hypothetical protein / (gDSL-like Lipase/Acylhydrolase [Eubacterium eligens CAG:72])	21877	22686	-					none		
				19	hypothetical protein	22727	23929	-							
				20	hypothetical protein	23993	24151	+							
				21	Phage protein	24205	25533	-							
				22	hypothetical protein	25785	25931	+							
				23	hypothetical protein	26230	26760	-							
				24	Glycogen synthase, ADP-glucose transglucosylase (EC 2.4.1.21)	26870	28345	-	GT5 (probably too short at N-term)						
				25	hypothetical protein	28356	28538	-							
				26	Uridine kinase family protein	28628	29410	+							
				27	hypothetical protein	29559	30704	+							
				28	hypothetical protein	30717	31187	+							
				29	hypothetical protein	31184	31819	+							
				30	hypothetical protein	31861	32052	+							
				31	hypothetical protein	32077	32844	+							
				32	hypothetical protein	32875	33657	+							
				33	hypothetical protein	33694	33984	+							
				34	GCN5-related N-acetyltransferase	34001	34447	+							
				35	hypothetical protein	34515	34706	+	GH3 (short fragment, too short at C-term)				none		
				36	hypothetical protein	34810	35031	+							

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

3F21	IVTE	1 (30,213 bp) Bacteroidales (Bacteroidetes)	AZCL-Barley beta glucan, wheat straw	1	Acyltransferase	3083	4321	+	GH16_Iam			
				2	FIG053235: Diacylglycosamine hydrolase like	4468	4971	+				
				3	Beta-glucosidase (EC 3.2.1.21)	5109	7262	-	GH3	4101 (9): 96% cov, 99% id	RecName: Full=Periplasmic beta-glucosidase; AltName: Full=Beta-D-glucoside glucohydrolase; AltName: Full=Celllobiase; AltName: Full=Gentiobiase; AltName: Full=T-cell inhibitor; Flags: Precursor (91%, 1e-179, 41%) Q56078.2	cellulose degradation
				4	Beta-glucanase precursor (EC 3.2.1.73)	7275	8324	-				
				5	hypothetical protein	8308	8442	-				
				6	FIG00412289: hypothetical protein	8446	9810	-				
				7	RagB/SusD domain protein	9831	11300	-				
				8	hypothetical protein	11254	11373	+				
				9	TonB family protein / TonB-dependent receptor	11374	14385	-				
				10	FIG00417103: hypothetical protein	14689	17538	-				
				11	Fe-S OXIDOREDUCTASE (1.8.-.-)	17691	18398	-				
				12	FIG00898434: hypothetical protein	18528	20591	-				
				13	hypothetical protein	20602	21285	-				
				14	FIG00896812: hypothetical protein	21518	22123	+				
				15	23S rRNA methyltransferase (EC 2.1.1.-)	22115	22645	-				
				16	Quinolinate synthetase (EC 2.5.1.72)	23167	24165	+				
				17	Immunoreactive 84kD antigen PG93	24169	26457	-				
				18	Nucleoside 5-triphosphatase RdgB (dHATP, dITP, XTP-specific) (EC 3.6.1.15)	26514	27089	-				
				19	putative transporter	27168	28043	-				
				20	Leucyl-tRNA synthetase (EC 6.1.1.4)	28043	30211	-				
6F24	IVTE	1 (21,494 bp) Prevotella (Bacteroidetes; Bacteroidia; Bacteroidales)	AZO-CM cellulose, AZCL- HE Cellulose, wheat straw	1	hypothetical protein	1027	2538	-				
				2	hypothetical protein	2568	3572	-				
				3	hypothetical protein	3586	3792	-				
				4	hypothetical protein	3841	4011	-				
				5	transposase	4169	5239	+				
				6	FIG00937925: hypothetical protein	5405	6115	-				
				7	Ynd	6108	6869	-				
				8	hypothetical protein	7408	7665	-				
				9	hypothetical protein	7640	7756	-				
				10	Hypothetical Nudix-like regulator	7910	8665	+				
				11	hydrolase, putative	8668	9465	+				
				12	Mobile element protein	9875	11110	-				
				13	Beta-glucosidase (EC 3.2.1.21)	11571	13883	-	GH3	30D14 (6): 100%cov, 100%id	RecName: Full=Beta-glucosidase BoGH3B; AltName: Full=Glycosyl hydrolase family protein 3B; Short=BoGH3B; Flags: Precursor (99%, 0.0, 58%) A7LXU3.1	
				14	hypothetical protein	13883	14623	-				
				15	FIG00937602: hypothetical protein	14762	16504	-				
				16	FIG01289191: hypothetical protein	16616	19795	-				
				17	DNA mismatch repair protein precursor (EC 3.2.1.4);B-glucanase	19849	21096	-	GH55		RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=EGCCA; AltName: Full=Endo-1,4-beta glucanase A; Flags: Precursor (95%, 5e-61, 33%) P17901.1	
				18	FIG00938042: hypothetical protein	21306	21458	-				

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

Sample ID	Genome	Genome Size (bp)	Phylum	Class	Order	Family	Genus	Species	Accession	Coverage	Identity	Gene Name	EC Number	Function	Enzyme Commission	Substrate	Product	Notes					
																			Cellulose degradation	Hemicellulose degradation			
10D6	IVTE	1 (21,977 bp) <i>Bacteroides uniformis</i> (Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides)	AZO-CM cellulose, AZO Galactomannan, wheat straw	1	hypothetical protein	104	265	+															
				2	Putative ion-channel protein	289	1293	-															
				3	Cellobiose phosphorylase (EC 2.4.1.-)	1312	3798	-														cellulose degradation	
				4	Beta-galactosidase (EC 3.2.1.23)	4022	6466	+															
				5	Transcriptional regulator of various polyols utilization, AraC family	6792	7658	+															
				6	FIG00937556: hypothetical protein	7779	9014	+															
				7	Endo-1,4-beta-mannosidase	9106	10398	+															
				8	Endoglucanase 5A (EC 3.2.1.4) (Endo-1,4-beta-glucanase 5A) (Alkaline cellulase	10556	11482	+															
				9	hypothetical protein	11546	11659	+															
				10	TonB family protein / TonB-dependent receptor	11717	14983	+															
				11	putative outer membrane protein, probably involved in nutrient binding	15004	16725	+															
				12	FIG00417257: hypothetical protein	16730	18043	+															
				13	FIG00410329: hypothetical protein	18065	20050	+															
				14	Mannan endo-1,4-beta-mannosidase B precursor (EC 3.2.1.78)	20118	21815	+															
2		2 (5,173 bp) Bacteroidales (Bacteroidetes)		1	hypothetical protein	403	1125	-															
				2	Di-/tripeptide transporter	1125	2687	-															
				3	Lipid A biosynthesis lauroyl acyltransferase (EC 2.3.1.-)	2786	3700	+															
				4	hypothetical protein	3712	4722	+															

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

1	hypothetical protein	38	664	+					
2	hypothetical protein	690	1625	+					
3	TonB family protein / TonB-dependent receptor	2265	5618	+					
4	FIG00937915: hypothetical protein	5627	7282	+					
5	FIG00938925: hypothetical protein	7301	8206	+					
6	hypothetical protein	8218	9012	+					
7	hypothetical protein	9019	10353	+					
8	hypothetical protein	10710	12101	+					
9	metallo-beta-lactamase family protein	12083	12892	+					
10	Cystathionine beta-lyase, Bsu PatB (EC 4.4.1.8)	12889	13116	-					
11	HIT family protein	13321	13722	-					
12	Transcription elongation factor GreA	13728	14192	-					
13	Aspartate aminotransferase (EC 2.6.1.1)	14409	15716	-					
14	Indolepyruvate oxidoreductase subunit IorB (EC 1.2.7.8)	15697	16275	-				RecName: Full=Indolepyruvate oxidoreductase subunit IorB; Short=IOR; AltName: Full=Indolepyruvate ferredoxin oxidoreductase subunit beta (98%, 1e-31, 36%) O58496.1	
15	Indolepyruvate oxidoreductase subunit IorA (EC 1.2.7.8)	16279	17931	-				RecName: Full=Indolepyruvate oxidoreductase subunit IorA; Short=IOR; AltName: Full=Indolepyruvate ferredoxin oxidoreductase subunit alpha (99%, 1e-109, 38%) O28783.1	
16	Aliphatic amidase AmiE (EC 3.5.1.4)	17910	18692	-					
17	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11)	18733	19743	-					
18	Probable exoglucanase a (1,4-beta-cellobiosidase) protein (EC 3.2.1.91)	20198	25573	+				RecName: Full=CaM kinase- like vesicle-associated protein (5%, 1e-11, 41%) Q8NCB2.2	
19	hypothetical protein	25580	26773	+					
20	hypothetical protein	26791	28212	+					
21	hypothetical protein	28260	28559	+					
22	Na ⁺ /H ⁺ antiporter	29313	31316	+					
23	Lipoprotein releasing system ATP-binding protein LolD	31345	32043	+					
24	FIG00938655: hypothetical protein	32070	32936	+					
25	llyltransferase (EC 2.5.1.1) / (2E,6E)-farnesyl diphosphate synthase (EC 2.5.1.10)	33028	33924	+					
26	Deoxyribose-phosphate aldolase (EC 4.1.2.4)	33986	34927	-					
27	D-tyrosyl-tRNA(Tyr) deacylase (EC 3.6.1.n1)	34945	35397	-					
28	Excinuclease ABC subunit C	35394	37256	-					
29	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	37299	39194	-					
30	nitroreductase	39303	39917	-					
31	Fumarate hydratase class I, anaerobic (EC 4.2.1.2)	40220	41872	+					
32	Ribosome recycling factor	41902	42465	+					
33	Ribosome small subunit-stimulated GTPase EngC	42473	43402	+					
34	hypothetical protein	43788	44300	-					
35	Glutamine synthetase type III, GlnN (EC 6.3.1.2)	44544	46541	-					

10N13

IVTE

1
(46,551 bp)
Bacteroidales
(Bacteroidetes)

ABTS, wheat
straw

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

Sample ID	Genome	Gene ID	Gene Name	EC	Accession	Length (bp)	GC (%)	Signal	Family	Coverage	Identity	Description	Notes	
														Function
12E13	IVTE	1 (13,214 bp) Bacteroides graminisolvens (Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroidaceae; Bacteroides)	AZO-galactomannan, AZCL-debranched arabinan	1	Alpha-N-arabinofuranosidase 2 (EC 3.2.1.55)	15	929	+	GH51					
				2	Transketolase (EC 2.2.1.1)	999	3011	+						
				3	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	3113	3544	+						
				4	Beta-xylosidase (EC 3.2.1.37)	3663	5552	-	GH43_4					
				5	Putative outer membrane protein, probably involved in nutrient binding	5622	7421	-						
				6	TonB-dependent receptor	7446	10580	-						
				7	putative outer membrane protein, probably involved in nutrient binding	10674	12440	-						
				8	TonB-dependent receptor	12446	13165	-						
			2 (1,071 bp) Bacteroides uniformis (Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides)		1	putative DNA-binding protein	318	1505	+					
					2	hypothetical protein	1548	1661	+					
					3	hypothetical protein	1828	2472	-					
					4	FIG00416167: hypothetical protein	2621	3490	-					
					5	putative outer membrane protein, probably involved in nutrient binding	3505	5274	-					
					6	SusC, outer membrane protein involved in starch binding	5287	8535	-					
					7	Mannan endo-1,4-beta-mannosidase B precursor (EC 3.2.1.78)	8941	10638	-	GH26	10D6 (14): 100% cov, 100% id	RecName: Full=Mannan endo-1,4-beta-mannosidase; AltName: Full=Endo-(1,4)-beta-mannanase (55%, 1e-35, 41%) P49425.3	hemicellulose degradation	
					13F22	IVTE	1 (24,657 bp) Bacteroidetes Bacteroides graminisolvens (Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides)	AZCL-arabinoxylan, wheat straw	1	putative outer membrane protein, probably involved in nutrient binding	4	1998	+	
2	FIG00938343: hypothetical protein	2041	3837	+										
3	FIG00939126: hypothetical protein	3868	4902	+										
4	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	4965	7139	+					GH10a-X-X-GH10b (X distantly related to CBM4)		RecName: Full=Endo-1,4-beta-xylanase C; Short=Xylanase C; AltName: Full=1,4-beta-D-xylan xylanohydrolase C; Flags: Precursor (52%, 1e-17, 29%) Q0H904.2			
5	hypothetical protein	7146	7274	-										
6	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	7228	9450	+					GH10-CE1		RecName: Full=Endo-1,4-beta-xylanase/feruloyl esterase; Includes: RecName: Full=Endo-1,4-beta-xylanase; Includes: RecName: Full=Feruloyl esterase; AltName: Full=Ferulic acid esterase; Flags: Precursor (95%, 0.0, 64%) D5EY13.1			
7	DNA-binding response regulator, AraC family	9595	13485	-										
8	Endoglucanase D precursor (EC 3.2.1.4)	13889	16414	+					GH9_dist		RecName: Full=Xyloglucan-specific endo-beta-1,4-glucanase BoGH9A; AltName: Full=Glycosyl hydrolase family protein 9A; Short=BoGH9A; Flags: Precursor (69%, 6e-06, 21%) A7LXT3.1	hemicellulose degradation		
9	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	16459	17583	+					CE1		RecName: Full=Carbohydrate acetyl esterase/feruloyl esterase; Includes: RecName: Full=Carbohydrate acetyl esterase; Includes: RecName: Full=Feruloyl esterase; AltName: Full=Ferulic acid esterase; Flags: Precursor (90%, 9e-75, 40%) D5EXZ4.1			
10	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	17638	20157	+					GH43_2-CBM6-GH8 (a bit too short at N-term)		RecName: Full=Reducing end xylose-releasing exo-oligoxylanase; Short=Rex (47%, 3e-88, 43%) Q9KB30.1			
11	Hypothetical glycoside hydrolase, family 43, similar to arabinosidase	20221	21183	-					GH43_35 (a bit too short at N-term)		RecName: Full=Arabinoxylan arabinofuranohydrolase; Short=AXH; AltName: Full=AXH-m2,3; Short=AXH-m2,3; AltName: Full=Alpha-L-arabinofuranosidase; Short=AF; Flags: Precursor (99%, 8e-35, 30%) Q45071.2			
		2 (4,084 bp) Bacteroidetes		12	Pyridoxamine 5'-phosphate oxidase (EC 1.4.3.5)	21674	22165	+						
				13	hypothetical protein	22187	22603	-						
				14	transcriptional regulator	22852	23058	+						
				1	SusD, outer membrane protein	1282	3024	-						
				2	putative outer membrane protein, probably involved in nutrient binding	3043	3864	-						

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

Sample ID	Phylum	Genus	Strain	Gene ID	Gene Size		GC Content	Signal	Gene Name	EC Number	Enzyme Class	Enzyme Activity	Enzyme Source	Enzyme Description			
					Start	End											
14J10	IVTE	1 (27,156 bp) Lachnospiraceae (Firmicutes; Clostridia; Clostridiales)	AZCL-Barley beta glucan, AZO-CM cellulose, Tween 20	1	32	625	+										
				2	644	1156	+										
				3	1179	1565	+										
				4	1754	3049	+										
				5	3190	4122	+										
				6	4174	5907	-										
				7	6107	7135	+				GHS_37						RecName: Full=Cellodextrinase A (100%, 2e-104, 46%) P16169.3
				8	7463	8911	+										
				9	8913	9482	+										
				10	9707	14236	+										
				11	14253	15716	+										
				12	15803	17020	+										
				13	17041	17622	+										
				14	17661	18488	-				CE*_CE1 (esterase)						RecName: Full=Acetyl esterase; AltName: Full=Acetylxylosidase (93%, 3e-58, 39%) P23553.1
				15	18687	20168	+										
				16	20178	20741	+										
				17	20871	22973	+										
				18	23031	24617	+										
				19	24829	25860	+										
				20	25844	26437	+										
15F15	IVTE	1 (1,164 bp) Firmicute	AZCL-Barley beta glucan, AZO-CM cellulose, AZCL Tamarind xyloglucan, AZCL HE cellulose, wheat straw	1	127	270	+										
				2	425	994	+										
				3	1370	2146	-										
				4	2146	2676	-										
				5	2705	2968	-										
				6	1140	3074	+										
				7	3046	4215	-										
				8	4248	4871	-										
				9	4881	6245	-				GHS_4	% cov, 100% id; 44C20 (4) 100%					RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=EGCCA; AltName: Full=Endo-1,4-beta glucanase A; Flags: Precursor (81%, 2e-96, 43%) P17901.1
				10	6299	7171	-										
		11	7183	9264	-												
		12	9292	10020	-												
		13	10066	11103	-												
		14	11103	11627	-												
		15	30	266	+												
		16	421	1866	+												
		17	2060	3382	+												
		18	3394	4275	+												
		19	4283	5428	+												
		20	6053	7747	+												
21	7809	9311	-														
22	9340	9486	+														
23	9527	10453	-														
24	10622	10792	-														
25	11135	11305	+														
26	11558	11692	-														
27	11912	12238	+														
28	12309	12548	+														
29	12558	13001	+														
30	13003	13710	+														
31	13745	14071	+														
32	14097	15809	+														
33	15814	16242	+														

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

20G16	IVTE	1 (35,118 bp) Bacteroidetes Bacteroidales	AZO-galactomannan, wheat straw	1	putative glycosyltransferase	10	870	-	GT14				
				2	Succinoglycan biosynthesis protein	871	1950	-					
				3	Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)	1930	2634	-					
				4	Cytoplasmic axial filament protein CafA and Ribonuclease G (EC 3.1.4.-)	2652	4223	-					
				5	DNA-binding protein HU	4452	4724	-					
				6	ATP-dependent protease La (EC 3.4.21.53) Type I	5010	7352	+					
				7	Uracil-DNA glycosylase, family 1	7352	8020	+					
				8	Phosphopantetheine adenylyltransferase (EC 2.7.7.3)	8017	8490	+					
				9	hypothetical protein	8492	8710	-					
				10	Thiol:disulfide interchange protein tlpA	8714	9961	+					
				11	includes chaperone protein YajL (former ThiJ), parkinsonism-associated protein	9950	10522	-					
				12	yl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-alanine ligase (EC 6.3.2.10) / A	10632	11744	-					
				13	probable metal-dependent peptidase	12103	12777	+					
				14	hypothetical protein	12774	13067	-					
				15	tRNA-specific adenosine-34 deaminase (EC 3.5.4.-)	13084	13518	-					
				16	hypothetical protein	13639	13869	-					
				17	Transcriptional regulator, AraC family	13954	14526	+					
				18	Endo-1,4-beta-mannosidase	14696	17095	+	GH5_7	0% cov, 100% id; 44D8 (13): 100%	RecName: Full=Mannan endo-1,4-beta-mannosidase 1; AltName: Full=Beta-mannanase 1; AltName: Full=Endo-beta-1,4-mannanase 1; AltName: Full=OsMAN1; Flags: Precursor (51%, 5e-49, 32%) Q0JKM9.2	mannan degradation	
				19	TonB family protein / TonB-dependent receptor	17153	20332	+					
				20	putative outer membrane protein, probably involved in nutrient binding	20345	22093	+					
				21	FIG00417257: hypothetical protein	22106	23410	+					
				22	hypothetical protein	23417	25843	+					
				23	Cellulase (EC:3.2.1.4)	25938	27464	+	GH5_4	0% cov, 100% id; 44D8 (18): 100%	RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=Endo-1,4-beta-D-glucanase A; Flags: Precursor (77%, 2e-54, 32%) O08342.1		
				24	Mannan endo-1,4-beta-mannosidase	27578	28609	+			RecName: Full=Mannan endo-1,4-beta-mannosidase; AltName: Full=Mannanase 26A; Short=Man26A; AltName: Full=Mannanase A; Short=ManA; Flags: Precursor (86%, 3e-41, 33%) P49424.2		
				25	COG2152 predicted glycoside hydrolase	28667	29842	+	GH130_1	0% cov, 100% id; 44F17 (12): 99% cov, 100% id	RecName: Full=4-O-beta-D-mannosyl-D-glucose phosphorylase; Short=MGP; Short=Mannosylglucose phosphorylase (98%, 0.0, 72%) Q5LH68.1		
				26	Xyloside transporter XynT	29863	31290	+					
				27	DNA polymerase I (EC 2.7.7.7)	31846	34494	+					
				28	RNA polymerase sigma factor RpoE	34466	35071	+					

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

21E20	IVTE	1 (37,465 bp) <i>Paenirhodobacter enshiensis</i> (Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae)	lignin alkali	1	OmpA family outer membrane protein	2606	3292	-				cellulose degradation
				2	Glucosamine-6-phosphate deaminase (EC 3.5.99.6)	3374	5296	-				
				3	hypothetical protein	5535	6131	-				
				4	putative outer membrane protein, probably involved in nutrient binding	6214	7995	-				
				5	SusC, outer membrane protein involved in starch binding	8002	11127	-				
				6	Hypothetical sugar kinase, ROK family	11212	12045	-				
				7	Xylose operon regulatory protein	12133	13290	-				
				8	Beta-hexosaminidase (EC 3.2.1.52)	13394	15634	-				
				9	Alkaline phosphatase (EC 3.1.3.1)	15667	18771	-	GH2_exoGn (over 190 residues missing at N-term)			
				10	Small protein containing transglutaminase-like domain	18870	19316	-				
				11	Lipid-A-disaccharide synthase (EC 2.4.1.182)	19343	20497	-	GT19			
				12	Protein of unknown function DUF1009 clustered with KDO2-Lipid A biosynthesis gene	20494	21312	-				
				13	UDP-N-acetylglucosamine O-acyltransferase (EC 2.3.1.1)	21309	22103	-				
				14	3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59)	22135	22602	-				
				15	Outer membrane protein H precursor	22745	23458	-				
				16	Outer membrane protein assembly factor YaeT precursor	23467	25848	-				
				17	Membrane-associated zinc metalloprotease	26009	27379	-				
				18	1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267)	27383	28606	-				
				19	Phosphatidate cytidylyltransferase (EC 2.7.7.41)	28608	29420	-				
				20	Undecaprenyl diphosphate synthase (EC 2.5.1.31)	29417	30169	-				
				21	Ribosome recycling factor	30185	30748	-				
				22	Uridine monophosphate kinase (EC 2.7.4.22)	30868	31599	-				
				23	tRNA dimethylallyltransferase (EC 2.5.1.75)	31681	32580	+				
				24	Transcriptional regulator, AraC family	32774	33535	+				
				25	hypothetical protein	33545	33913	-				
				26	Aldehyde dehydrogenase (EC 1.2.1.3)	34122	35651	-		RecName: Full=Acetaldehyde dehydrogenase 2; AltName: Full=Acetaldehyde dehydrogenase II; Short=ACDH-II (97%, 0.0, 73%) P46368.1		
				27	hypothetical protein	35706	35867	+				
				28	Succinate semialdehyde dehydrogenase [NAD] (EC 1.2.1.24); Succinate-semialdehyde dehydrogenase	35988	37379	-		RecName: Full=Succinate-semialdehyde dehydrogenase [NAD(+)]; Short=SSDH (97%, 9e-152, 49%) Q32507.1		

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

			1	Chromate transporter	33	506	+					
			2	probable chromate transport protein, putative	503	1078	+					
			3	putative integral membrane transport protein	1307	2515	+					
			4	Sugar phosphate isomerases/epimerases	2560	3453	+					
			5	hypothetical protein	3802	3933	+					
			6	Putative phosphate ABC transporter, phosphate-binding component	4003	4836	-					
			7	Ferric siderophore transport system, periplasmic binding protein TonB	4839	5654	-					
			8	Biopolymer transport protein ExbD/TolR	5676	6332	-					
			9	Biopolymer transport protein ExbD/TolR	6347	6952	-					
			10	MotA/TolQ/ExbB proton channel family protein	6978	7790	-					
			11	hypothetical protein	7910	8311	-					
			12	hypothetical protein	11678	11797	+					
			13	hypothetical protein	14289	14420	+					
			14	Probable poly(beta-D-mannuronate) O-acetylase (EC 2.3.1.-)//Aminobenzoate degradation : Ethylbenzene degradation	14832	16247	-				RecName: Full=Peptidoglycan O- acetyltransferase; Short=PG acetyltransferase (75%, 2e-80, 38%) O25526.1	
			15	putative periplasmic protein	16288	17133	-					
			16	putative periplasmic protein	17180	18439	-					
			17	FIG00936253: hypothetical protein	18618	18911	-					
			18	cobalt-zinc-cadmium resistance protein CzCA; Cation efflux system protein CusA	18908	22066	-					
			19	Membrane fusion protein of RND family multidrug efflux pump	22079	23095	-					
			20	Outer membrane efflux protein precursor	23130	24467	-					
			21	Transcriptional regulator, TetR family	24502	25134	-					
			22	FIG00405371: hypothetical protein	25263	25895	-					
			23	LSU ribosomal protein L17p	26423	26914	-					
			24	DNA-directed RNA polymerase alpha subunit (EC 2.7.7.6)	26918	27910	-					
			25	SSU ribosomal protein S4p (S9e)	27922	28527	-					
			26	SSU ribosomal protein S11p (S14e)	28634	29023	-					
			27	SSU ribosomal protein S13p (S18e)	29035	29415	-					
			28	LSU ribosomal protein L36p	29449	29565	-					
			29	Translation initiation factor 1	29574	29792	-					
			30	Methionine aminopeptidase (EC 3.4.11.18)	29807	30598	-					
			31	Preprotein translocase secY subunit (TC 3.A.5.1.1)	30607	31956	-					
			32	LSU ribosomal protein L15p (L27Ae)	31961	32407	-					
			33	LSU ribosomal protein L30p (L7e)	32435	32611	-					
			34	SSU ribosomal protein S5p (S2e)	32623	33144	-					
			35	LSU ribosomal protein L18p (L5e)	33144	33488	-					
			36	LSU ribosomal protein L6p (L9e)	33511	34065	-					
			37	SSU ribosomal protein S8p (S15Ae)	34083	34478	-					
			38	ribosomal protein S14p (S29e) @ SSU ribosomal protein S14p (S29e), zinc-indepe	34531	34800	-					
			39	LSU ribosomal protein L5p (L11e)	34805	35362	-					
			40	LSU ribosomal protein L24p (L26e)	35365	35682	-					
			41	LSU ribosomal protein L14p (L23e)	35702	36067	-					
			42	SSU ribosomal protein S17p (S11e)	36070	36327	-					
			43	LSU ribosomal protein L29p (L35e)	36341	36535	-					
			44	LSU ribosomal protein L16p (L10e)	36541	36975	-					
			45	SSU ribosomal protein S3p (S3e)	36998	37723	-					
			46	LSU ribosomal protein L22p (L17e)	37735	38139	-					
			47	SSU ribosomal protein S19p (S15e)	38166	38435	-					
			48	LSU ribosomal protein L2p (L8e)	38455	39279	-					
			49	LSU ribosomal protein L23p (L23Ae)	39281	39571	-					
			50	LSU ribosomal protein L4p (L1e)	39584	40210	-					
			51	LSU ribosomal protein L3p (L3e)	40210	40827	-					
			52	SSU ribosomal protein S10p (S20e)	40848	41153	-					

26117

IVTE

1
(41,253 bp)
Bacteroidaceae (fam)
(Bacteroidetes;
Bacteroidia;
Bacteroidales)

ABTS, wheat
straw

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

28D5	IVTE	1 (17,825 bp) Bacteroidetes; Bacteroidia; Bacteroidales	AZO-galactomannan, wheat straw	1	hypothetical protein	167	2233	+																				
				2	Cellulase(EC:3.2.1.4)	2328	3854	+	X153-GH5_4 (a bit too short at N-term)	38A17 (23): 100% cov, 100% id	RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=Endo-1,4-beta-D-glucanase A; Flags: Precursor (77%, 2e-54, 32%) O08342.1																	
				3	Mannan endo-1,4-beta-mannosidase	3968	4999	+	GH26_3 (a bit too short at N-term)	d; 44D8 (19): 100% cov, 100% id	RecName: Full=Mannan endo-1,4-beta-mannosidase; AltName: Full=Mannanase 26A; Short=Man26A; AltName: Full=Mannanase A; Short=ManA; Flags: Precursor (86%, 3e-41, 33%) P49424.2																	
				4	COG2152 predicted glycoside hydrolase	5077	6252	+	GH130_1	id; 44F17 (12): 99% cov, 100% id	RecName: Full=4-O-beta-D-mannosyl-D-glucose phosphorylase; Short=MGP; Short=Mannosylglucose phosphorylase (98%, 0.0, 72%) Q5LH68.1																	
				5	Xyloside transporter XynT	6273	7700	+																				
				6	DNA polymerase I (EC 2.7.7.7)	8185	9375	+																				
				7	DNA polymerase I (EC 2.7.7.7)	9372	10916	+																				
				8	RNA polymerase sigma factor RpoE	10888	11493	+																				
				9	rRNA small subunit 7-methylguanosine (m7G) methyltransferase GidB	11490	12152	+																				
				10	ZINC PROTEASE (EC 3.4.99.-)	12631	15408	+																				
				11	iron (III) ABC transporter ATP-binding	15422	16081	-																				
				12	Vitamin B12 ABC transporter, permease component BtuC	16078	17079	-																				
				13	hypothetical protein	17088	17759	-																				
	2 (15,524 bp) Bacteroidetes; Bacteroidia; Bacteroidales				1	hypothetical protein	422	1354	+																			
					2	putative secretory protein	1869	2651	-	GH92 (approx 750 residues missing at N-term)		none																
					3	Alpha-1,2-mannosidase	2661	4937	-	GH92 (possibly a few residues too short at N-term)		RecName: Full=Uncharacterized glycosidase Rv0584; Flags: Precursor (97%, 8e-95, 32%) O86365.1																
					4	putative outer membrane protein, probably involved in nutrient binding	4990	6924	-																			
					5	TonB-dependent receptor	6972	10127	-																			
					6	DNA-binding response regulator, AraC family	10312	14082	-																			
					7	RNA methyltransferase, TrmA family	14192	14503	+																			

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

30D14	IVTE	1 (23,779 bp) Prevotella (genus) (Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae)	AZO-CM cellulose, wheat straw	1	FIG00938042: hypothetical protein	13	1584	+				
				2	DNA mismatch repair protein precursor (EC 3.2.1.4)	1791	3041	+	GH5_4		RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=EGCCA; AltName: Full=Endo-1,4-beta glucanase A; Flags: Precursor (98%, 3e-61, 32%) P17901.1	
				3	FIG01289191: hypothetical protein	3095	6274	+				
				4	FIG00937602: hypothetical protein	6386	8128	+				
				5	hypothetical protein	8267	9007	+				
				6	Beta-glucosidase (EC 3.2.1.21)	9007	11319	+	GH3	6F24 (13): 100% cov, 100% id	RecName: Full=Beta- glucosidase BoGH3B; AltName: Full=Glycosyl hydrolase family protein 3B; Short=BoGH3B; Flags: Precursor (99%, 0.0, 58%) A7LXU3.1	
				7	Mobile element protein	11780	13015	+				
				8	hydrolase, putative	13425	14222	-				
				9	Hypothetical Nudix-like regulator	14225	14980	-				
				10	hypothetical protein	15134	15250	-				
				11	hypothetical protein	15210	15482	+				
				12	Ynd	16102	16782	+				
				13	FIG00937925: hypothetical protein	16775	17485	+				
				14	transposase	17651	18721	-				
				15	hypothetical protein	19098	19304	+				
				16	hypothetical protein	19318	20322	+				
				17	hypothetical protein	20322	21863	+				
				18	hypothetical protein	22629	23006	+				

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

41J20	IVTE	1 (23,867 bp) Bacteroides graminisolvens (Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides)	AZO-galactomannan, wheat straw	1	FIG00410329: hypothetical protein	25	2022	+					
				2	Mannan endo-1,4-beta-mannosidase B precursor (EC 3.2.1.78)	2244	3524	+	GH26 (a bit too short at N-term)	RecName: Full=Mannan endo-1,4-beta-mannosidase; AltName: Full=Endo-(1,4)-beta-mannanase (80%, 6e-54, 35%) P49425.3		mannan degradation	
				3	Beta-glucosidase (EC 3.2.1.21)	3635	5917	+	GH3	RecName: Full=Periplasmic beta-glucosidase; AltName: Full=Beta-D-glucoside glucohydrolase; AltName: Full=Cellobiase; AltName: Full=Gentiobiase; AltName: Full=T-cell inhibitor; Flags: Precursor (94%, 8e-139, 34%) Q56078.2			
				4	Mannan endo-1,4-beta-mannosidase	6010	7044	+	GH26 (too short at N-term)	RecName: Full=Mannan endo-1,4-beta-mannosidase; AltName: Full=Mannanase 26A; Short=Man26A; AltName: Full=Mannanase A; Short=ManA; Flags: Precursor (99%, 2e-49, 33%) P49424.2			
				5	COG2152 predicted glycoside hydrolase	7077	8267	+	GH130_1	RecName: Full=4-O-beta-D-mannosyl-D-glucose phosphorylase; Short=MGP; Short=Mannosylglucose phosphorylase (98%, 0.0, 84%) Q5LH68.1			
				6	Cation symporter	8287	9720	+					
				7	N-acylglucosamine 2-epimerase (EC 5.1.3.8)	9717	10907	+					
				8	hypothetical protein	10914	11399	-					
				9	Ferredoxin	11507	12526	-					
				10	Transcriptional regulator, AraC family	12538	13347	-					
				11	Isochorismatase (EC 3.3.2.1)	13351	13908	-					
				12	Predicted nucleotide-binding protein	14921	16189	-					
				13	hypothetical protein	16203	16391	-					
				14	FIG00896557: hypothetical protein	16741	17868	-					
				15	FIG00403577: hypothetical protein	17980	20352	-					
				16	TonB-dependent receptor	20404	23046	-					
				17	NADH-ubiquinone oxidoreductase chain N (EC 1.6.5.3)	23311	23688	-					
		2	(4,322 bp) Lachnospiraceae (Firmicutes; Clostridia;	1	Transcriptional regulator, PadR family	86	424	+					
		2	hypothetical protein	435	1415	+							
		3	hypothetical protein	1425	2180	+							
		4	two-component response regulator	2233	4257	-							

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

4101	IVTE	1 (22,594 bp) Bacteroidetes; Bacteroidia; Bacteroidales	AZCL-Barley beta glucan, wheat straw	1	Osmosensitive K+ channel histidine kinase KdpD (EC 2.7.3.-)	347	1810	+						
				2	Two-component system response regulator	1807	2499	+						
				3	CRISPR-associated protein, Csn1 family	2724	7178	+						
				4	CRISPR-associated protein Cas1	7252	8193	+						
				5	CRISPR-associated protein Cas2	8235	8540	+						
				6	Acyltransferase	12475	12966	+						
				7	Acyltransferase	12957	13712	+						
				8	FIG053235: Diacylglycosamine hydrolase like	13859	14362	+						
				9	Beta-glucosidase (EC 3.2.1.21)	14403	16652	-	GH3		RecName: Full=Periplasmic beta-glucosidase; AltName: Full=Beta-D-glucoside glucohydrolase; AltName: Full=Cellobiase; AltName: Full=Gentiobiase; AltName: Full=T-cell inhibitor; Flags: Precursor (95%, 0.0, 42%) Q56078.2	cellulose degradation		
				10	Beta-glucanase precursor (EC 3.2.1.73)	16665	17762	-	GH16	RecName: Full=Beta-glucanase; AltName: Full=1,3-1,4-beta-D-glucan 4-glucanohydrolase; AltName: Full=Endo-beta-1,3-1,4 glucanase; AltName: Full=Lichenase; Flags: Precursor (63%, 4e-67, 46%) P45798.1				
				11	FIG00412289: hypothetical protein	17836	19200	-						
				12	RagB/SusD domain protein	19221	20645	-						
						13	hypothetical protein	20644	20763	+				
						14	TonB family protein / TonB-dependent receptor	20764	22581	-				
	2 (1,024 bp) Bacteroidetes; Bacteroidia; Bacteroidales	1	L-rhamnose-proton symporter	17	808	+				pectin degradation				
		2	Rhamnulose-1-phosphate aldolase (EC 4.1.2.19)	845	1639	+								
		3	putative glycosylhydrolase	1650	4871	+	GH106							
		4	Alfa-L-rhamnosidase (EC 3.2.1.40)	4916	6715	+	GH78	none						
		5	Alfa-L-rhamnosidase (EC 3.2.1.40)	6785	9319	+	GH78	none						

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

Sample ID	Phylum	Genome Size (bp)	Phylum	Gene ID	Gene Name	Length (bp)	Start	End	Start	End	GC (%)	Annotations	RecName	Flags				
44C18	IVTE	1 (17,351 bp) Firmicute	AZCL-Barley beta glucan, AZO-CM cellulose, AZCL Tamarind xyloglucan, AZCL HE cellulose, wheat straw	1	hypothetical protein	18	665	+										
				2	Alanine racemase (EC 5.1.1.1)	637	1821	-										
				3	hypothetical protein	1839	2462	-										
				4	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	2472	3836	-										RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=EGCCA; AltName: Full=Endo-1,4-beta-glucanase A; Flags: Precursor (81%, 2e-96, 43%) P17901.1
				5	Ribosome small subunit-stimulated GTPase EngC	3890	4762	-										
				6	Serine/threonine protein kinase PrkC, regulator of stationary phase	4774	6855	-										
				7	Protein serine/threonine phosphatase PrpC, regulation of stationary phase	6883	7611	-										
				8	Ribosomal RNA large subunit methyltransferase N (EC 2.1.1.-)	7657	8694	-										
				9	16S rRNA (cytosine(967)-C(5))-methyltransferase (EC 2.1.1.176)	8694	9026	-										
				10	Ribosomal RNA small subunit methyltransferase B (EC 2.1.1.-)	9026	10042	-										
				11	Methionyl-tRNA formyltransferase (EC 2.1.2.9)	10039	10983	-										
				12	Peptide deformylase (EC 3.5.1.88)	10983	11468	-										
				13	Helicase PriA essential for oriC/DnaA-independent DNA replication	11487	13709	-										
				14	Single-stranded-DNA-specific exonuclease RecJ (EC 3.1.-.-)	13755	15530	-										
				44C20	IVTE	1 (17,351 bp) Firmicute	AZCL-Barley beta glucan, AZO-CM cellulose, AZCL Tamarind xyloglucan, AZCL HE cellulose, wheat straw	1	hypothetical protein	18	665	+						
2	Alanine racemase (EC 5.1.1.1)	637	1821					-										
3	hypothetical protein	1839	2462					-										
4	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	2472	3836					-									RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=EGCCA; AltName: Full=Endo-1,4-beta-glucanase A; Flags: Precursor (81%, 2e-96, 43%) P17901.1	
5	Ribosome small subunit-stimulated GTPase EngC	3890	4762					-										
44C20	IVTE	2 (5,528 bp) Bacteroidetes; Bacteroidia; Bacteroidales	AZCL-Barley beta glucan, AZO-CM cellulose, AZCL Tamarind xyloglucan, AZCL HE cellulose, wheat straw	6	Serine/threonine protein kinase PrkC, regulator of stationary phase	4774	6855	-										
				7	Protein serine/threonine phosphatase PrpC, regulation of stationary phase	6883	7611	-										
				8	Ribosomal RNA large subunit methyltransferase N (EC 2.1.1.-)	7657	8694	-										
				9	16S rRNA (cytosine(967)-C(5))-methyltransferase (EC 2.1.1.176)	8694	9026	-										
				10	Ribosomal RNA small subunit methyltransferase B (EC 2.1.1.-)	9026	10042	-										
				11	Methionyl-tRNA formyltransferase (EC 2.1.2.9)	10039	10983	-										
				12	Peptide deformylase (EC 3.5.1.88)	10983	11468	-										
				13	Helicase PriA essential for oriC/DnaA-independent DNA replication	11487	13709	-										
				14	Single-stranded-DNA-specific exonuclease RecJ (EC 3.1.-.-)	13755	15530	-										
				15	membrane protein SecD (TC 3.A.5.1.1) / Protein-export membrane protein Sec	15534	17303	-										
44C20	IVTE	2 (5,528 bp) Bacteroidetes; Bacteroidia; Bacteroidales	AZCL-Barley beta glucan, AZO-CM cellulose, AZCL Tamarind xyloglucan, AZCL HE cellulose, wheat straw	1	putative outer membrane protein, probably involved in nutrient binding	25	2934	+										
				2	putative outer membrane protein, probably involved in nutrient binding	2971	4782	+										
				3	hypothetical protein	4779	4913	+										

Functional exploration of naturally and artificially enriched rumen microbiomes reveals novel enzymatic synergies involved in polysaccharide breakdown

45H9	IVTE	1 (14,385 bp) Firmicute Clostridiales (fam)	AZO-CM cellulose, wheat straw	1	widely conserved MoxR-like protein in magnesium chelatase family	286	1284	-						
				2	FIG028593: membrane protein	1397	2362	-						
				3	Guanine deaminase (EC 3.5.4.3)	2377	2835	-						
				4	hypothetical protein	2850	3833	-						
				5	Alanyl-tRNA synthetase (EC 6.1.1.7)	3850	5085	-						
				6	hypothetical protein	5381	5557	-						
				7	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)	5801	7261	-						
				8	hypothetical protein	7312	7827	-						
				9	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	7860	9128	-	GHS_4		RecName: Full=Endoglucanase A; AltName: Full=Cellulase A; AltName: Full=EGCCA; AltName: Full=Endo-1,4-beta glucanase A; Flags: Precursor (85%, 4e-109, 45%) P17901.1		cellulose degradation	
				10	Metallo-beta-lactamase family protein, RNA-specific	9137	10744	-						
				11	Transcription regulator [contains diacylglycerol kinase catalytic domain]	10981	11877	-						
				12	hypothetical protein	11890	12042	-						
				13	Pyruvate:ferredoxin oxidoreductase, beta subunit (EC 1.2.7.1)	12098	13036	-						
				14	Pyruvate:ferredoxin oxidoreductase, alpha subunit (EC 1.2.7.1)	13065	14246	-						
	1	Cof protein:HAD-superfamily hydrolase, subfamily IIB	1552	2385	+									
	2	Cell division protein FtsI [Peptidoglycan synthetase] (EC 2.4.1.129)	2688	4841	-									
	3	hypothetical protein	4874	4990	-									
	4	Transcriptional regulator, RpiR family	5147	6007	-									
	5	Threonyl-tRNA synthetase (EC 6.1.1.3)	6217	8160	-									
	6	hypothetical protein	8157	8270	-									
	7	ocrotonate tautomerase (EC 5.3.2.-); Xylose transport system permease protei	8496	8711	-									
	8	cadmium-translocating P-type ATPase	8722	10545	-									
		2	(1,161 bp) Enterococcus faecium (Firmicutes; Bacilli; Lactobacillales; Enterococcaceae; Enterococcus)											

References:

- Ausec, L., Zakrzewski, M., Goesmann, A., Schlüter, A., Mandic-Mulec, I., 2011. Bioinformatic analysis reveals high diversity of bacterial genes for laccase-like enzymes. *PLoS ONE* 6, e25724. doi:10.1371/journal.pone.0025724
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formisano, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. doi:10.1186/1471-2164-9-75
- Bhat, A., Riyaz-Ul-Hassan, S., Ahmad, N., Srivastava, N., Johri, S., 2013. Isolation of cold-active, acidic endocellulase from Ladakh soil by functional metagenomics. *Extremophiles*. doi:10.1007/s00792-012-0510-8
- Boonen, F., Vandamme, A.-M., Etoundi, E., Pigneur, L.-M., Housen, I., 2014. Identification and characterization of a novel multicopper oxidase from *Acidomyces acidophilus* with ferroxidase activity. *Biochimie* 102, 37–46. doi:10.1016/j.biochi.2014.02.009
- Bruhl, J.M., Antonopoulos, D.A., Berg Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., Edwards, R.E., Frank, E.D., Emerson, J.B., Wacklin, P., Coutinho, P.M., Henrissat, B., Nelson, K.E., White, B.A., 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences* 106, 1948–1953. doi:10.1073/pnas.0806191105
- Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré, J., Henrissat, B., Rемаud-Siméon, M., Monsan, P., Potocki-Véronèse, G., 2013. Functional Metagenomics Reveals Novel Pathways of Prebiotic Breakdown by Human Gut Bacteria. *PLoS ONE* 8, e72766. doi:10.1371/journal.pone.0072766
- Cheng, F., Sheng, J., Dong, R., Men, Y., Gan, L., Shen, L., 2012. Novel Xylanase from a Holstein Cattle Rumen Metagenomic Library and Its Application in Xylooligosaccharide and Ferulic Acid Production from Wheat Straw. *Journal of Agricultural and Food Chemistry* 60, 12516–12524. doi:10.1021/jf302337w
- Goel, G., Dagar, S.S., Raghav, M., Bansal, S., 2015. Rumen: an underutilised niche for industrially important enzymes, in: *Rumen Microbiology: From Evolution to Revolution*. Springer India, pp. 247–263.
- Hall, J.B., Silver, S., 2009. *Nutrition and Feeding of the Cow-Calf Herd: Digestive System of the Cow*.
- Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z., Rubin, E.M., 2011. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* 331, 463–467. doi:10.1126/science.1200387

- Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. *Genome Research* 17, 377–386. doi:10.1101/gr.5969107
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi:10.1093/bioinformatics/btu031
- Jørgensen, H., Kristensen, J.B., Felby, C., 2007. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels, Bioproducts and Biorefining* 1, 119–134. doi:10.1002/bbb.4
- Ko, K.-C., Han, Y., Choi, J.H., Kim, G.-J., Lee, S.-G., Song, J.J., 2011. A novel bifunctional endo-/exo-type cellulase from an anaerobic ruminal bacterium. *Applied Microbiology and Biotechnology* 89, 1453–1462. doi:10.1007/s00253-010-2949-9
- Kruger, N.J., 1994. The Bradford Method for Protein Quantitation, in: *Basic Protein and Peptide Protocols*. Humana Press, New Jersey, pp. 9–16.
- Kulkarni, S., Patil, S., Satpute, S., 2013. Microbial Esterases: An overview. *International Journal of Current Microbiology and Applied Sciences* 2, 135–146.
- Lazuka, A., Auer, L., Bozonnet, S., Morgavi, D.P., O'Donohue, M., Hernandez-Raquet, G., 2015. Efficient anaerobic transformation of raw wheat straw by a robust cow rumen-derived microbial consortium. *Bioresource Technology* 196, 241–249. doi:10.1016/j.biortech.2015.07.084
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–495. doi:10.1093/nar/gkt1178
- Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S., 2002. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiology and Molecular Biology Reviews* 66, 506–577. doi:10.1128/MMBR.66.3.506-577.2002
- Menon, V., Rao, M., 2012. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. *Progress in Energy and Combustion Science* 38, 522–550. doi:10.1016/j.pecs.2012.02.002
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386. doi:10.1186/1471-2105-9-386
- Morgavi, D.P., Kelly, W.J., Janssen, P.H., Attwood, G.T., 2013. Rumen microbial (meta)genomics and its application to ruminant production. *animal* 7, 184–201. doi:10.1017/S1751731112000419
- Nelson, N., 1944. A photometric adaptation of the Somogyi method for the determination of glucose. *The Journal of Biological Chemistry* 375–380.

- Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–214. doi:10.1093/nar/gkt1226
- Patel, D.D., Patel, A.K., Parmar, N.R., Shah, T.M., Patel, J.B., Pandya, P.R., Joshi, C.G., 2014. Microbial and Carbohydrate Active Enzyme profile of buffalo rumen metagenome and their alteration in response to variation in the diet. *Gene* 545, 88–94. doi:10.1016/j.gene.2014.05.003
- Regalado, V., Perestelo, F., Rodr#x00ED;guez, A., Carnicero, A., Sosa, F.J., De la Fuente, G., Falc#x00F3;n, M.A., 1999. Activated oxygen species and two extracellular enzymes: laccase and aryl-alcohol oxidase, novel for the lignin-degrading fungus *Fusarium proliferatum*. *Applied Microbiology and Biotechnology* 51, 388–390. doi:10.1007/s002530051407
- Ross, E.M., Moate, P.J., Bath, C.R., Davidson, S.E., Sawbridge, T.I., Guthridge, K.M., Cocks, B.G., Hayes, B.J., 2012. High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genetics* 13, 53. doi:10.1186/1471-2156-13-53
- Schwarz, W.H., 2001. The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiology and Biotechnology* 56, 634–649. doi:10.1007/s002530100710
- Song, L., Laguerre, S., Dumon, C., Bozonnet, S., O'Donohue, M.J., 2010. A high-throughput screening system for the evaluation of biomass-hydrolyzing glycoside hydrolases. *Bioresource Technology* 101, 8237–8243. doi:10.1016/j.biortech.2010.05.097
- Strachan, C.R., Singh, R., VanInsberghe, D., Ievdokymenko, K., Budwill, K., Mohn, W.W., Eltis, L.D., Hallam, S.J., 2014. Metagenomic scaffolds enable combinatorial lignin transformation. *Proceedings of the National Academy of Sciences* 111, 10143–10148. doi:10.1073/pnas.1401631111
- Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., Dore, J., Monsan, P., Remaud-Simeon, M., Potocki-Veronese, G., 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* 20, 1605–1612. doi:10.1101/gr.108332.110
- Terrapon, N., Henrissat, B., 2014. How do gut microbes break down dietary fiber? *Trends in Biochemical Sciences* 39, 156–158. doi:10.1016/j.tibs.2014.02.005
- Ufart#eacute;, L., Laville, E., Duquesne, S., Morgavi, D., Robe, P., Klopp, C., Rizzo, A., Bozonnet, S., Potocki-Veronese, G., 2016a. Discovery of carbamate degrading enzymes by functional metagenomics. to be submitted.
- Ufart#eacute;, L., Potocki-Veronese, G., Cecchini, D., Rizzo, A., Morgavi, D., Cathala, B., Moreau, C., Cl#eacute;ret, M., Robe, P., Klopp, C., Bozonnet, S., Laville, E., 2016b. Highly promiscuous oxidases discovered by functional exploration of the bovine rumen microbiome. to be submitted.

- Wang, L., Hatem, A., Catalyurek, U.V., Morrison, M., Yu, Z., 2013. Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS ONE* 8, e78507. doi:10.1371/journal.pone.0078507
- Warner, C.D., Go, R.M., García-Salinas, C., Ford, C., Reilly, P.J., 2011. Kinetic characterization of a glycoside hydrolase family 44 xyloglucanase/endoglucanase from *Ruminococcus flavefaciens* FD-1. *Enzyme and Microbial Technology* 48, 27–32. doi:10.1016/j.enzmictec.2010.08.009
- White, B.A., Lamed, R., Bayer, E.A., Flint, H.J., 2014. Biomass Utilization by Gut Microbiomes*. *Annual Review of Microbiology* 68, 279–296. doi:10.1146/annurev-micro-092412-155618
- Wilson, D.B., 2011. Microbial diversity of cellulose hydrolysis. *Current Opinion in Microbiology* 14, 259–263. doi:10.1016/j.mib.2011.04.004
- Wu, S., Baldwin, R.L., Li, W., Li, C., Connor, E.E., Li, R.W., 2012. The Bacterial Community Composition of the Bovine Rumen Detected Using Pyrosequencing of 16S rRNA Genes. *Metagenomics* 1, 1–11. doi:10.4303/mg/235571
- Ze, X., Ben David, Y., Laverde-Gomez, J.A., Dassa, B., Sheridan, P.O., Duncan, S.H., Louis, P., Henrissat, B., Juge, N., Koropatkin, N.M., Bayer, E.A., Flint, H.J., 2015. Unique Organization of Extracellular Amylases into Amyloosomes in the Resistant Starch-Utilizing Human Colonic *Firmicutes* Bacterium *Ruminococcus bromii*. *mBio* 6, e01058–15. doi:10.1128/mBio.01058-15

Conclusion and perspectives

In the introduction section, we have seen that the bovine rumen ecosystem is a highly diverse microbiota, producing a large range of enzymes, able to act on many different kinds of polymeric structures issued from dietary constituents. Up to now, this ecosystem, especially its bacterial fraction, has been mostly studied to understand better its ability to degrade plant cell wall polysaccharides, in particular cellulose and hemicelluloses. However, none of the numerous ruminal CAZymes that were discovered these last years through culture-independent methods like metagenomics, were tested on crude plant cell wall extracts representing industrial substrates for biorefineries. In addition, the structural originality of the newly discovered enzymes is relative, since none of them belong to a real new enzyme family and/or display new functions of carbohydrate modification. Furthermore, the potential role of ruminal bacterial oxidoreductases in lignin degradation has never been investigated, while this polymer is an important plant cell wall constituent, which negatively impacts nutritional efficiency. Finally, the substrate flexibility of the main classes of ruminal enzymes (glycoside-hydrolases, esterases, oxidoreductases, proteases, phytases...) has rarely been exploited for degradation of unnatural substrates, such as pollutants.

In this context, this thesis work aimed at several objectives, among which are:

- The discovery of novel pollutant degrading activities from metagenomic clones.
- Enrichment of the bovine ruminal microbiome in lignocellulosic functions in order to boost the discovery of original enzymes working synergistically to break down an industrial lignocellulosic substrate
- The exploitation of the functional diversity of this complex ecosystem to isolate novel enzymes of interest for depollution

Before concluding on the major results obtained from this thesis project, it is important to mention how enrichment of the ruminal microbiome was performed. Wheat straw was chosen as enrichment substrate because of its industrial interest, several European biorefineries using this agricultural waste, which is not in competition with food or feed applications, as raw material. It is a substrate particularly rich in cellulose, hemicelluloses and lignin that can be used to supplement cow's diet without causing too much digestive troubles. Because of its composition, it was chosen to enrich the native bovine microbiome in several activities of interest for this project, namely glycoside-hydrolases, esterases, and potentially also, bacterial oxidases. The metagenome of the consortium obtained by *in vivo* enrichment was thus the most studied in the frame of this thesis. The *in vivo* enrichment duration was only of seven weeks, and the library was immediately constructed and available at the beginning of my project. The optimization of the *in vitro* enrichment, which was conducted under the responsibility of Dr G. Hernandez-Raquet, lasted one year to obtain a stable lignocellulolytic consortium, which was subsequently used for the second library construction. As soon as the second library was available, the same primary screening strategy was applied to it, allowing comparison of the functional profiles of the *in vitro* and

in vivo enriched libraries. The taxonomical analysis of both libraries by 16S rDNA analysis is underway, together with that of the initial metagenomics sample before any enrichment.

To find enzymes of interest for bioremediation, it might have been easier to enrich the ecosystem with target pollutants, the native microbes and their enzymes not being naturally adapted to such synthetic compounds. Nevertheless, the *in vivo* enrichment approach would not have been possible because of the toxicity of the targeted compounds for mammals. In addition, conducting several *in vitro* enrichments on various pollutants would have taken too much time to allow result exploitation in only three years, and would have been too expensive. Moreover, toxicity for humans of most of the compounds targeted in this study (dyes and phytosanitary products) would have rendered much more difficult the steps of enrichment, DNA recovery, and library construction. We thus decided to abandon this option and to exploit a more 'natural' functional diversity. The multi-steps screening strategy we used allowed us to first isolate batteries of glycoside-hydrolases, esterases and oxidoreductases, of which the specificity towards structurally different substrates, including toxic ones, was screened in a second step, allowing us to select the most interesting clones for in-depth functional characterization.

In this context, we will conclude on the main results, discuss the advances, along with providing the perspectives of these works.

1. Discovery of new bacterial oxidoreductases

As described in the introduction section and in chapter V, bacterial ligninases were only recently discovered and very few of them have been isolated. Because lignin is a major plant constituent, it represents a significant part of a bovine's diet. However, at the beginning of this thesis, nothing was known on the potential ability of bacterial ruminal oxidoreductases to break down lignin.

1.1. Set up of novel screening assays

In order to investigate this question, we had to design novel protocols for screening this activity on solid media. Automated screening in solid format indeed allows a 500 to 1,000 times higher screening throughput (up to 400,000 assays per week) than screening in liquid format, as described in two of our papers published in 2015 (Ufarté et al. annex 1 and chapter 1).

First, we tested lignin alkali, otherwise known as kraft lignin (depolymerized lignin), as a substrate to look for enzymes involved in lignin metabolization. Counting on the fact that a battery of catabolic enzymes could be encoded by the same multi-genic cluster, then compatible with our fosmid clone libraries which harbor metagenomics inserts sizing several

dozen kb, the screening was performed in minimal medium, allowing growth of the sole clones that were able to use lignin alkali as carbon source. This positive selection strategy, which gave positive results, would probably not be successful with native, insoluble lignin, of which particles would not be sufficiently accessible to the heterologous enzymes mostly produced intracellularly by *E. coli*.

In order to increase our chances of finding ruminal oxidoreductases, we also adapted a classical protocol used for screening in micro-plate format libraries of laccases variants obtained by combinatorial protein engineering. In order to perform this screen in solid medium, we indeed had to first grow the *E. coli* clones on a membrane deposited on a rich culture medium, and to transfer them onto screening agar plates containing the ABTS chromogenic substrate and high amounts of copper, which are lethal for *E. coli*. This protocol also gave positive clones, at a yield comparable with that obtained from compost microbiomes in the frame of another team project. By optimizing this screening protocol and by increasing by 1,000 times its throughput, we rendered it highly generic and compatible with the functional exploration of very large sequence spaces, issued from metagenomes or directed evolution.

The results of the primary screening performed here on the bovine rumen indicated the presence of ruminal bacterial oxidoreductases, which may be involved in lignin degradation. Nevertheless, at this stage, we did not find evidence of their ability to break down polymeric lignin.

With this aim, we took advantage of the collaboration between the 'Discomics' group ('Metaomics for Enzyme Discovery') of the LISBP, in which this thesis work was done, and B. Cathala and C. Moreau (BIA, INRA Nantes, France), experts in polymer sciences and nanotechnologies. They indeed developed an innovative technology for the construction of nano-layers of polysaccharides, of which the color change depended on the layer thickness. In the frame of the ANR project 'REFLEX', the Discomics group set up an automated screening approach, based on the use of these polymer layers, for CAZymes discovery in metagenomes for instance. In the frame of my thesis, we thus asked our partners to create nano-layers of polymeric lignin, in order to test our metagenomic clones displaying oxidoreductase activities. Sulfonated lignin was used to create the films, in order to obtain stable polymer chain assemblies, which would not have been possible with insoluble native lignin particles. The positive results, presented in chapter V, establish the proof of concept of this screening methodology. Once optimized (as it was done recently for polysaccharide layers in the frame of the REFLEX project) for a better control of homogeneity and stability of the films, this technology will offer a very robust, sensitive, and quantitative method for high-throughput screening of lignin-degrading enzymes. With the other screening methods developed in this thesis project, it would in particular be highly interesting to explore the lignolytic potential of uncultivated ruminal fungi, thanks to activity-based

metatranscriptomic approaches, such as those recently developed by the Discomic group (paper in preparation).

1.2. Properties of the new ruminal oxidoreductases

In order to characterize substrate specificity of the newly discovered enzymes, optimization of the reaction conditions was performed by using ABTS as substrate. Experiments with or without copper, and with or without hydrogen peroxide, showed that none of the eight hits absolutely needed copper to be active, indicating that they do not belong to laccase families. While none exhibited higher activity with hydrogen peroxide from the *in vivo* enrichment, one enzyme from the *in vitro* enrichment was active in the presence of H₂O₂, indicating the possibility of it being a peroxidase. Further characterization of the temperature and pH profiles of activity was necessary, to find the optimal conditions of use on the substrates of interest (dyes and lignin).

This showed that these enzymes are thermoactive, being still active at 50°C, and also active in acidic conditions (pH 4.5-5.0). In their respective optimal conditions of use, the three clones selected for in-depth characterization were able to discolor structurally different dyes, especially AZO dyes. Further experiments are nevertheless necessary to conclude on the exact activity of each hit enzyme. Structural characterization of the reaction products (the soluble and insoluble ones) by NMR and MS should allow us to determine the substrate modification sites. This step will need purification of the targeted enzymes, in order to eliminate the *E. coli* oxidoreductase background activity that we detected in the metagenomic clones. In order to identify the genes responsible for the detected activities, there are three different approaches.

The first one, which is a random one, is transposition mutagenesis. It consists of disrupting genes by randomly inserting a transposon sequence in the fosmid insert, and in screening again the transformants to identify those which lost the activity. The Discomics group used this approach many times in other studies. However, the results were often disappointing, since activity is sometimes abolished because of the insertion of the transposon in sequence regions that act as transcription regulators for *E. coli*. A recent transcriptomic study of the team, performed on *E. coli* metagenomic clones indeed allowed us to demonstrate that heterologous gene expression is due to the random presence of those sequences in the metagenomic inserts, and not to the native promoters (paper in preparation).

The second approach consists of rational truncation of the fosmid inserts around the genes that have been identified by sequence analysis to be the most probable targets. This strategy is especially powerful, even though time consuming, if several genes encode

enzymes that are involved in the screened activity. Nevertheless, it does not allow one to produce high yields of heterologous proteins, and to tag them for further purification.

The last approach consists of sub-cloning only the best gene targets in overexpression vectors. In our case, functional annotation allowed identification, for each hit clone, of genes encoding oxidoreductases, along with a genomic context that seems to imply a synergic activity of different enzymes. Two of the targets did not correspond to any known genes involved in polyaromatic compound degradation, highlighting the originality of these new sequences, but unfortunately, the impossibility of determining the exact enzyme family. The presence in the third clone (clone 32D12) of genes encoding putative red-ox enzyme with four subunits will make the over-production of this target a crucial step, especially during the protein folding phase in the cytoplasm. Different vectors and strains would be necessary to test in order to find the best production conditions. For the monomeric and homo-multimeric enzymes, usual overexpression vectors, containing the tags of interest for purification would be the first to test, while for the four sub-units enzyme, a special cloning system, used to produce the whole enzyme in one step would be necessary. A strain designed to help folding the protein it produces might be necessary, like the ArcticExpress competent cells (Agilent) that produce cold-adapted chaperonin to help in the folding of expressed enzymes for example. Of course, a large number of experiments and multiple combinations would be necessary to find the best cloning and expression conditions. Once purified, these enzymes will be further characterized to investigate their role in the structural modification of the polyaromatic substrates, to investigate their structure/function relationships, and to quantify their potential for dye bioremediation and the valorization of lignin derivatives from different industries (in terms of efficiency, stability, yield of recombinant expression, and toxicity/recovery facility/biodegradability and recyclability of the released products).

Dye discoloration is indeed of interest, especially for the textile industry, considering the amount of dyes leaking into the environment. The formation of colored precipitates could be used to remove this part of waste by centrifugation, filtration or sedimentation for example. Enzyme immobilization could be envisaged to eventually stabilize the biocatalysts, to recycle them and decrease their production costs. Dye discoloration would in this case be performed in batch enzyme reactors only, precipitation of the reaction products preventing the use of fixed-bed reactors. Microbial reactors based on enzyme production in heterologous strains performing efficient enzyme secretion, if possible GRAS (Generally Recognized as Safe) ones like *Bacillus subtilis*, could also be envisaged to circumvent the economic bottleneck due to enzyme production and purification.

The same could be said for lignin alkali, a waste produced in paper industries. The creation of new processes to degrade it that could be used in small or medium sized paper plant would be of high interest, considering that they cannot afford the usual expensive

treatment of wastewater. The positive impact on the environment could be quite high, considering that Kraft process is predominant in the pulping industry. The fact that these enzymes are also able to degrade sulfonated lignin, a byproduct of the sulfite pulping method is also of high interest. Indeed, while the process has been mostly replaced by Kraft process, it is still used nowadays. Further analyses are needed to determine the yield of substrate depolymerisation and the structure of the degradation products. The ability of our new red-ox enzymes to degrade lignin extracts from various botanical origins, as well as lignin wastes that are recalcitrant to plant biomass pretreatment processes, will also have to be tested. These last aspects require complementary expertise in lignin structural characterization. Collaboration has been initiated with L. Eltis (University of British Columbia), to this aim.

Finally, if we discovered several new red-ox bacterial enzymes that are able to break down various polyaromatic substrates, including lignin derivatives, their physiologic role in the rumen is still unclear. We indeed have no experimental proof of action of these enzymes on native lignin, even if at least two (produced by clones 10N15 and 32D12) are encoded by multigenic systems, encoding also hemicellulase involved in the breakdown of the plant cell wall polysaccharidic network. The characterization of native lignin structure before and after incubation with those enzymatic extracts should soon allow us to confirm or not this hypothesis.

2. Ruminal esterases as bioremediation tools

A large diversity of esterases, active on many different substrates, was also discovered thanks to the multi-step screening strategy we used in this study.

More in depth characterization was performed after primary and secondary screening on chromogenic esters, in order to study their potential for carbamate pollutant degradation. Two classes of substrates were selected: Impranil DLN, a commercial polyurethane, two insecticides carbamates (fenoxycarb and fenobucarb) and one herbicide thiocarbamate (prosofocarb). Screening in solid minimal and rich media on Impranil DLN allowed the isolation of 57 clones from the 85 primary hits. Two clones were chosen for their efficiency to form a degradation halo on rich medium, and to grow on minimal medium supplemented with Impranil DLN as sole carbon source. Clone 44112 is also active on the three tested *p*NP-substrates, while 12F23 is only weakly active on *p*NPB only. Efficiency of polyurethane degradation and these differences in substrate specificities guided our choice, in order to avoid sequence redundancy, which often occurs after activity-based screening of metagenomics libraries.

HPSEC and MALDI-TOF results confirmed that Impranil DLN structural modifications occurred during hydrolysis. However, since the exact structure of the lateral chains of this

commercial polymer is not available (even though a potential structure has been recently proposed) (Biffinger et al., 2015), it is impossible at this stage to identify the exact cleavage sites of this substrate. Reaction buffer and conditions of HPSEC analysis will have to be optimized, in order to avoid aggregation (due to lyophilisation in PBS buffer), and to increase product recovery yields. Then, NMR and MS analysis of the reaction products should allow us to decipher the mechanism of action of these new esterases on Impranil DLN.

Reactions on smaller carbamates showed that only clone 44I12 is able to convert the tested phytosanitary compounds, in particular fenobucarb, into 2-*sec*-butylphenol. This highlights the large substrate flexibility of CE2_Ubrb, which can degrade the three tested *p*NP-substrates, Tween20, and two structurally different carbamate pollutants, compared to CE1_Ubrb that can only degrade one *p*NP-substrate and Impranil DLN. Sequence analysis indicated that CE1 and CE2_Ubrb belong to lipolytic family V and IV respectively. These two enzymes are today the only members of these families presenting carbamate degrading activities. Indeed, until the present study, only family VII was known to possess carbamate degrading members. Let's notice that CE1_Ubrb best characterized homolog is a PLA degrading enzyme. Preliminary assays did not allow us to detect any PLA degrading activity for the CE1 and CE2_Ubrb producing clone. Nevertheless, degradation of PLA and structurally homologous plastics will have to be re-tested soon by using recently optimized methods, plastic degradation being a very important research field for plastic recycling.

In any case, like the red-ox enzymes described before, the next step will be to sub-clone the CE1 and CE2_Ubrb encoding genes in an over-expression vector in order to purify and crystallize them to identify the structural determinants of substrate flexibility. This will also allow us to determine their catalytic efficiency and stability, which are two important parameters for enzyme industrialization. Indeed, polyurethanes are a non-negligible part (6-7%) of the total mass of plastics produced worldwide, and cover more than half of the automotive and commercial coating application market. Among these, Impranil DLN is used for textile, leather, and aircraft fabric coatings. As such, the degradation of such a dispersed pollutant is of high interest. The fact that the clones are able to grow on Impranil DLN selective medium shows their ability to completely metabolize the substrate, which is very interesting for microbial bioremediation processes. The same can be said for fenobucarb. Indeed, fenobucarb is an insecticide carbamate, used predominantly in paddy fields to control rice thrips, leafhoppers and plant hoppers in many rice growing countries worldwide. Unfortunately, it has been demonstrated that a part of the insecticides used move from the rice paddy effluent to receiving water bodies, among which can be found surface water and more dangerously ground water. The degradation of fenobucarb has then become a serious environmental need. Bioremediation of fenobucarb thanks to CE2_Ubrb-based enzymatic or microbial process is thus of particular interest, especially because this enzyme is the first member of the lipolytic family IV to catalyze this reaction, which would allow one to

circumvent intellectual property covering carbamate degrading enzymes belonging to family VII.

3. Effect of enrichment on the functional profile of the rumen microbiome

The two metagenomic libraries obtained after *in vivo* and *in vitro* enrichment on wheat straw were screened for polysaccharide degrading enzymes, esterases and oxidoreductases. The results highlighted drastically different functional profiles expressed in these libraries. A much lower yield of hit clones active on Tween 20 and AZCL-xylan was observed, along with a higher hit yield on AZCL-galactomannan, AZCL- β -glucan and AZO-CMCellulose, by comparing the results obtained from the IVTE library with those from the IVVE one. This trend could be explained by the duration and the efficiency of substrate degradation. While *in vitro*, the reaction lasts seven days before a new inoculation is made, *in vivo*, the transit time varies from 20 to 48h. As such, the hypothesis is that in the rumen, the readily accessible biomass is degraded first, and then, hemicelluloses are being degraded, with a little of amorphous cellulose, which might explain the higher number of hits on hemicellulolytic substrates, and the low number of hits on cellulases. On the other hand, for the *in vitro* enrichment, the degradation lasts longer, and the amorphous cellulose might be the one that is being degraded at the sampling time, along hemicelluloses, which would explain the higher number of cellulases.

Degradation assays on different kinds of chromogenic celluloses show that no matter the enrichment, the more crystalline cellulose was not degraded, while the more amorphous ones gave a higher number of hits from the IVTE, validating the hypothesis on cellulose degradation. This result is not surprising, since to our knowledge, no enzyme able to degrade crystalline cellulose has been isolated from mammal or insect gut microbiomes by using activity-based metagenomics. This is probably due to the fact that although crystalline cellulose is chemically homogeneous, no single enzyme is able to hydrolyze it, whereas soluble cellulose derivatives can easily be degraded by a single endo- β -1,4-glucanase. The crystalline cellulose is only hydrolyzed by a cluster of interacting enzymes or by multi-enzyme complex like cellulosomes. In the present study, the strategy of metagenomic library construction allowed the cloning of entire multigenic clusters, but concomitant expression in *E. coli* of all genes from a same gene cluster is highly improbable. Even if some of our CAZy sequences seem to harbor dockerin modules, that are the witnesses of the presence of cellulosome-like structures (like in clone 49M8), it is not sure at all that the enzymatic complex targets cellulose in the native bacterium, since cellulosome-like assemblies are not specific to cellulose degradation. In addition, chromogenic substrates and chemically treated ones (like Avicel) are not structurally identical to native semi-crystalline cellulose, which might induce a bias in substrate specificity characterization. Testing our clones on native

crystalline cellulose, like cotton fibers for instance, would allow us to definitively assess the potential of these enzymes for degradation of crystalline materials.

Another point of interest is that from the 172 hits from the primary screening, 68 were active on more than one polysaccharidic substrate. This trait has already been observed several times for fosmid clones expressing CAZyme activities, whatever the gut ecosystem screened was. In bacteria, synergistic systems of transport and degradation of complex carbohydrate structures, like plant cell wall polysaccharides and eukaryotic glycans, are indeed encoded by multigenic systems, like the Polysaccharide Utilization Loci (PULs) described for bacteria from the Bacteroidetes phylum. In the present study, many CAZyme encoding multigenic clusters have been identified, most of them being issued from uncultivated Firmicutes. It would thus be false to name them real PULs, since PUL description is specific for Bacteroidetes, and since we won't be able to prove that in the native strain, all genes of a genomic locus involved in polysaccharide harvesting are co-transcribed. However, our data, together with the ones previously published by the Discomics group and others, indicates that (meta)genomic loci involved in the degradation of complex polysaccharide structures are not at all specific for Bacteroidetes. Only transcriptomic studies performed on native culturable strain harboring similar loci would allow one to definitively prove it. In addition, some of the multigenic clusters involved in plant cell wall degradation discovered in the present study also contain red-ox enzymes that are probably able to act on lignin. This is very unusual, since bacterial CAzymes and ligninases, like laccases, are not known to be encoded on the same loci. Synthetic analyses of the metagenomic loci discovered here, together with transcriptomic studies performed both on the metagenomic clones and on cultivated strains harboring homolog genomic structures would allow us to support this hypothesis, and to revisit the concept of co-expression of bacterial ligninase and CAZyme encoding genes.

As it is the case for other gut microbiomes, more than one hundred hit clones producing CAZyme activities were identified, from only few Gb of metagenome cDNA. Even if secondary screening allowed us to discriminate the clones depending on their functional profile, and thus to avoid too much sequence redundancy, it was still difficult to identify those that present the highest potential for degradation of plant biomass. For this purpose, we quantified the efficiency of lignocellulose degradation for each of the hit clones, by measuring the amount of reducing sugars released from the crude plant biomass substrate used for enrichment, wheat straw. We had to optimize the method previously described by Song et al. to screen in micro-plate format activity of hemicellulase mutants over-expressed in plasmid under the control of a strong promoter. In contrast, in fosmids and Epi100 *E. coli* host, recombinant gene expression is much lower, requiring highly sensitive screening methods. Here, by modifying the method of reducing sugar quantification, we increased the sensitivity of the assay by 10 times. This allowed us to detect very low activities of degradation of real plant cell wall matrixes. It is the first time that metagenomic clones have

been tested on such crude lignocellulosic biomass used in biorefineries. The straw degradation efficiency of the best clones obtained from both libraries was similar. In addition, their efficiency was not correlated to polyspecificity of the enzymatic extracts, the most active clones being active on only one or two polysaccharide structures. In these metagenomic clones, efficiency of plant cell wall degradation is thus rather due to a highest expression level and/or activity of one particular cellulase or hemicellulase, rather than to synergistic CAzyme activities. Of course, this experiment is not representative of the functioning of lignocellulolytic microbial consortia, which simultaneously produce a large battery of lignocellulases to tackle the complexity of plant cell wall structure.

We thus tried to reproduce the effects of such synergistic enzyme actions, in order to exploit the diversity of activities we discovered for designing lignocellulolytic cocktails, and to study the specific role of esterases and red-ox enzymes in straw degradation efficiency. Several hit clones were chosen to create lignocellulase mixtures, according to two criteria: the production of several complementary activities required for plant cell wall degradation, due to multi-genic clusters encoding a battery of CAZymes, and potentially also of oxidases; and the clone efficiency of wheat straw degradation, which had to correspond at least to 1.5 times more released reducing sugars than with the negative control. Different mixes were created, adding separately the oxidoreductases and, if possible, the esterase clones.

This study allowed the conception of enzymatic cocktails with real synergic effects, since at least one mix from each library was able to free around twice the amount of reduced sugars as a theoretical amount calculated from the sum of individual clone activities. While esterases positively impacted the efficiency of straw degradation, the red-ox enzymes surprisingly tended to inhibit the CAZymes. This might be due to modification of the lignin network, and eventually to its depolymerization/repolymerization, which would affect the binding and accessibility of cellulases and hemicellulases. Nevertheless, this hypothesis is still highly speculative, the mode of action of the newly identified red-ox enzymes being not identified at this stage. The performances of our cocktails are of course limited compared to those of commercial lignocellulolytic cocktails and of microbial consortia, because of the very low expression level of the heterologous enzymes. Most of the proteins produced by these hit clones are indeed those of *E. coli* itself. In addition, the entire combinations of the hit clones isolated in this study was far from having been exploited, since we selected only a few clones for the design of the mixes. Nevertheless, here we established the proof of concept of the strategy, which could be applied to the hundreds of lignocellulase hit clones issued from this thesis work, but also from previous studies targeted on other microbiomes. Once the best cocktails of hit clones are identified, the lignocellulase encoding genes they harbor will have to be sub-cloned in over-expression vectors in order to produce and purify them for in-depth characterization of their catalytic efficiency, individually and with their plant cell wall degrading partners.

In addition, at this stage of the study, we do not know if the lignocellulases identified after activity-based screening are the same as those that are produced in the best lignocellulolytic consortia, designed by A. Lazuka and G. Hernandez-Raquet (Lazuka et al., 2015). Metaproteomic analysis of the artificial consortia should allow us to compare the hit sequences, and to precisely evaluate the bias of the activity-based metagenomic strategy. In addition, analysis of the CAZome of the ruminal bacteriome before and after the two enrichments should be possible soon, thanks to the development of a functional chip harboring all the known CAZy sequences available in genomic databases. Even biased by the content of the public databases, which is far from representing the CAZy diversity of all cultivated and uncultivated bacteria, this approach should allow one to contribute to the understanding of the effect of enrichments on the functional profile of the ruminal ecosystem. These data will be integrated to the coming 16S rDNA analysis ones, to also investigate the effect of enrichments on the consortia taxonomical diversity.

More generally, the enzymes identified here seem to be the key actors of lignocelluloses degradation in the rumen, since several of our CAZy encoding genes, even entire gene clusters, were already found in other metagenomic studies targeting other cow cohort worldwide. The analysis of prevalence and of abundance of our hit sequence by mapping the data on catalogs of genes obtained by massive random sequencing of the bovine rumen metagenome should allow us to highlight those of our enzymes that are the most important for the ecosystem soon. These results will guide the selection of the most pertinent targets for further in depth functional and structural investigations, together with sequence originality.

The number of unknown CAZy modules found from both libraries is indeed particularly high, opening very interesting perspectives for characterization of new enzyme families and possibly new functions. We also identified some clones harboring activities that are not related to any of the known CAZy families annotated in these sequences. We will thus have to rapidly identify these genes which are responsible for these functions, in order to enlarge the panel of functions of carbohydrate modification of already characterized or not enzyme families. The analysis of the genomic context of these targets is underway, to guide the in-depth characterization of their substrate specificity.

To conclude, our work has highlighted the interest of the cow rumen microbiota as a source of very original bacterial enzymes that are key actors of dietary constituent catabolism, and of which the promiscuity can be exploited for remediation. Several of these enzymes present a particular interest for improving the understanding of the relationships between enzyme structure and substrate specificity, along with being targets of high potential for biorefineries and bioremediation. Moreover, while bringing hope for a better understanding of the cow rumen ecosystem functioning from the molecular to the ecosystem scales, the approach developed here could be applied to other substrates and

ecosystems, particularly in order to decipher the mechanisms of xenobiotic degradation in mammal guts, soil or wastewater.

Artwork and Table contents

1. Artwork content

Figure 1: Strategies for the functional exploration of metagenomes, metatranscriptomes and metaproteomics to discover new functions and protein families.....	34
Figure 2: Microfluidic strategies for new enzyme screening.	41
Fig. 3. Schematic representation of sequence-based and activity-based functional metagenomics approaches.....	61
Figure 4: Schematic representation of the bovine digestive tract.....	85
Figure 5: Composition and organisation of primary cell wall	87
Figure 7: Relative distribution of the most abundant phyla in three chambers of cow stomach	91
Figure 6: Phylum level composition depending on the age of the animals.....	91
Figure 8: Relative distribution of phyla of the cow abomasum.....	92
Figure 9: Bacteroidetes polysaccharide utilisation system: example of the xyloglucan degradation pathway of <i>Bacteroidetes ovatus</i> ATCC 8483	101
Figure 10: Schematic representation of the <i>C. thermocellum</i> cellulosome.....	102
Figure 11: The three types of reactions catalyzed by cellulases on cellulose.....	107
Figure 12: Schematic representation of plant xylan structure, with its different substituent groups, showing sites of attack by microbial xylanolytic enzymes.	108
Figure 13: Pectinolytic activities.....	109
Figure 14: Mechanism of phytate hydrolysis by phytases.....	111
Figure 15: Hydrolytic reaction by tannase	112
Figure 16: Predominant pathways of biohydrogenation of dietary linoleic and linolenic acids in the rumen.....	113
Figure 17: Functional profiles of the hit metagenomic clones, assayed on polysaccharides (dark blue), Tween 20 (green), Impranil (brown: minimal medium; orange: rich medium), pNPA (purple), pNPB (red), pNPP (bright blue).	149
Figure 18: Reaction media of 12F23 and 44I12 cytoplasmic extracts incubated with Impranil at 30°C during 48 h.....	151
Figure 19: HPSEC chromatogram of the reaction between Impranil and the clones of interest.	152
Figure 20: MALDI-TOF spectra of reaction media (A) before and (B) after 24h reaction with clone 44I12.....	153
Figure 21: HPLC analysis from the reaction of fenobucarb with 44I12 enzymatic extract at the beginning of the reaction (blue) and after 90 hours reaction (red)	154
Figure 22: Sequence alignment between CE1_Ubrb and enzymes from the lipolytic family V	161
Figure 23: Sequence alignment between CE2_Ubrb and enzymes from the lipolytic family IV	163

Figure 24: Determination of optimal pH of ABTS oxidation by the metagenomic clones.at 30°C.....	192
Figure 25: Determination of optimal temperature of ABTS oxidation by the metagenomic clones, at optimal pH.....	193
Figure 26 Determination of the optimal temperature of ABTS oxidation by the metagenomic clones at optimal pH, after 17 hours of incubation at each temperature.	193
Figure 27: Effect of the mediator type on oxidative activity of the metagenomic clones after 17 hours of incubation, at optimal pH and temperature..	195
Figure 28: Degradation of semi-reflective layers of xyloglucan and sulfonated lignin by the hit metagenomics clones.	196
Figure 29: Dye degradation by the hit clones, in solid and liquid media: quantification of the yields of liquid medium discoloration after reaction	198
Figure 30: Dye discoloration by the hit metagenomic clones.	199
Figure 31: Activity of the IVVE hit clones on wheat straw..	227
Figure 32: Activity of the IVTE hit clones on wheat straw.	228
Figure 33: Activity of the different lignocellulolytic mixes on wheat straw.	234

2. Table content

Table 1: Examples of databases specialized in enzymatic functions of biotechnological interest.....	35
Table 2: Discovery of pollutant degrading enzymes using sequence based metagenomics approaches.....	68
Table 3: Discovery of pollutant degrading enzymes by activity-based metagenomics.....	69
Table 4: Comparison of the abundancy of major GH families in the gut metagenomes of herbivores, ruminant or not	104
Table 5: List of studies mentioning the discovery of enzymes from the cow rumen through sequence-based metagenomics, including an experimental proof of function.....	119
Table 6: List of studies mentioning the discovery of enzymes from the cow rumen through activity-based metagenomics.	119
Table 7: Structures of Tween 20 and of the carbamates used in this study.....	150
Table 8: ORFs of interest from the sequenced clones active on Tween 20, Impranil and/or pNPA/B/P, annotated as putative esterases, ureases or proteases.....	156
Table 9: Structures of the tested dyes (type I, indigo dye; AZO, azo dye; A, anthraquinonic dye; T, triarylmethane dye), and their maximal wavelength of absorbance.	186
Table 10: Optimal conditions for ABTS oxidation by the hit metagenomic clones.	191
Table 11: Number of hits obtained from the primary screening of both metagenomic libraries.	223
Table 12: Number of hits obtained from secondary screening.	226

Table 13: List of the clones included in the lignocellulase cocktails, their functional properties and CAZome, deduced from sequence analysis. 231

Résumé en français

L'écosystème du rumen bovin est composé d'un microbiote très diversifié, produisant une large gamme d'enzymes, capables d'agir sur plusieurs types de structures polymériques issues de constituants alimentaires. Jusqu'à maintenant, cet écosystème, et plus particulièrement sa fraction bactérienne, a été étudié majoritairement pour mieux comprendre sa capacité à dégrader les polysaccharides de la paroi cellulaire des végétaux, en particulier la cellulose et les hémicelluloses. Cependant, aucune des nombreuses CAZymes issues du rumen qui furent découvertes ces dernières années par des méthodes ne nécessitant pas la culture directe du microbiote, comme la métagénomique, n'ont été testées sur des extraits bruts de la paroi végétale représentatifs des substrats industriels utilisés en bioraffinerie. De plus, l'originalité structurale des enzymes nouvellement découvertes est relative, puisque aucune d'entre elles n'appartient à de réelles nouvelles familles d'enzymes et/ou ne montrent de nouvelles fonctions de modification des sucres. En outre, le rôle potentiel des oxydoréductases bactériennes du rumen dans la dégradation de la lignine n'a jamais été investiguée, bien que ce polymère soit un important constituant de la paroi végétale, qui impacte négativement sur l'efficacité nutritionnelle. Enfin, la flexibilité de substrats des classes majeures des enzymes du rumen (glycoside-hydrolases, estérases, oxydoréductases, protéases, phytases...) n'a été exploitée que rarement pour la dégradation de substrats non-naturels, comme les polluants.

Dans ce contexte, ce travail de thèse a eu pour objectif de :

- La découverte de nouvelles activités de dégradation de polluants issues de clones métagénomiques.
- L'enrichissement du microbiome du rumen bovin en fonctions lignocellulosiques pour augmenter la découverte d'enzymes originales travaillant en synergie pour dégrader un substrat lignocellulosique industriel.
- L'exploitation de la diversité fonctionnelle de cet écosystème complexe pour isoler de nouvelles enzymes d'intérêt pour la dépollution.

Il est important aussi de mentionner comment l'enrichissement a été réalisé. La paille de blé a été choisie comme substrat d'enrichissement grâce à son intérêt industriel, plusieurs bioraffineries européennes utilisant ce déchet issu de l'agriculture, qui n'entre pas en compétition avec les applications alimentaires, en tant que matériel brut. C'est un substrat particulièrement riche en cellulose, hémicelluloses et lignine qui peuvent compléter le régime des vaches sans causer trop de problèmes digestifs. Du fait de cette composition, nous l'avons choisie pour enrichir le microbiome natif en activités d'intérêts pour ce projet, notamment les glycoside-hydrolases, estérases, et aussi potentiellement, les oxydases bactériennes. Le métagénome du consortium obtenu par enrichissement *in vivo* a ainsi été le plus étudié dans le cadre de cette thèse. La durée de l'enrichissement *in vivo* a seulement été de deux semaines, et la banque a été immédiatement construite et disponible au début de mon projet. L'optimisation de l'enrichissement *in vitro*, qui a été réalisé sous la responsabilité du Dr. G. Hernandez-Raquet, a duré un an pour obtenir un consortium

lignocellulolytique stable, qui a par la suite été utilisé pour la construction de la seconde banque. Dès que celle-ci a été disponible, la même stratégie de criblage primaire a été utilisée, permettant la comparaison des profils fonctionnels des banques enrichies *in vivo* et *in vitro*. L'analyse taxonomique des deux banques par analyse des ADNr 16S est en cours, ainsi que celle de l'échantillon initial avant enrichissement.

Pour trouver des enzymes d'intérêt pour la bioremédiation, il aurait pu être plus simple d'enrichir l'écosystème avec les polluants ciblés, les microbes natifs et leurs enzymes n'étant pas naturellement adaptés à de tels composés synthétiques. Cependant, la méthode d'enrichissement *in vivo* n'aurait pas été possible à cause de la toxicité des composés ciblés pour les mammifères. De plus, réaliser plusieurs enrichissements *in vitro* sur différents polluants aurait pris trop de temps pour permettre l'exploitation des résultats en seulement trois ans, et aurait été trop onéreux. En outre, la toxicité pour les humains de la plupart des composés ciblés par cette étude (colorants, produits phytosanitaires) aurait rendu les étapes d'enrichissement, de récupération d'ADN et de construction de banque plus difficiles. Nous avons donc décidé d'abandonner cette option et d'exploiter une diversité fonctionnelle plus « naturelle ». La stratégie de criblage multi-étapes que nous avons utilisé a permis d'isoler des batteries de glycoside-hydrolases, estérases et oxydoréductases, desquelles la spécificité envers des substrats structuralement différents, y compris des substrats toxiques, a été criblée dans une seconde étape, nous permettant de sélectionner les clones les plus intéressants pour une caractérisation fonctionnelle approfondie.

Régime initial des vaches: Ration standard des vaches laitières en ensilage de maïs, foin et concentrés		
Echantillonnage initial du contenu du rumen Mix d'échantillons de rumen de deux animaux		
Enrichissement <i>in vivo</i> (IVVE): Les deux mêmes vaches sont nourries majoritairement avec de la paille de blé pendant sept semaines	Enrichissement <i>in vitro</i> (IVTE): Echantillon de rumen (mix issu de deux animaux) avec de la paille de blé comme seule source de carbone dans un réacteur (dix cycles de sept jours)	
Extraction d'ADN pour la création de banques métagénomiques		
Banque IVVE	Banque IVTE	
<i>Criblage primaire</i>		
Estérases/lipases: Tween 20	Oxydoréductases: Lignine alkali	Carbohydrate Active Enzymes (CAZymes): Polysaccharides chromogéniques
<i>Criblage secondaire</i>		
Estérases/lipases: <i>p</i> NP-acétate, <i>p</i> NP-butyrates, <i>p</i> NP-palmitate	Oxydoréductases: ABTS	CAZymes: 6 polysaccharides chromogéniques de structure proche (celluloses modifiées chimiquement, hémicelluloses ramifiées ou linéaires)
<i>Caractérisation de la spécificité de substrat et quantification de l'activité</i>		
<i>Banque IVVE</i>		<i>Banques IVVE et IVTE</i>
Estérases/lipases: Polymères et monomères de la famille des carbamates	Oxydoréductases: Colorants synthétiques, dérivés de lignine	Paille de blé, Conception de cocktails enzymatiques
Séquençage, annotations fonctionnelle et taxonomique	Séquençage, annotations fonctionnelle et taxonomique	Séquençage, annotations fonctionnelle et taxonomique
		Comparaison de banques
Nouvelles enzymes pour les bioraffineries et la bioremédiation		

Chapitre IV

Chapitre V

Chapitre VI & Annexe I (Ufarté et al., Springer methods, 2016)

Les résultats principaux sont présentés ci-après, ainsi que la discussion des avancées réalisées, et les perspectives envisagées.

1. Découverte de nouvelles oxydoréductases bactériennes

Les ligninases bactériennes n'ont été mises en évidence que récemment, et seul un faible nombre ont été isolées. Comme la lignine est un constituant majeur des plantes, elle représente une part non négligeable du régime bovin. Cependant, au début de cette thèse, rien n'était connu sur les capacités potentielles des oxydoréductases bactériennes du rumen pour dégrader la lignine.

1.1. Mise en place d'un nouveau test de criblage

Afin d'investiguer cette question, nous avons dû concevoir de nouveaux protocoles de criblages de cette activité sur milieu solide. En effet, le criblage automatique en format solide permet un débit de criblage 500 à 1000 fois plus élevé (jusqu'à 400000 tests par semaine) que le criblage en format liquide, comme décrit dans deux de nos papiers publiés en 2015 (Ufarté et al., 2015 annexe 1 et chapitre I).

En premier lieu, nous avons testé la lignine alkali, aussi appelée kraft lignine (lignine dépolymérisée), comme substrat pour trouver des enzymes impliquées dans la métabolisation de la lignine. En comptant sur le fait qu'une batterie d'enzymes cataboliques peut être encodée par le même groupe multigénique, et compatible avec les clones de notre banque fosmidique, contenant des inserts métagénomiques de plusieurs dizaines de kb, le criblage a été réalisé en milieu minimum, permettant la seule croissance de clones capables d'utiliser la lignine alkali comme source de carbone. Cette stratégie de sélection positive, qui a donné des résultats positifs, n'aurait certainement pas réussi avec la lignine insoluble native, dont les particules ne seraient pas suffisamment accessibles aux enzymes hétérologues majoritairement produites sous forme intracellulaires chez *E. coli*.

Afin d'augmenter nos chances de trouver des oxydoréductases du rumen, nous avons aussi adapté un protocole classique utilisé pour le criblage en format microplaques de banques de mutants de laccases obtenus par ingénierie combinatoire des protéines. Afin de réaliser ce criblage en milieu solide, nous avons en effet d'abord dû faire croître les clones *E. coli* sur des membranes déposées sur milieu riche, et de les transférer sur un criblage en boîtes agar contenant de l'ABTS, substrat chromogénique, et de grandes quantités de cuivre, qui sont létales pour *E. coli*. Ce protocole a aussi donné des clones positifs, à un rendement comparable avec celui obtenu à partir des microbiomes de compost dans le cadre d'un autre projet de l'équipe. En optimisant ce protocole, et en augmentant par 1000 son débit, nous l'avons rendu hautement générique et compatible avec l'exploration fonctionnelle de très larges espaces de séquences, issues des métagénomomes ou de l'évolution dirigée.

Les résultats du criblage primaire réalisés sur le rumen bovin indiquent la présence d'oxydoréductases bactériennes du rumen, qui pourraient être impliquées dans la dégradation de la lignine. Toutefois, à cette étape, nous n'avons pas mis en évidence la dégradation de la lignine polymérique.

Dans ce but, nous avons tiré profit de la collaboration entre le groupe « Discomics » (« Métaomiques pour la Découverte d'Enzymes ») du LISBP, dans lequel ce travail de thèse a été effectué, et B. Cathala et C. Moreau (BIA, INRA Nantes), experts en science des polymères et nanotechnologies. En effet, ils ont développé une technologie innovante de construction de nano-couches de polysaccharides, dont la couleur dépend de l'épaisseur de la couche. Dans le cadre d'un projet ANR « REFLEX », le groupe Discomics a mis en place une approche de criblage automatisée, basée sur l'utilisation de ces couches de polymères, pour la découverte de CAZymes dans les métagénomomes par exemple. Dans le cadre de ma thèse, nous avons donc demandé à nos collaborateurs de créer des nano-couches de lignine polymérique, afin de tester nos clones métagénomiques produisant une activité oxydoréductase. La lignine sulfonée a été utilisée pour créer ces films, afin d'obtenir des assemblages stables de chaînes de polymères, ce qui n'aurait pas été possible avec des particules de lignine native, insoluble. Les résultats positifs, présentés dans le chapitre V, établissent la preuve de concept de cette méthode de criblage. Une fois optimisée (comme cela fut le cas récemment pour les couches de polysaccharides dans le cadre du projet REFLEX), pour un meilleur contrôle de l'homogénéité et de la stabilité des films, cette technologie offrira une méthode très robuste, sensible, et quantitative, pour le criblage haut débit d'enzymes capables de dégrader la lignine. Avec les autres méthodes de criblages développées dans ce projet de thèse, il serait très intéressant, en particulier, d'explorer le potentiel lignolytique des champignons du rumen non cultivables, grâce à des approches de métatranscriptomique basée sur l'activité, comme celles développée récemment par le groupe Discomics (papier en préparation).

1.2. Propriétés des nouvelles oxydoréductases du rumen

Afin de caractériser la spécificité de substrat de ces enzymes nouvellement découvertes, l'optimisation des conditions réactionnelles a été réalisée en utilisant l'ABTS comme substrat. Des expériences avec ou sans cuivre, et avec ou sans peroxyde d'hydrogène, ont montré qu'aucun des huit clones positifs n'a besoin de cuivre pour être actifs, indiquant qu'elles n'appartiennent pas à la famille des laccases. Bien qu'aucun des clones de l'enrichissement *in vivo* ne montrait d'activité plus importante en présence de peroxyde d'hydrogène, une enzyme de l'enrichissement *in vitro* était active en présence d' H_2O_2 , ce qui indique qu'elle pourrait être une peroxydase. Une caractérisation plus approfondie des profils de température et pH d'activité a été nécessaires pour trouver les conditions optimales d'utilisation avec les substrats d'intérêt (colorants et lignine).

Cela a permis de montrer que ces enzymes sont thermoactives, étant toujours actives à 50°C, et aussi actives en conditions acides (pH 4,5-5,0). Dans leurs conditions optimales d'utilisation respectives, les trois clones choisis pour être caractérisés de façon plus approfondie étaient capables de décolorer des colorants structuralement différents, en particulier des colorants AZO. D'autres expériences sont néanmoins nécessaires pour conclure sur l'activité exacte de chaque enzyme cible. La caractérisation structurale des produits de réaction (solubles et insolubles) par RMN et spectrométrie de masse devrait nous permettre de déterminer les sites de modification du substrat. Cette étape nécessite la purification des enzymes cibles, afin d'éliminer le bruit de fond de l'activité des oxydoréductases d'*E. coli* qui a été détecté dans les clones métagénomiques. Afin d'identifier les gènes responsables des activités détectées, il existe trois approches différentes.

La première, qui est aléatoire, est la mutagénèse de transposition. Elle consiste à insérer aléatoirement des transposons dans les inserts fosmidiques, et à cribler une nouvelle fois les transformants pour identifier ceux qui ont perdu l'activité. Le groupe Discomics a utilisé cette approche de nombreuses fois dans d'autres études. Cependant, les résultats ont souvent été décevants, puisque l'activité est souvent abolie à cause de l'insertion du transposon dans des régions de la séquence qui code pour des régulateurs de transcription pour *E. coli*. Une étude de transcriptomique récente de l'équipe, réalisée sur des clones métagénomiques *E. coli*, a en effet permis de démontrer que l'expression hétérologue de gènes est due à la présence aléatoire de ces séquences dans les inserts métagénomiques, et pas aux promoteurs natifs (papier en préparation).

La seconde approche consiste à tronquer de façon rationnelle des inserts fosmidiques autour des gènes qui ont été identifiés par analyse de séquence comme les cibles les plus probables. Cette stratégie est particulièrement puissante, même si ardue, si plusieurs gènes encodent des enzymes qui sont impliquées dans l'activité ciblée. Néanmoins, elle ne permet pas de produire les protéines hétérologues avec un haut rendement, ni de les taguer pour leur future purification.

La dernière approche consiste à sous-cloner uniquement les meilleurs gènes cibles dans des vecteurs de surexpression. Dans notre cas, l'annotation fonctionnelle a permis d'identifier, pour chaque clone positif, des gènes codant pour des oxydoréductases ainsi qu'un contexte génomique qui semble impliquer une activité synergique des différentes enzymes. Deux des cibles ne présentent pas d'homologie significative avec des enzymes de dégradation des composés polyaromatiques, soulignant l'originalité de ces nouvelles séquences, mais malheureusement aussi, l'impossibilité de déterminer exactement leur famille protéique. La présence dans le troisième clone (32D12) de gènes codant pour une enzyme redox putative de quatre sous-unités fera de l'étape de surproduction de cette cible une étape clé, notamment durant la phase de repliement de la protéine dans le cytoplasme.

Différents vecteurs et souches devraient être testés afin de trouver les meilleures conditions de production. Pour les enzymes monomériques et homo-multimériques, les vecteurs habituels de surexpression, contenant des tags d'intérêt pour la purification seraient les premiers à tester, alors que pour l'enzyme avec quatre sous-unités, un système spécial de clonage, utilisé pour produire l'enzyme entière en une seule étape, serait nécessaire. Une souche conçue pour aider au repliement des protéines qu'elle produit pourrait être utilisée, comme les cellules compétentes ArcticExpress (Agilent), qui produisent des protéines chaperonnes adaptées au froid pour aider au repliement des protéines exprimées, par exemple. Bien sûr, un grand nombre d'expériences et de multiples combinaisons seraient nécessaires pour trouver les meilleures conditions de clonage et d'expression. Une fois purifiées, ces enzymes seront caractérisées plus profondément afin de chercher à comprendre leurs relations structure/fonction et de quantifier leur potentiel de bioremédiation des colorants et de valorisation des dérivés ligneux issus de différentes industries (en terme d'efficacité, stabilité, rendement d'expression de protéines recombinantes, et toxicité/facilités de récupération/biodégradabilité et recyclage des produits réactionnels).

La décoloration des colorants est en effet particulièrement intéressante pour l'industrie textile, considérant la quantité de colorant qu'elle libère dans l'environnement. La formation de précipités colorés pourrait être utilisée pour éliminer cette partie des déchets par centrifugation, filtration ou sédimentation par exemple. L'immobilisation d'enzymes pourrait aussi être envisagée pour stabiliser les biocatalyseurs, les recycler et diminuer leurs coûts de production. La décoloration des colorants serait, dans ce cas, réalisée en réacteur à fonctionnement discontinu seulement, la précipitation des produits de réaction empêchant l'utilisation de réacteurs à lit fixe. Les réacteurs microbiens basés sur la production d'enzymes dans des souches hétérologues permettant une sécrétion efficace des enzymes, si possible des souches GRAS (Generally Recognized As Safe), comme *Bacillus subtilis*, pourraient aussi être envisagés pour contourner les obstacles économiques dus à la production et purification des enzymes.

Il en va de même pour la lignine alkali, un déchet produit dans l'industrie papetière. L'impact positif sur l'environnement pourrait être très élevé, puisque le procédé Kraft est prédominant dans l'industrie papetière. Le fait que ces enzymes soient aussi capables de dégrader la lignine sulfonée, un sous-produit de la méthode de fabrication de pâtes au bisulfite est aussi très intéressant. En effet, bien que ce procédé ait été majoritairement remplacé par le procédé Kraft, il est tout de même encore utilisé de nos jours. Il sera nécessaire de déterminer le rendement de dépolymérisation du substrat et la structure des produits de dégradation. La capacité de nos nouvelles enzymes redox à dégrader les extraits de lignines issues de différentes sources végétales, ainsi que les déchets de lignine récalcitrants aux procédés de prétraitement de la biomasse végétale, devra aussi être testée. Ces derniers aspects nécessitent une expertise supplémentaire en caractérisation structurale

de la lignine. Dans ce but, une collaboration avec L. Eltis (Université de Vancouver), a été initiée.

Enfin, si nous avons découvert plusieurs nouvelles enzymes redox bactériennes capables de dégrader différents substrats polyaromatiques, y compris des dérivés de la lignine, leur rôle physiologique dans le rumen est toujours obscur. Nous n'avons en effet aucune preuve expérimentale de l'action de ces enzymes sur la lignine native, même si deux d'entre elles (produites par les clones 10N15 et 32D12) sont codées par des systèmes multigéniques codant aussi pour des hémicellulases impliquées dans la dégradation du réseau polysaccharidique de la paroi végétale. La caractérisation de la structure de lignines natives, avant et après incubation avec ces extraits enzymatiques, devrait permettre de confirmer ou non cette hypothèse rapidement.

2. Des estérases issues du rumen comme outils de bioremédiation

Une grande diversité d'estérases, actives sur plusieurs substrats différents, a aussi été découverte grâce à la stratégie de criblage multi-étapes utilisée dans cette étude.

Une caractérisation plus en profondeur a été réalisée après les criblages primaires et secondaires sur des esters chromogéniques, afin d'étudier leur potentiel de dégradation des polluants de la famille des carbamates. Deux classes de substrats ont été choisies : l'Impranil DLN, un polyuréthane commercial, deux insecticides de la famille des carbamates (fenoxycarb et fenobucarb), et un herbicide de la famille des thiocarbamates (prosulfocarb). Le criblage sur Impranil DLN en milieu minimum solide et milieu riche solide a permis l'isolation de 57 clones issus des 85 du criblage primaire. Deux clones ont été choisis pour leur efficacité à former un halo de dégradation sur milieu riche et à croître sur milieu minimum avec Impranil DLN comme seule source de carbone. Le clone 44112 est aussi actif sur les trois substrats *p*NP, alors que 12F23 est seulement faiblement actif sur *p*NPB. L'efficacité de dégradation du polyuréthane, et ces différences de spécificité de substrats ont guidé notre choix, afin d'éviter d'obtenir des séquences redondantes, ce qui arrive souvent après criblage de banques métagénomiques.

Les résultats HPSEC et MALDI-TOF ont confirmé que des modifications structurales de l'Impranil DLN se sont produites pendant l'hydrolyse. Cependant, puisque la structure exacte des chaînes latérales de ce polymère commercial ne sont pas accessibles (malgré une structure potentielle qui a récemment été proposée), il est impossible, à cette étape, d'identifier les sites de coupure exacts de ce substrat. Le tampon réactionnel, ainsi que les conditions d'analyses HPSEC vont devoir être optimisés, afin d'éviter l'agrégation du polymère (liée à la lyophilisation dans le tampon PBS), et d'augmenter les rendements d'élution. Ensuite, des analyses RMN et de spectrométrie de masse des produits réactionnels

devraient permettre de déchiffrer le mécanisme d'action de ces nouvelles estérases sur Impranil DLN.

Les réactions sur de plus petits carbamates montrent que seul le clone 44I12 est capable de convertir les composés phytosanitaires testés, en particulier le fenobucarb, en 2-sec-butyphénol. Cela souligne la grande flexibilité de substrats de CE2_Ubrb, qui peut dégrader les trois substrats *p*NP testés, le Tween 20, et deux polluants de la famille des carbamates structuralement différents. CE1_Ubrb, quant à elle, ne peut dégrader qu'un substrat *p*NP et Impranil DLN. Les analyses de séquences ont montré que CE1 et CE2_Ubrb appartiennent aux familles lipolytiques V et IV respectivement. Ces deux enzymes sont, aujourd'hui, les seuls membres de ces familles à présenter une activité de dégradation des carbamates. En effet, jusqu'à ce travail de thèse, seule la famille VII était connue pour avoir des membres capables de dégrader les carbamates. Il faut aussi noter que le meilleur homologue caractérisé de CE1_Ubrb est une enzyme capable de dégrader le PLA. Des tests préliminaires ne nous ont pas permis de détecter une activité de dégradation du PLA par les clones produisant CE1 et CE2_Ubrb. Néanmoins, la dégradation du PLA et d'autres plastiques avec des structures homologues seront re-testées en utilisant des méthodes récemment optimisées, la dégradation de plastiques étant un domaine de recherche capital pour leur recyclage.

Comme les enzymes redox décrites précédemment, l'étape suivante sera de sous-cloner les gènes codant pour CE1 et CE2_Ubrb dans un vecteur de surexpression, afin de les purifier et les cristalliser pour identifier les facteurs structuraux liés à la flexibilité de substrat. Cela permettra aussi de déterminer leur efficacité catalytique et leur stabilité, qui sont deux paramètres cruciaux pour l'industrialisation des enzymes. En effet, les polyuréthanes constituent une part non-négligeable (6-7%) de la masse totale de plastiques produite à travers le monde, et couvre plus de la moitié du marché des revêtements automobiles. Parmi eux, l'Impranil DNL est utilisé dans les textiles, le cuir et les revêtements d'avion. C'est pourquoi la dégradation d'un polluant à utilisations aussi diverses est d'un grand intérêt. Le fait que les clones soient capables de pousser sur un milieu sélectif avec l'Impranil DLN comme seule source de carbone montre leur capacité à métaboliser complètement ce substrat, ce qui est très intéressant pour des procédés de bioremédiation microbienne. Il en va de même pour le fenobucarb. En effet, cet insecticide de la famille des carbamates est utilisé dans plusieurs pays producteurs de riz comme élément de contrôle de la population de thrips, cicadelles et autres sauterelles. Malheureusement, il a été démontré qu'une large part des insecticides est transférée des effluents des sols aux eaux souterraines et aux nappes phréatiques. La dégradation du fenobucarb est donc devenue capitale pour l'environnement. Sa bioremédiation grâce à des procédés microbiens ou enzymatiques basés sur l'utilisation de CE2_Ubrb est donc d'un intérêt particulier, notamment puisque cette enzyme est le premier membre de la famille lipolytique IV à catalyser cette réaction, ce

qui permettrait de contourner la propriété intellectuelle qui couvre les enzymes appartenant à la famille VII capables de dégrader les carbamates.

3. Effet de l'enrichissement sur le profil fonctionnel du microbiome du rumen

Les deux banques métagénomiques obtenues après enrichissement *in vivo* (IVVE) et *in vitro* (IVTE) sur paille de blé ont été criblées pour rechercher des enzymes dégradant des polysaccharides, des estérases et des oxydoréductases. Les résultats soulignent l'importante différence des profils exprimés dans ces banques. Un rendement bien plus faible de clones actifs sur Tween 20 et AZCL-xylane a été observé, ainsi qu'un rendement plus élevé sur AZO-galactomannane, AZCL- β -glucane et AZO-CMCellulose, en comparant les résultats obtenus sur la banque IVTE par rapport à IVVE. Cette tendance peut être expliquée par la durée et l'efficacité de la dégradation du substrat. Alors qu'*in vitro*, la réaction dure sept jours, avant une nouvelle inoculation, *in vivo*, le temps de passage varie entre 20 et 48h. Ainsi, dans le rumen, la biomasse facilement accessible pourrait être dégradée en premier, suivie des hémicelluloses, et d'une faible fraction de cellulose amorphe, ce qui pourrait expliquer le nombre plus important de clones positifs obtenus sur les substrats hémicellulolytiques, et le faible nombre de cellulases. D'un autre côté, pour l'enrichissement *in vitro*, la dégradation du substrat durant plus longtemps, le réseau lignocellulolytique pourrait être plus largement attaqué, rendant la cellulose plus accessible aux cellulases, alors très actives au moment de l'échantillonnage.

Les tests de dégradation de différentes celluloses chromogéniques ont montré que quelle que soit la méthode d'enrichissement, la cellulose cristalline n'est pas dégradée par les clones isolés. Ce résultat, bien que décevant, n'est pas surprenant, puisqu'à notre connaissance, aucune enzyme capable de dégrader la cellulose cristalline n'a été isolée de microbiomes digestifs de mammifères ou d'insectes, en utilisant des approches similaires. C'est probablement dû au fait que bien que la cellulose cristalline est chimiquement homogène, aucune enzyme n'est capable de l'hydrolyser seule, alors que les dérivés solubles de la cellulose peuvent l'être facilement sous l'action d'endo- β -1,4-glucanases. La cellulose cristalline peut être hydrolysée par un groupe d'enzymes synergiques ou des complexes multi-enzymatiques comme les cellulosomes. Dans notre étude, la stratégie de construction de la banque métagénomique a permis le clonage de groupes multigéniques entiers, mais l'expression concomitante chez *E. coli* de tous les gènes d'un même locus est hautement improbable. De plus, même si certaines des séquences CAZy semblent contenir des modules dockérine, qui témoignent de la présence de structures semblables aux cellulosomes (comme dans le clone 49M8), il n'est pas certain que ces complexes enzymatiques ciblent la cellulose dans la bactérie native, puisque les structures semblables aux cellulosomes ne sont pas spécifiques de la dégradation de la cellulose. De plus, les substrats chromogéniques et ceux traités chimiquement (comme Avicel) ne sont pas structurellement identiques à la

cellulose semi-cristalline native, ce qui pourrait induire un biais dans la caractérisation de la spécificité de substrat. Tester nos clones sur de la cellulose semi-cristalline native, comme les fibres de coton par exemple, nous permettrait d'évaluer le potentiel de ces enzymes pour la dégradation de matériaux hautement cristallins.

De plus, à partir des 172 clones positifs issus du criblage primaire, 68 sont actifs sur plus d'un substrat polysaccharidique. Cette caractéristique a déjà été observée plusieurs fois pour des clones fosmidiques exprimant des activités CAZymes, quel que soit l'écosystème digestif criblé. Dans les bactéries, les systèmes synergiques de transport et de dégradation de structures glucidiques complexes, comme celles de la paroi végétale et des glycanes eucaryotes, sont en effet codés par des systèmes multigéniques, comme les Polysaccharide Utilization Loci (PULs) décrits pour les bactéries du phylum Bacteroidetes. Dans notre étude, beaucoup de groupes multigéniques codant pour des CAZymes ont été identifiés, la plupart d'entre eux issus de Firmicutes non cultivés. Cela serait donc faux de les nommer PULs, puisque la description des PUL est spécifique des Bacteroidetes, et que nous ne pouvons pas prouver que dans la souche native, tous les gènes d'un locus impliqué dans ces activités sont co-transcrits. Cependant, nos données, avec celles publiées précédemment par plusieurs groupes dont le groupe Discomics, indiquent que les loci (méta)génomiques impliqués dans la dégradation de structures polysaccharidiques complexes ne sont pas spécifiques aux Bacteroidetes. Seules des études transcriptomiques, réalisées sur des souches cultivables natives contenant des loci similaires nous permettraient de le prouver. De plus, certains groupes multigéniques impliqués dans la dégradation de la paroi végétale découverts dans cette étude codent aussi pour des enzymes redox qui sont probablement capables d'agir sur la lignine. Ceci est très original, puisque les CAZymes et ligninases bactériennes, comme les laccases, ne sont pas connues pour être codées par les mêmes loci. Des analyses de syntonie des loci métagénomiques découvert dans le cadre de cette étude, ainsi que des études transcriptomiques réalisées sur les clones métagénomiques et sur des souches cultivables contenant des structures génomiques homologues, nous permettraient de soutenir cette hypothèse, et de revisiter le concept de co-expression de gènes codant pour des ligninases et CAZymes bactériennes.

Comme c'est le cas pour d'autres microbiomes digestifs, plus de cent clones positifs produisant des activités CAZymes ont été identifiés, à partir de quelques Gb d'ADNc métagénomique seulement. Même si le criblage secondaire nous a permis de discriminer les clones selon leur profil fonctionnel, et ainsi d'éviter trop de redondance au niveau des séquences, il restait difficile d'identifier ceux qui présentaient le plus fort potentiel de dégradation de la biomasse végétale. Pour cela, nous avons quantifié l'efficacité de la dégradation de la lignocellulose pour chaque clone positif, en mesurant la quantité de sucres réducteurs libérés à partir du substrat végétal brut utilisé pour l'enrichissement, la paille de blé. Nous avons dû optimiser une méthode décrite précédemment par Song et al. pour cribler au format microplaque l'activité de mutants d'hémicellulases surexprimés dans un

plasmide sous contrôle d'un promoteur fort. En revanche, dans les fosmidés et l'hôte Epi100 (*E. coli*), l'expression de gènes recombinants est beaucoup plus faible, ce qui nécessite une grande sensibilité des méthodes de criblage. Dans notre cas, en modifiant la méthode de quantification des sucres réducteurs, nous avons augmenté la sensibilité de notre crible par dix. Cela a permis de détecter de très faibles activités de dégradation de matrices végétales réelles. C'est la première fois que des clones métagénomiques ont été testés sur ce type de biomasse lignocellulosique brute, utilisée en bioraffinerie. L'efficacité de dégradation de la paille des meilleurs clones obtenus à partir des deux banques est similaire. De plus, leur efficacité n'est pas corrélée à la polyspécificité des extraits enzymatiques, les clones les plus actifs n'étant actifs que sur une ou deux structures polysaccharidiques. Pour ces clones métagénomiques, l'efficacité de dégradation de la paroi végétale est donc due plutôt à un niveau d'expression et/ou d'activité plus élevé, plutôt qu'à une activité synergique de CAZymes. Il est évident que cette expérience n'est pas représentative du fonctionnement de consortia microbien lignocellulolytiques, qui produisent simultanément une batterie de lignocellulases pour s'attaquer à la complexité de la structure de la paroi végétale.

Nous avons donc essayé de reproduire ces effets synergiques, afin d'exploiter la diversité des activités que nous avons découvertes pour créer des cocktails lignocellulolytiques, et étudier le rôle spécifique des estérases et enzymes redox dans l'efficacité de dégradation de la paille.

Plusieurs clones positifs ont été choisis pour créer des mélanges de lignocellulases selon deux critères : la production de plusieurs activités complémentaires nécessaires pour la dégradation de la paroi végétale, due aux clusters multigéniques codant pour des batteries de CAZymes, et potentiellement aussi d'oxydases, et l'efficacité des clones pour dégrader la paille de blé. Différents mélanges ont été créés, en ajoutant séparément les oxydoréductases et, si possible, les clones estérases.

Cette étude a permis de concevoir des cocktails d'enzymes aux réels effets synergiques, puisqu'au moins un cocktail issu de chaque banque est capable de libérer autour de deux fois plus de sucres réducteurs que la quantité théorique calculée à partir la somme des activités de chaque clone, individuellement. Bien que les estérases impactent positivement l'efficacité de dégradation de la paille, les enzymes redox, étrangement, ont tendance à inhiber les CAZymes. Cela peut être dû à la modification du réseau de lignine, et éventuellement à sa depolymérisation /repolymérisation, ce qui affecterait l'accessibilité des cellulases et hémicellulases au réseau polysaccharidique. Néanmoins, cette hypothèse est hautement spéculative, le mode d'action des nouvelles enzymes redox n'étant pas encore identifié à ce stade. Les performances de nos cocktails sont bien sûr limitées en comparaison des cocktails lignocellulosiques commerciaux et des consortia microbiens, à cause du très faible niveau d'expression des enzymes hétérologues. La plupart des protéines produites par ces clones positifs sont en effet celles d'*E. coli* elle-même. De plus, la combinatoire totale des

clones positifs isolés lors de cette étude est loin d'avoir été exploitée, puisque nous avons choisi seulement quelques clones pour la création de ces cocktails. Néanmoins, nous avons établi la preuve de concept de cette stratégie, qui pourrait être mise en place pour exploiter les centaines de clones lignocellulolytiques issus de ce travail de thèse, mais aussi de travaux précédents ciblant d'autres microbiomes. Une fois que les meilleurs cocktails de clones positifs auront été identifiés, les gènes codant pour les lignocellulases qu'ils contiennent seront sous-clonés dans des vecteurs de surexpression afin de les produire et les purifier pour pouvoir caractériser en profondeur leur efficacité catalytique, seuls ou avec leurs partenaires de dégradation de la paroi végétale.

De plus, à cette étape de l'étude, nous ne savons pas si les lignocellulases identifiées sont les mêmes que celles produites dans le meilleur consortia lignocellulolytique créé par A. Lazuka et G. Hernandez-Raquet. Des analyses de métaprotéomique sur les consortia artificiels devraient nous permettre de comparer les séquences des clones positifs et d'évaluer précisément le biais de la stratégie de métagénomique basée sur la détection de l'activité. De plus, des analyses du CAZome du bactériome du rumen avant et après les deux enrichissements devraient être possibles rapidement grâce au développement d'une puce fonctionnelle contenant toutes les séquences CAZy disponibles dans les bases de données génomiques. Même biaisée par le contenu des bases de données publiques, qui est loin de représenter la diversité des CAZy chez les bactéries cultivables et non-cultivables, cette approche devrait permettre de contribuer à la compréhension de l'effet des enrichissements sur le profil fonctionnel de l'écosystème du rumen. Ces données seront intégrées dans les futures analyses d'ADNr 16S, pour chercher à comprendre les effets des enrichissements sur la diversité taxonomique des consortia.

Plus généralement, les enzymes identifiées ici semblent être les facteurs clé de la dégradation de la lignocellulose dans le rumen, puisque plusieurs de nos séquences CAZy, et même des clusters de gènes entiers, ont déjà été isolés à partir d'autres études métagénomiques ciblant d'autres cohortes de vaches dans le monde entier. L'analyse de la prévalence et de l'abondance de nos séquences dans les catalogues de gènes obtenus par séquençage massif aléatoire du métagénome du rumen bovin devrait bientôt nous permettre de mettre en évidence les enzymes qui jouent un rôle clef pour l'écosystème. Ces résultats, ainsi que l'originalité des séquences, vont guider la sélection des cibles les plus pertinentes pour les analyses structurales et fonctionnelles ultérieures.

Le nombre de modules CAZy inconnus, trouvés dans les deux banques, est en effet particulièrement élevé, ouvrant des perspectives très intéressantes pour la caractérisation de nouvelles familles d'enzymes et éventuellement de nouvelles fonctions. Nous avons aussi identifié des clones contenant des activités qui ne sont liées à aucune des familles CAZy connues annotées dans ces séquences. Nous allons donc devoir rapidement identifier les gènes responsables de ces fonctions, afin d'élargir le panel de fonctions de modification des

glucides de familles d'enzymes déjà caractérisées ou non. L'analyse du contexte génomique de ces cibles est en cours, afin de guider la caractérisation approfondie de leur spécificité de substrat.

Pour conclure, notre travail a mis en lumière l'intérêt du microbiote du rumen bovin comme source d'enzymes bactériennes très originales, qui constituent des acteurs clés du catabolisme des constituants alimentaires, et dont la promiscuité peut être exploitée pour la dépollution. Plusieurs de ces enzymes sont des cibles particulièrement pertinentes pour améliorer la compréhension des relations entre structure enzymatique et spécificité de substrats, tout en présentant à un potentiel élevé pour les bioraffineries et la bioremédiation. De plus, tout en permettant de mieux comprendre le fonctionnement de l'écosystème du rumen bovin, des échelles moléculaires à écosystémiques, l'approche développée ici pourrait être utilisée pour cibler d'autres substrats et microbiotes, en particulier pour déchiffrer les mécanismes de dégradation de xénobiotiques dans l'intestin des mammifères, les sols ou les eaux usées.

Annex I:

Functional metagenomics: construction and high-throughput screening of fosmid libraries for discovery of novel carbohydrate-active enzymes

Lisa Ufarté^{1,2,3}, Sophie Bozonnet^{1,2,3}, Elisabeth Laville^{1,2,3}, Davide A. Cecchini^{1,2,3}, Sandra Pizzut-Serin^{1,2,3}, Samuel Jacquiod⁴, Sandrine Demanèche⁴, Pascal Simonet⁴, Laure Franqueville⁴, Gabrielle Potocki-Veronese^{1,2,3}

¹Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

²INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France

³CNRS, UMR5504, F-31400 Toulouse, France

⁴Université de Lyon, Laboratoire Ampère, CNRS UMR5005, Ecole Centrale de Lyon, F-69134 Ecully, France

Springer protocols, Microbial Environmental Genomics (2016)

Functional Metagenomics: Construction and High-Throughput Screening of Fosmid Libraries for Discovery of Novel Carbohydrate-Active Enzymes

Lisa Ufarté, Sophie Bozonnet, Elisabeth Laville, Davide A. Cecchini, Sandra Pizzut-Serin, Samuel Jacquiod, Sandrine Demanèche, Pascal Simonet, Laure Franqueville, and Gabrielle Potocki Veronese

Abstract

Activity-based metagenomics is one of the most efficient approaches to boost the discovery of novel biocatalysts from the huge reservoir of uncultivated bacteria. In this chapter, we describe a highly generic procedure of metagenomic library construction and high-throughput screening for carbohydrate-active enzymes. Applicable to any bacterial ecosystem, it enables the swift identification of functional enzymes that are highly efficient, alone or acting in synergy, to break down polysaccharides and oligosaccharides.

Key words Metagenomic DNA, Fosmidic libraries, High-throughput screening, Carbohydrate-active enzymes, Complex glycans

1 Introduction

Early metagenomic studies focused on exploring microbial diversity through sequencing of ribosomal RNA sequences and later, with the emergence of powerful sequencing technologies, of functional DNA recovered from environmental samples. Generated data usually cover several gigabases of sequence information in the form of short sequences, which need to undergo an assembly pipeline in order to extract useful gene information. However, gene annotation provides only a functional potential to the annotated genes, according to their sequence homology, the real activity requiring an experimental demonstration.

Activity-based metagenomics allows to by-pass these challenges, as proven by numerous studies dedicated to the discovery of novel enzymes, in particular Carbohydrate-Active Enzymes or CAZymes [1]. Indeed, carbohydrates, in particular glycans, assure key and highly versatile functions in the living world, for energy storage,

cell signaling, recognition, or shape maintain. Carbohydrate metabolism is thus crucial for all organisms, and requires a large panel of CAZymes to cleave, modify, or create osidic linkages. Moreover, as plant polysaccharides constitute the main source of renewable carbon, the extraordinary diversity and natural efficiency of microbial CAZymes can be exploited to develop green processes of plant biomass conversion into biofuels or sugar-based materials including surfactants, fine-chemicals, secondary metabolites, drugs, vaccines, among others.

As a result, in the last years, numerous studies have been published, exploiting the immense potential of functional metagenomics to explore various bacterial ecosystems. Indeed, many ecosystems like mammal and insect guts, soil and composts, are more or less specialized in polysaccharide breakdown, enabling the discovery of carbohydrate catabolism-related enzymes, like glycoside hydrolases.

Functional metagenomics consists in (1) constructing large libraries of thousands to several hundred thousands of recombinant clones, carrying metagenomic DNA fragments sizing between 2 and 200 kbp, cloned into plasmids, cosmids, fosmids or even bacterial artificial chromosomes, (2) screening them for the targeted activities, and (3) sequencing the screening hits in order to identify the genes that are responsible for the observed phenotypes (Fig. 1). By using this approach, several hundreds of novel CAZymes were

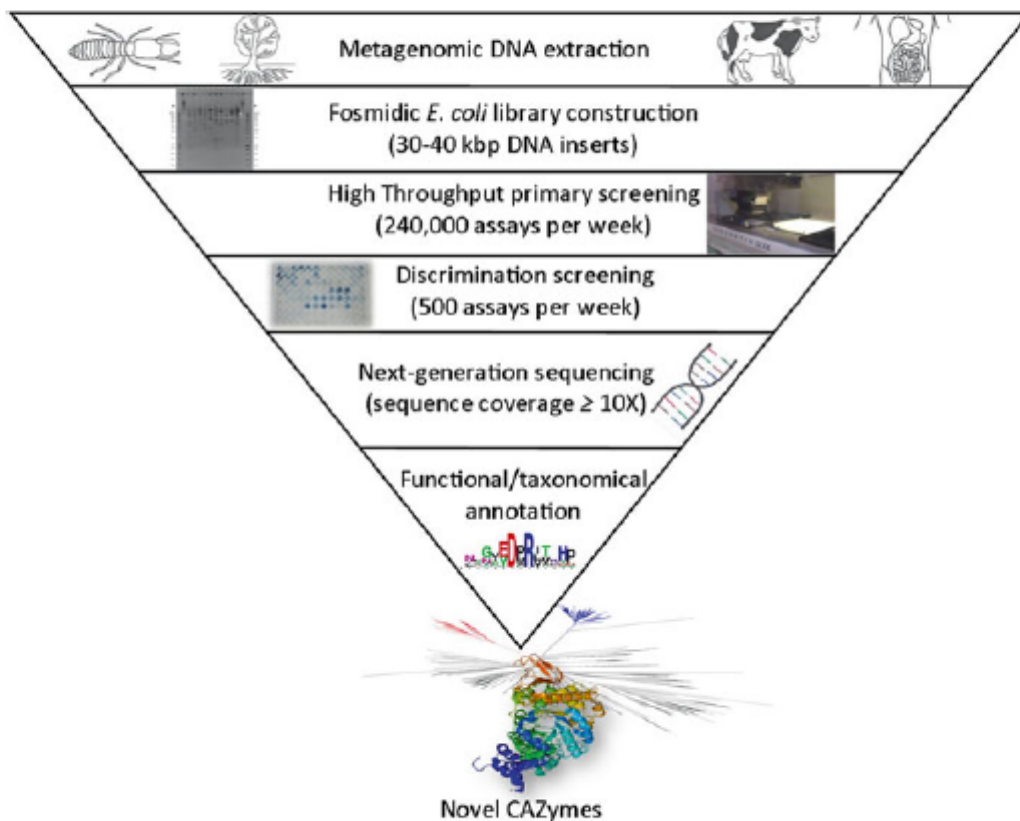


Fig. 1 Multistep activity-based metagenomic strategy for CAZyme discovery

retrieved from metagenomes these last few years [2, 3], most of them presenting very original sequences [4], sometimes belonging to novel protein families [5] and/or displaying inherited key functions of carbohydrate foraging [6] which would not have been predicted by genomic or metagenomic sequence analysis.

In this chapter, we describe a robust and inexpensive procedure of high-throughput functional exploration of bacterial ecosystems (soil being used as an example), to drive in-depth metagenomic sequencing and focus on genes encoding catabolic CAZymes. Even if alternative strains with different expression and secretion capabilities can be used [7, 8], we detail the construction of *E. coli* fosmidic metagenomic libraries, as it allows to easily and rapidly explore extremely large sequence spaces, covering several Gbp of metagenomic DNA. Search for functional CAZyme encoding genes consists here in applying a multistep screening approach to (1) isolate clones producing catalysts with the desired specificity toward polysaccharidic and oligosaccharidic substrates, (2) discriminate endo- and exo-hydrolytic activities (Fig. 2a, b) and even discover enzyme cocktails, encoded by multigenic clusters that are frequently found on large metagenomic fosmidic inserts (sizing between 30 and 50 kbp) that are involved in the break-

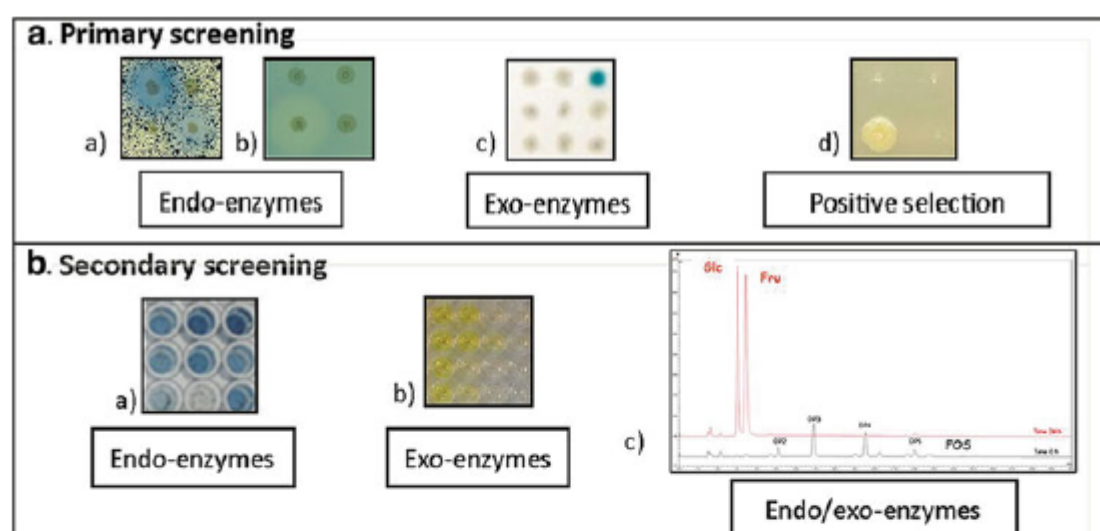


Fig. 2 High-throughput screening of *E. coli* metagenomic libraries for endo- and exo-acting glycanases. (a) Pictures of primary screening results (solid medium) on (a) insoluble AZCL-polysaccharides, *blue halos* showing the release of soluble AZCL-oligosaccharides; (b) solubilized Azo-polysaccharides, *clear halo* showing the degradation of the colored polymer; (c) X-mono/oligosaccharides, *blue color* showing the release of free X-compounds; (d) minimal medium supplemented with oligosaccharides as carbon source, the sole growing clone being able to degrade the targeted oligosaccharides. (b) Pictures of secondary screening results (liquid medium) on (a) AZCL-polysaccharides, *blue color* showing the release of soluble AZCL-oligosaccharides in the reaction medium; (b) *pNP*-mono/oligosaccharides, *yellow color* showing the release of free *pNP*-compounds in the reaction medium; (c) HPAEC-PAD analysis of, in *black*, oligosaccharides substrate (fructo-oligosaccharides as an example) before enzymatic hydrolysis; in *red*, hydrolysis reaction products (glucose and fructose as an example) after 24 h of reaction

down of plant cell wall. If the automated solid plate assays used in the primary screens can be easily carried out by only one person at a throughput of 240,000 assays per week (corresponding to the screening of 20,000 clones for 12 different activities) for only few k€, the discriminating assays in liquid media are used with a lower throughput of around 500 assays per week. However, they are highly recommended in order to avoid total or partial sequence redundancy between the hits presenting the same activities. Sequence redundancy can also be avoided by choosing to screen several little libraries (of few dozens of thousands clones) constructed from different samples rather than only a large one (of hundred thousand clones) issued from one unique sample.

This highly generic approach is applicable to mine all complex bacterial communities for novel catabolic CAZymes. Depending on their ability to face the structural complexity of plant cell wall, and to use it as the main carbon source for growing and maintaining themselves in their habitat, the hit rates will vary from less than 0.2‰ (for example for soil communities) to more than 4‰ (for highly specialized ecosystems like termite guts [9]). In any case, in order to increase hit yield, we recommend (1) increasing the number of primary screens by using a large diversity of polysaccharidic and oligosaccharidic substrates, varying in nature of glycosidic residues, type of osidic linkages, polymerization degree and ramification content; (2) increasing the library size, preferably constructed from several metagenomic DNA samples, in order to minimize sequence redundancy.

After hit recovery, fosmid sequencing with high coverage (more than 10×) allows to easily identify the genes, or the gene clusters, that are responsible for the screened activity, and their taxonomic origin, sometimes up to species level. As the functional and taxonomical annotation procedures do not differ from those developed for sequenced-based metagenomics and genomics, they will not be detailed in this chapter.

2 Materials

All plastics used are certified free of DNase and DNA. All tools and materials used should be washed and cleaned with 70 % ethanol solution (*see Note 1*). Glass or metal materials, as well as solutions when specified, are sterilized before use (121 °C, 20 min). Prepare all solutions using ultrapure water (18 MΩ cm at 25 °C) and analytical grade reagents. Follow all waste disposal regulations when disposing waste materials.

2.1 DNA Sampling

1. 4-mm sterilized glass beads.
2. 0.2 % Sodium hexametaphosphate (HMP): Add about 400 mL water to a 500-mL measuring tube. Weight 1.0 g HMP and

transfer to the cylinder. Add water to a volume of 500 mL. Mix and transfer to a 1-L glass bottle. Sterilize and store at room temperature.

3. Two 250-mL sterilized polypropylene Nalgene tubes.
4. Sterile gauze.
5. 0.8 % Sodium chloride (NaCl): Add about 700 mL water to a 1-L measuring tube. Weigh 8.0 g NaCl and transfer to the cylinder. Add water up to 1 L. Mix and transfer 500 mL to two 1-L glass bottles. Sterilize and store at room temperature.
6. 1.3 g/mL 5-(*N*-2,3-dihydroxy propylacetamido)-2,4,6-tri-iodo-*N,N'*-bis (2,3 dihydroxypropyl) isophthalamide (Nycodenz[®]) (Axis-Shield): in order to obtain a density of 1.3 g/mL, mix 50 mL water and 40 g Nycodenz[®] in a 100-mL glass bottle with a magnetic stirrer. Stir and heat to 50 °C to dissolve Nycodenz[®]. Remove the stirrer, sterilize and store at room temperature.
7. Tris-HCl-EDTA buffer (TE): 50 mM Tris-HCl (pH 8.0) with 100 mM EDTA buffer.
8. InCert[®] agarose (BMA).
9. Plug molds (Bio-Rad).
10. Lysis buffer A (LA): 50 mM Tris-HCl (pH 8.0), 100 mM EDTA, 5 mg/mL lysozyme, 0.5 mg/mL achromopeptidase.
11. Lysis buffer B (LBB): 50 mM Tris-HCl (pH 8.0), 100 mM EDTA, 1 % lauryl sarcosyl, 2 mg/mL proteinase K.
12. Storage buffer: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA.
13. 0.1 mM Phenylmethanesulfonyl fluoride (PMSF) (Sigma): dilute the weighed powder directly in storage buffer.

2.2 Fosmid Library Construction

1. Low-melting-temperature agarose (Bio-Rad).
2. Tris-acetate-EDTA (TAE): dilute 10× stock solution ten times (Promega).
3. PFGE ladder: lambda bacteriophage DNA (NEB).
4. 1 µg/mL Ethidium bromide: dilute in water.
5. GELase (Epicentre Technologies).
6. EpiFOS[™] Fosmid Library Production Kit (Epicentre, Illumina[®]).
7. 50 mg/mL Chloramphenicol: prepare stock solution in ethanol and filter-sterilize before aliquots storage at -20 °C.
8. Freezing medium: Luria-Bertani medium supplemented with 20 % (w/v) glycerol and 12.5 µg/mL chloramphenicol.

2.3 Media and Solutions for Functional Screening

For libraries using the pEpiFOS-5 Fosmid Vector (EPICENTRE).

1. 1000× Cm stock solution: 12.5 mg/mL chloramphenicol (Cm) in ethanol, stored at -20 °C.

2. 5× Salts stock solution: 18 g/L Na₂HPO₄, 12H₂O, 3.31 g/L KH₂PO₄, 0.53 g/L NaCl, 2.11 g/L NH₄Cl, deionized water (dH₂O) up to 1 L. Autoclave.
3. 500× MgSO₄ stock solution: 1 M MgSO₄ in dH₂O. Autoclave.
4. 333× CaCl₂ stock solution: 0.01 M in dH₂O. Autoclave.
5. 1000× Salts stock solution: 15 g/L Na₂EDTA-2H₂O, 4.5 g/L ZnSO₄-7H₂O, 3 g/L CoCl₂-6H₂O, 1 g/L MnCl₂-4H₂O, 1 g/L H₃BO₃, 0.4 g/L Na₂MoO₄-2H₂O, 3 g/L FeSO₄-7H₂O, 0.3 g/L CuSO₄-5H₂O. Dissolve EDTA and ZnSO₄ in 800 mL of deionized water, adjust pH to 6.0 with HCl/NaOH. Dissolve the other compounds one by one and keep the pH at 6.0. Adjust pH to 4 and the volume to 1 L (*see Note 2*). The solutions are filter-sterilized.
6. 1000× Leucine: 40 g/L in dH₂O (*see Note 3*). Filter-sterilize.
7. 100× Thiamine hypochloride: 10 g/L in dH₂O. After dissolution, adjust pH to 2.0 with 2 N HCl. Filter sterilization and preservation at 4 °C hidden from light.
8. Luria-Bertani-Chloramphenicol (LB-Cm) medium: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 1 mL/L 1000× Cm stock solution. Autoclave LB medium and let it cool at 50 °C before adding Cm stock solution. Prepare 200 mL LB-Cm medium for the underlay and 100 mL for the overlay for each large agar plate (QTray, 24.5 cm × 24.5 cm) (*see Note 4*). For solid medium, add 15 g/L agar.
9. Minimal (M9) medium: 15 g/L agar, 200 mL/L salts (5×) stock solution, 2 mL/L MgSO₄ (500×) stock solution, 3 mL/L CaCl₂ (333×) stock solution, 1 mL/L salts (1000×) stock solution, 10 mL/L thiamine hypochloride (100×) stock solution, 1 mL/L leucine (1000×) stock solution, 1 mL/L Cm (1000×) stock solution. For each QTray (24.5 cm × 24.5 cm plates), prepare 200 mL medium for the underlay and 100 mL for the overlay, 24.5 cm × 24.5 cm (*see Notes 5 and 6*).
10. pEpiFOS-5 library: Set of *Escherichia coli* EPI100 clones arrayed in 384-well plates, each well containing one copy of a fosmid clone in LB + glycerol 8 %.
11. QTrays (24.5 cm × 24.5 cm), sterile (Corning Incorporated).
12. LB + 8 % glycerol: Autoclave separately 500 mL 2× LB and 500 mL glycerol at 16 % (w/v) in deionized water, cool to room temperature, mix and add 1 mL/L Cm stock solution.
13. Azurine-Crosslinked (AZCL)/Azo substrates (used for identification of endo-acting CAZymes): Autoclave separately 500 mL 2× LB agar and 500 mL 2× AZCL/Azo-substrate in water. Cool to 60 °C, and mix the two preparations. The final screening medium contains 2 g/L of these chromogenic substrates.

14. 1000× 5-Bromo-4-chloro-3-indolyl (X-) substrates (used for identification of exo-acting CAZymes): 60 mg/mL in dimethyl sulfoxide (DMSO). The final screening medium contains 60 mg/L of X-substrates.
15. Glycerol stock solution: 30 % glycerol (w/v) in deionized water, autoclaved.
16. Omnitrays (86 cm×128 cm), sterile (Thermo scientific Nunc).
17. Clear 96-well microplates, sterile (Corning Incorporated).
18. Cryotubes, 2 mL sterile (Thermo scientific Nunc).
19. Dinitrosalicylic acid (DNS) solution: 10 g/L DNS, 300 g/L potassium sodium tartrate, 16 g/L NaOH in dH₂O (*see* Notes 7 and 8).
20. 10× *para*-Nitrophenyl (*p*NP-) substrates stock solution: 10 mM in dH₂O (*see* Note 9).
21. AZCL/Azo-substrate stock solution for discrimination screening: 0.2 % (w/v) final concentration in deionized water.
22. Lysozyme stock solution (10×): 5 g/L in activity buffer. Store at -20 °C.
23. 50 mM Potassium phosphate buffer, pH 7.0.
24. 1 M Na₂CO₃ stock solution.
25. 5× Oligosaccharide stock solution: 5 % oligosaccharide (w/v) in deionized water.
26. 1 M NaOH stock solution: in water.
27. Eluent A: 150 mM NaOH.
28. Eluent B: 150 mM NaOH, 500 mM CH₃COONa.

2.4 Equipments

1. Eppendorf centrifuge 5810 R, with swing-bucket rotor A-4-81, fixed angle-rotor F-34-6-38 (+ adaptors for 50 and 15 mL Falcon tubes), and a rotor for Eppendorf tube.
2. Pulsed-field CHEFDRII electrophoresis system (Bio-Rad).
3. Liquid handling automat operating in sterile conditions (Biomek 2000, Beckman, Fullerton, CA).
4. Pump PM600 Jouan.
5. Colony picker QPixII (Genetix, Hampshire, UK). Colony picking and microplate replicating.
6. Microtiterplate shaker incubator (Multitron, Infors, Massy, France).
7. Automated microplate gridder (K2, KBiosystem, Basildon, UK): replication of microtiterplate organized libraries (384 well plates) on solid agar plates (QTrays).
8. Bioblock Scientific Vibra-Cell 72412 ultrasonic processor.

9. Microplate spectrophotometer (e.g., Sunrise, TECAN, Männedorf, Switzerland).
10. Dionex ICS-3000 system (Dionex Corp., Sunnyvale, CA).
11. CarboPac PA100 46250 column and guard column (Dionex).

3 Methods

3.1 DNA Sampling

3.1.1 Soil Sampling

1. Collect soil core samples of 6 cm in diameter from surface soils (0–20 cm) by using geostatistical methods as described for example by Atteia [10] on a grid of 6.20 × 3.20 m (*see Note 10*).
2. Transfer soil cores as soon as possible to the laboratory in plastic bags.
3. Sieve soil at 2 mm and store it at 4 °C until rapid processing (within a week).

3.1.2 Bacterial Cells Recovery

1. Refrigerate HMP, NaCl and Nycodenz solutions at 4 °C and perform following procedures on ice unless otherwise specified. Mix the equivalent of 50 g of dry soil with 180 mL of HMP and about 20 glass beads and stir strongly (CATSSO stirrer, set to position 1/min) for 2 h at 22 °C.
2. Centrifuge in a swing rotor (Eppendorf A-4-81) at 18 × *g* for 1 min at 10 °C to eliminate coarse particles.
3. Filter supernatant on sterile gaze into a new 250-mL Nalgene tube and centrifuge in a swing rotor at 3,220 × *g* for 20 min at 10 °C.
4. Eliminate supernatant and suspend pellet in 35 mL NaCl (*see Note 11*).
5. Fill two 50-mL falcon tubes with 11 mL Nycodenz solution and carefully add on surface half of the soil suspension in both tubes (*see Note 12*) [11].
6. Centrifuge in a swing rotor at 3,220 × *g* for 40 min at 10 °C without acceleration and deceleration (set to 0).
7. Pipette the white bacterial ring (approximately 4 mL) without disturbing Nycodenz gradient at the interface between Nycodenz and NaCl (*see Note 13*), pool both rings in a single 50-mL falcon tube and fill with NaCl up to 40 mL.
8. Centrifuge the falcon tube with a fixed-angle rotor at 9,000 × *g* for 20 min at 10 °C (*see Note 14*).
9. Eliminate supernatant, wash pellet with 10 mL NaCl and transfer to a 15-mL Falcon tube.
10. Centrifuge with a fixed-angle rotor at 9,000 × *g* for 15 min at 10 °C.

11. Eliminate supernatant, wash pellet with 1 mL NaCl and transfer to a 1.5-mL Eppendorf tube.
12. Centrifuge at $13,000 \times g$ for 5 min at 10 °C.
13. Eliminate supernatant and suspend pellet in 50 μ L TE buffer.

3.1.3 High Molecular Weight DNA Extraction

1. Mix bacterial pellets with an equal volume of molten 1.6 % InCert[®] agarose (*see Note 15*) [12], transfer into disposable plug mold.
2. Let it stand at 4 °C until solidification, then unmold the solidified cell suspension and transfer into a 50-mL Falcon tube.
3. Add 45 mL of LA solution and incubate at 37 °C for 6 h.
4. Transfer the plug into a new 50-mL Falcon tube, add 45 mL of LBB solution and incubate at 55 °C for 24 h.
5. Repeat the operation: transfer plug in 45 mL of fresh LBB solution into a new 50-mL Falcon tube and incubate at 55 °C for 24 h.
6. Wash plug in 10 mL of storage buffer containing PMSF for 2×1 h at 50 °C (*see Note 16*).
7. Dialyze the plug in three successive 10 mL storage buffer baths for 8 h and store at 4 °C until use.

3.2 Fosmid Library Construction

1. Prepare 150 mL of 0.8 % low-melting-temperature agarose gel (*see Note 17*) in $1 \times$ TAE buffer (*see Note 18*) and wait for solidification.
2. Transfer high-molecular-weight bacterial DNA trapped in the agarose plug using a sterile pipette tip into wells of solidified gel, place in the pulsed-field electrophoresis system, load PFGE ladder (*see Note 19*) and fill it with $1 \times$ TAE buffer.
3. Migrate for 18 h at 4.5 V/cm with 5–40-s pulse times in $1 \times$ TAE buffer cooled at 12 °C for the whole migration time.
4. After electrophoresis, stain the gel in a solution of ethidium bromide for 30 min at room temperature.
5. Cut DNA fragments between 30 and 50 kbp and recover DNA with GELase following manufacturer's procedure.
6. Clone the extracted metagenomic DNA into fosmid and transform in the *E. coli* strain EPI100 as recommended by the manufacturer.
7. Using a colony picker, select transformants grown on plate supplemented with 12.5 μ g/mL chloramphenicol and transfer them to 384 multiwell plates containing 70 μ L freezing medium per well and incubate at 37 °C for 22 h (*see Note 20*).
8. Duplicate the library and store in two -80 °C different freezers for safety reasons.

3.3 Replication of the Metagenomic Library Prior to Functional Screening

1. The day before replication, place the metagenomic library at 4 °C to allow gentle thawing from -80 °C storage.
2. New 384-well microtiter plates are filled with LB + 8 % Glycerol solution, 70 µL per well, using an automated liquid handling station.
3. The metagenomic library is replicated, using a QPixII colony picker (3 h for 54 microplates, totalizing 20,736 clones).
4. The mother plates are stored back at -80 °C, covered with microplate aluminum sealing tapes. The copy plates are incubated overnight (about 16 h) at 37 °C, covered by porous adhesive membranes.

3.4 High-Throughput Primary Screening

3.4.1 QTray Preparation

1. Sterilize the autoclavable tubing of the PM600 Jouan pump, as well as deionized water to wash the tubing between two different media distribution. The quantity of water depends on the number of substrates (*see Note 21*).
2. Calibrate the pump using sterile water.
3. Pour 200 mL of underlay for each substrate using the pump (*see Note 22*). For chromogenic substrates, use LB medium; for selective growth, use M9 medium. Leave them to dry under the hood, lid off, for ~30 min.
4. Pour 100 mL of overlay medium for each substrate (*see Note 22*). For chromogenic substrates, use LB medium containing 2 g/L of AZCL/Azo-substrate, or 60 mg/L of X-substrates; for selective growth, use M9 medium containing a final concentration of 0.5 % (w/v) oligosaccharidic carbon source. Leave them to dry, stacked under the hood, lid off, for ~30 min (*see Notes 23 and 24*).
5. Until the day of the gridding, stock the plates at 4 °C, upside down.

3.4.2 Gridding

1. When storing the plates at 4 °C, place them at room temperature the day before the gridding.
2. Microtiter plates are gridded on large LB agar plates, using a K2 automated plate replication system. One QTray can accommodate the clones from six 384-well plates, for a total number of 2304 fosmid clones per plate. In 7 h, 54 microtiter plates can be gridded on 12 different substrates.
3. Plates containing arrayed clones are incubated at 37 °C.

3.4.3 Hit Isolation, Selection, and Validation

1. Positive clones are recognized: (1) for AZCL-substrates, thanks to the blue-colored halo formed around the colonies (Fig. 2a.a); (2) on Azo-substrates by the appearance of a discoloration halo around positive clones (Fig. 2a.b); (3) as blue-colored colonies on X-substrates (Fig. 2a.c); (4) as the sole growing colonies on M9 media supplemented with targeted carbon source (*see Note 25*) (Fig. 2a.d).

2. Positive clones are picked from the QTray and streaked on Petri dishes containing LB and chloramphenicol, and grown overnight at 37 °C. For each selected clone, three isolated colonies are selected to inoculate three adjacent wells of a 96 microtiter plate, filled with 200 µL LB Cm (*see Note 26*). The plate is incubated at 37 °C with 200 rpm shaking (shaking throw 25 mm) for ~16 h.
3. This microplate is then gridded on omnitrays containing the same medium used for the primary screening, and incubated at 37 °C, until the awaited phenotype is observed.
4. Colonies from validated wells are streaked on fresh LB Cm plate, and after colony growth, two isolated colonies are picked to inoculate two 3 mL of liquid LB medium. Cells are incubated overnight at 37 °C, under shaking at 200 rpm (shaking throw 25 mm).
5. After overnight growth, 500 µL of culture are mixed into two cryotubes with 500 µL of glycerol stock solution (30 %). The two copies of each hit clones are stored at -80 °C in different freezers for safety concern.

3.5 Discrimination Screening of Validated Hits

3.5.1 Liquid Assays Using AZCL- Polysaccharides and pNP-Sugars

1. From an isolated colony, inoculate a 20 mL culture in liquid LB medium, and cultivate at 37 °C with shaking at 200 rpm overnight (shaking throw 25 mm) (*see Note 27*).
2. Measure the OD at 600 nm with a spectrophotometer.
3. Centrifuge the culture at 12,857 × *g*, for 5 min at 4 °C. Discard supernatant.
4. Suspend the pellet in activity buffer to obtain a final OD at 600 nm of 80 (*see Note 28*).
5. To break the cells, use the sonication method: with the probe at 30 % of the maximal power, do five cycles of 20 s separated by 4 min in ice.
6. Centrifuge the samples at 21,728 × *g* for 10 min.
7. Filter the supernatant with a 0.2 µm filter (Minisart). The solution obtained is called enzymatic extract from now on.
8. To test the activity of enzymatic extracts on AZCL-substrates, mix in hemolysis tubes 500 µL of enzymatic extract, 100 µL of AZCL-substrate solution (0.2 % (w/v) final concentration), and 400 µL of activity buffer. Incubate at 37 °C with regular shaking. After reaction times of 0, 15, 30 min, 1 h and 24 h, transfer 120 µL of reaction in an Eppendorf tube, centrifuge for 1 min at 21,728 × *g*, transfer 100 µL of supernatant into the well of a polystyrene microplate and read the OD at 590 nm, with a plate reader. Positive hits present an increase of the OD over time (Fig. 2b.a).

9. To test the activity of enzymatic extracts on *p*NP-substrates, mix in hemolysis tubes 100 μ L of 10 mM *p*NP-substrate solution, 200 μ L of enzymatic extract, and 200 μ L of activity buffer (the same as suspension buffer). Incubate at 37 °C. After reaction time of 0, 10 and 30 min, mix 50 μ L of the reaction medium with 250 μ L of 1 M Na₂CO₃. Transfer 200 μ L of this medium into another polystyrene microplate and read the OD at 405 nm, with a plate reader. Positive hits present an increase of the OD over time (Fig. 2b.b).

3.5.2 HPAEC-PAD
Analysis to Analyze
Reaction Products
of Oligosaccharide
Degradation [13]

1. Clones are grown at 37 °C in 5 mL LB Cm medium, with shaking at 120 rpm for 24 h (shaking throw 25 mm).
2. Centrifuge the culture for 5 min at 3,214 $\times g$.
3. Resuspend in 1 mL activity buffer, containing 0.5 g/L of lysozyme. Incubate at 37 °C for an hour. Complete cell lysis with a freeze (−80 °C) and thaw (30 °C) cycle.
4. Centrifuge cell debris at 21,728 $\times g$ for 10 min and filter the cytoplasmic extracts with a 0.2 μ m filter (Minisart).
5. Enzymatic reaction medium contains 0.2 mL of the oligosaccharide stock solution and 0.8 mL of cytoplasmic extract. Incubate at 37 °C for 24 h.
6. After 30 s and 24 h of reaction, take a 100 μ L sample out of the reaction medium, and heat at 90 °C for 5 min to deactivate enzymes.
7. Dilute samples 200 times with ultrapure water.
8. Perform HPAEC-PAD analyses on a Dionex ICS-3000 system, equipped with a CarboPac PA100 4 \times 250 column connected to the corresponding guard column. Oligosaccharides are separated at 30 °C, with a flow rate of 1 mL/min with a multistep gradient: 0–30 min (0–60 % B), 30–32 min (60–90 % B), 32–36 min (90–0 % B), and 36–46 min (0 % B). Samples of monosaccharides and oligosaccharides at 5, 10, 15, and 20 mg/L are used as standards (Fig. 2b.c). One unit of activity is defined as the amount of enzyme releasing 1 μ mol of product per minute.

4 Notes

1. The main strategy for prokaryote cell segregation consists in applying a density gradient through centrifugation. Because of their size and density, bacteria will cluster apart from eukaryote cells into a specific fraction of the gradient which can easily be recovered. However, co-extraction of low-density eukaryote cells such as fungi spores and pico-eukaryotes (*see ref. 14*) is a possible source of contamination.

2. The solution is green and becomes dark yellow during preservation at 4 °C.
3. To help leucine dissolution, add 5 M NaOH.
4. Not only the system of overlay/underlay enable the use of less substrate, hence decreasing costs, but it is very important in the case of insoluble substrate such as AZCL-polysaccharide because they naturally sediment at the bottom of the plate during solidification of the agar medium. Pouring 200 or 300 mL of such medium leads to the accumulation of all the substrate far from the surface colonies, and functional enzymes liberated when *E. coli* cells die are located too far away from the substrate. A top layer of 100 mL AZCL-substrate medium brings the insoluble substrate much closer to the recombinant clone.
5. Water and agar are sterilized alone. Other sterile components are added one by one.
6. To avoid solidification of M9 medium before time, warm the bottle of 5× salts stock solution before mixing solutions.
7. Shake overnight at room temperature for better solubilization, inside a volumetric flask.
8. Cover the flask entirely with aluminum paper: the DNS solution is light-sensitive.
9. Some *p*NP substrates are difficult to solubilize. You might need as long as a night for them to be totally dissolved.
10. For a large field, a nested sampling of the kind devised and elaborated by Oliver and Webster (*see ref. 15*) may ensure that the important variation is captured.
11. Depending on soil type, pellet can be difficult to suspend. However, soil solution imperatively needs to be homogeneous without any fragment. Using an ultrasonic bath can help.
12. Nycodenz solution must not be disturbed: soil solution has to lay on surface. To avoid disturbance, use the gravity function of pipet-aid to add soil suspension.
13. Bacterial rings are sometimes difficult to see. Soil pellet usually fill the tube up to the 10-mL graduation, above the Nycodenz solution reach the 15-mL graduation. Bacterial cells are usually near the 15-mL graduation on Nycodenz surface and above cells is the NaCl solution.
14. Do not forget to readjust the centrifuge parameters to maximal acceleration and deceleration values.
15. Warm briefly the cell suspension at 37 °C to avoid premature gelation of agarose.
16. This step can also be performed overnight at room temperature (22 °C) with gentle shaking.
17. Use caution when handling low-melting agarose gel because it is very fragile.

18. TAE buffer is recommended for subsequent enzymatic reactions.
19. Use embedded ladder supplied in a GelSyringe dispenser.
20. This very high-throughput work can benefit from dedicated facilities, gathering all the automats useful for such experimentation: colony picking, microplate replication, liquid transfer, etc. The functional screening described in this chapter has been performed mainly using the ICEO facility (LISBP, INSA Toulouse) dedicated to enzyme screening and discovery, and part of the Integrated Screening Platform of Toulouse (PICT, IBiSA).
21. If you do not have a pump, use a measuring tube to pour the right volume of agar medium.
22. Make sure that the plate is fully horizontal to have an even overlay. Be careful not to pour the overlay on cold agar, as it will solidify too quickly and impedes the obtention of a regular layer. To overcome this, incubate your QTrays containing the solidified underlay at 37 °C for an hour, just before pouring the overlay.
23. Do not let the medium inside the pipe solidify: if you need to, make a closed system within the bottle containing the medium.
24. If you have some medium left, use it to pour omnitrays (media with substrates) and Petri dishes (LB agar) that will be used for validation.
25. Positive clones on Azo, AZCL and X-substrates are usually observed rapidly, between 2 and 7 days of incubation. Growth of the hits obtained by positive selection on oligosaccharides as sole carbon sources is visualized between 5 and 20 days.
26. It is easier to streak the Petri dishes and wait to pick the isolated colonies for a large group of positive clones, so that you can fill the microplates all at once and arrange them as you wish.
27. You can also make a pre-culture the day before in 3 mL of liquid LB medium from a freeze-d sample.
28. An OD at 600 nm of 80 corresponds to the most efficient concentration for sonication.

Acknowledgements

This research was funded by the European Union project MetaExplore, the French Research Agency (Agence Nationale de la Recherche) ANR project Metasoil, and the INRA metaprogramme M2E (project Metascreen).

References

1. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490–D495
2. André I, Potocki-Véronèse G, Barbe S, Moulis C, Remaud-Siméon M (2014) CAZyme discovery and design for sweet dreams. *Curr Opin Chem Biol* 19:17–24
3. Li L-L, McCorkle SR, Monchy S, Taghavi S, van der Lelie D (2009) Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol Biofuels* 2:10
4. Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Martins Dos Santos VAP, Yakimov MM, Timmis KN, Golyshin PN (2005) Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem Biol* 12:895–904
5. Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C, Cantarel BL, Coutinho PM, Henrissat B, Leclerc M, Doré J, Monsan P, Remaud-Simeon M, Potocki-Veronese G (2010) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res* 20:1605–1612
6. Ladevèze S, Tarquis L, Cecchini DA, Bercovici J, André I, Topham CM, Morel S, Laville E, Monsan P, Lombard V, Henrissat B, Potocki-Véronèse G (2013) Role of glycoside phosphorylases in mannose foraging by human gut bacteria. *J Biol Chem* 288:32370–32383
7. Ekkers DM, Cretoiu MS, Kielak AM, van Elsas JD (2012) The great screen anomaly—a new frontier in product discovery through functional metagenomics. *Appl Microbiol Biotechnol* 93:1005–1020
8. Taupp M, Mewis K, Hallam SJ (2011) The art and design of functional metagenomic screens. *Curr Opin Biotechnol* 22:465–472
9. Bastien G, Arnal G, Bozonnet S, Laguerre S, Ferreira F, Fauré R, Henrissat B, Lefèvre F, Robe P, Bouchez O, Noirot C, Dumon C, O'Donohue M (2013) Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnol Biofuels* 6:78
10. Atteia O, Dubois JP, Webster R (1994) Geostatistical analysis of soil contamination in the Swiss Jura. *Environ Pollut (Barking Essex)* 86:315–327
11. Courtois S, Frostegård A, Göransson P, Depret G, Jeannin P, Simonet P (2001) Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ Microbiol* 3:431–439
12. Ginolhac A, Jarrin C, Gillet B, Robe P, Pujic P, Tuphile K, Bertrand H, Vogel TM, Perriere G, Simonet P, Nalin R (2004) Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl Environ Microbiol* 70:5522–5527
13. Cecchini DA, Laville E, Laguerre S, Robe P, Leclerc M, Doré J, Henrissat B, Remaud-Siméon M, Monsan P, Potocki-Véronèse G (2013) Functional metagenomics reveals novel pathways of prebiotic breakdown by Human gut bacteria. *PLoS One* 8:e72766
14. Moreira D, López-García P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol* 10:31–38
15. Oliver MA, Webster R (2010) Combining nested and linear sampling for determining the scale and form of spatial variation of regionalized variables. *Geogr Anal* 18:227–242

Annex II:

Inventory of the GH70 enzymes encoded by *Leuconostoc citreum* NRRL B-1299 – identification of three novel α -transglucosylases

Delphine Passerini^{1,2,3,*}, Marlène Vuillemin^{1,2,3,*}, Lisa Ufarté^{1,2,3},
Sandrine Morel^{1,2,3}, Valentin Loux⁴, Catherine Fontagné-Faucher⁵,
Pierre Monsan^{1,2,3}, Magali Remaud-Siméon^{1,2,3} and
Claire Moulis^{1,2,3}

¹ Université de Toulouse, Institut National des Sciences Appliquées (INSA), Université Paul Sabatier (UPS), Institut National Polytechnique (INP), Laboratoire d'Ingénierie des Systèmes biologiques et des Procédés (LISBP), Toulouse, France

² Centre National de la Recherche Scientifique, UMR5504, Toulouse, France

³ Institut National de la Recherche Agronomique, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France

⁴ Institut National de la Recherche Agronomique, UMR1077 Mathématique, Informatique et Génome, Jouy-en-Josas, France

⁵ Laboratoire de Biologie Appliquée à l'Agroalimentaire et à l'Environnement, Institut Universitaire de Technologie – Université Paul Sabatier, Auch, France

*These authors contributed equally to this work

FEBS Journal (2015)

Inventory of the GH70 enzymes encoded by *Leuconostoc citreum* NRRL B-1299 – identification of three novel α -transglucosylases

Delphine Passerini^{1,2,3,*}, Marlène Vuillemin^{1,2,3,*}, Lisa Ufarté^{1,2,3}, Sandrine Morel^{1,2,3}, Valentin Loux⁴, Catherine Fontagné-Faucher⁵, Pierre Monsan^{1,2,3}, Magali Remaud-Siméon^{1,2,3} and Claire Moulis^{1,2,3}

1 Université de Toulouse, Institut National des Sciences Appliquées (INSA), Université Paul Sabatier (UPS), Institut National Polytechnique (INP), Laboratoire d'Ingénierie des Systèmes biologiques et des Procédés (LISBP), Toulouse, France

2 Centre National de la Recherche Scientifique, UMR5504, Toulouse, France

3 Institut National de la Recherche Agronomique, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France

4 Institut National de la Recherche Agronomique, UMR1077 Mathématique, Informatique et Génome, Jouy-en-Josas, France

5 Laboratoire de Biologie Appliquée à l'Agroalimentaire et à l'Environnement, Institut Universitaire de Technologie – Université Paul Sabatier, Auch, France

Keywords

α -transglucosylase; branching sucrose;
dextranucrase; glucansucrase;
Leuconostoc citreum NRRL B-1299

Correspondence

M. Remaud-Simeon and C. Moulis
Université de Toulouse, Institut National des
Sciences Appliquées (INSA), Université
Paul Sabatier (UPS), Institut National
Polytechnique (INP), Laboratoire d'Ingénierie
des Systèmes biologiques et des Procédés
(LISBP), 135 Avenue de Rangueil 31077,
Toulouse cedex 4, France
Fax: +33 561 559 400
Tel: +33 561 559 446
E-mail: remaud@insa-toulouse.fr (M.
Remaud-Simeon)
E-mail: moulis@insa-toulouse.fr (C. Moulis)

*These authors contributed equally to this
work

(Received 20 November 2014, revised 4
March 2015, accepted 6 March 2015)

doi:10.1111/febs.13261

Leuconostoc citreum NRRL B-1299 has long been known to produce α -glucans containing both α -(1 \rightarrow 6) and α -(1 \rightarrow 2) linkages, which are synthesized by α -transglucosylases of the GH70 family. We sequenced the genome of *Leuconostoc citreum* NRRL B-1299 to identify the full inventory of GH70 enzymes in this strain. Three new genes (*brsA*, *dsm* and *dspd*) putatively encoding GH70 enzymes were identified. The corresponding recombinant enzymes were characterized. Branching sucrose A (BRS-A) grafts linear α -(1 \rightarrow 6) dextran with α -(1 \rightarrow 2)-linked glucosyl units, and is probably involved in the α -(1 \rightarrow 2) branching of *L. citreum* NRRL B-1299 dextran. This is the first report of a naturally occurring α -(1 \rightarrow 2) branching sucrose. DSR-M and DSR-DP are dextranucrases that are specific for α -(1 \rightarrow 6) linkage synthesis and mainly produce oligomers or short dextrans with molar masses between 580 and 27 000 g·mol⁻¹. In addition, DSR-DP contains sequences that diverge from the consensus sequences that are typically present in enzymes that synthesize linear dextran. Comparison of the genome with five other *L. citreum* genomes further revealed that *dspd* is unique to *L. citreum* NRRL B-1299. The presence of this gene in a prophage represents the first evidence of phage-mediated horizontal transfer of genes encoding such enzymes in lactic acid bacteria. Finally, *brsA* and *dsm* are located in a chromosomal region in which genes encoding strain-specific GH70 enzymes are consistently located. This region may be a good target on which to focus in order to rapidly evaluate the diversity of GH70 enzymes in *L. citreum* strains.

Abbreviations

BRS, branching sucrose; CD, catalytic domain; CDS, coding sequence; DSR, dextranucrase; GBD, glucan binding domain; GH, glycoside hydrolase; GS, glucansucrase; HPSEC, high-performance size-exclusion chromatography; IS, insertion sequence.

Introduction

Due to their renewability, biodegradability and carbon neutrality, bio-sourced polymers represent attractive alternatives to polymers derived from carbon fossil fuels, with a broad range of applications in food and feed products (e.g. prebiotics and additives), agriculture (e.g. fertilizers), healthcare products (e.g. vaccines, antibiotics, antivirals and biomaterials) and chemical industries (e.g. bioplastics and biosurfactants) [1–3]. In this context, the α -glucans of high molar mass produced by lactic acid bacteria are receiving increasing attention. Such polymers, including dextran, reuteran, mutan and alternan, are all synthesized from sucrose by α -transglucosylases, also named glucansucrases (GSs) [4]. These enzymes belong to family 70 of the glycoside hydrolases (GH70). They are structurally and mechanistically related to the GH13 and GH77 families, with which they form the clan GH-H of the Glycoside Hydrolases and share the same α -retaining mechanism. Cleavage of the osidic bond occurs through formation of a β -D-glucosyl enzyme intermediate, and involves two catalytic residues, namely a nucleophile and an acid/base proton donor, and then a third invariant aspartic acid that plays a role in stabilizing the substrate transition state [5,6]. Depending on the enzyme specificity, polymers or oligomers of various sizes with various types and arrangements of α -osidic linkages may be formed [5,7]. To date, 59 GH70 enzymes isolated from the *Lactobacillus*, *Leuconostoc*, *Streptococcus* and *Weissella* genera have been biochemically characterized, and four crystallographic structures are available. These enzymes adopt a unique U-shaped fold organized into five distinct structural domains A, B, C, IV and V. All the domains, except domain C, are formed by non-contiguous sequence fragments [8–11]. In addition, domains A, B and C are closely related to their counterparts in the GH13 family [8]. The $(\beta/\alpha)_8$ barrel of domain A shows a circular permutation compared with that of GH13 enzymes [12]. In particular, the four signature regions of the GH13 catalytic barrel, referred to as motifs I, II, III and IV, are conserved in the GH70 family but are found in a different order along the sequence, with motifs II, III and IV preceding motif I [12,13]. Domains IV and V are specific to GH70 enzymes. Domain V is rich in repeated motifs known as the YG repeats, which have been proposed to be involved in glucan binding and enzyme processivity [14,15].

The strain *Leuconostoc citreum* NRRL B-1299 (also named ATCC 11449 or NCDO 1875) was first described in the pioneering work by Jeanes *et al.* [16], and was found to produce a dextran polymer containing up to 35% α -(1→2) branching linkages [17,18]. The secreted GSs of this strain have been industrially exploited for

production of α -(1→2)-branched oligosaccharides with prebiotic properties. In particular, oligosaccharide resistance to digestive enzymes has been shown to be mainly related to the presence of the α -(1→2) glucosidic linkage, which is rarely found in nature [19,20]. Initially identified as *Leuconostoc mesenteroides*, the strain was recently reclassified as *Leuconostoc citreum* on the basis of repetitive sequence-based PCR and fermentation profiles [21]. More than ten years ago, three GS-encoding genes were identified in this strain, namely *dsrA*, *dsrB* and *dsrE*. These genes were expressed in *Escherichia coli*. Using a sucrose substrate, recombinant dextransucrases DSR-A and DSR-B were shown to synthesize dextrans mainly comprising α -(1→6) osidic bonds with several α -(1→3) branching linkages [22,23]. In contrast, DSR-E dextransucrase is more atypical [24,25]. Indeed, this protein of 2835 amino acids is the only GH70 enzyme that possesses two active catalytic domains (named CD1 and CD2) connected by a glucan binding domain (GBD). Fabre *et al.* further showed that the CD1 domain acts as a polymerase and is specific for α -(1→6) linkage synthesis, whereas CD2 is dedicated to α -(1→2) linkage synthesis [25], as revealed by the specificity of two truncated forms of DSR-E, GBD-CD2 and ΔN_{123} -GBD-CD2, from which CD1 has been deleted [10,25]. Dextran synthesized by recombinant DSR-E was structurally characterized. The polymer was shown to contain only 5% α -(1→2) branching linkages, indicating that this enzyme cannot be solely responsible for synthesis of the highly α -(1→2) branched dextran produced by *L. citreum* NRRLB-1299 [25,26].

In the present study, the genome of *L. citreum* NRRL B-1299 was sequenced to identify the full inventory of GH70-encoding genes in this strain and to perform a comparative genome analysis with other available *L. citreum* genome sequences. This approach enabled the discovery and biochemical characterization of three new GH70 enzymes with unusual sequences and properties. The findings shed new light on the catalytic machinery of *L. citreum* NRRL B-1299, particularly on enzymes exhibiting α -(1→2) linkage specificity. The results also provided insights into the acquisition of GH70 enzyme-encoding genes, providing the first evidence of phage-mediated horizontal transfer.

Results

Leuconostoc citreum NRRL B-1299 genome features

The complete genome of *L. citreum* NRRL B-1299 was sequenced and assembled. The strain harbors one circular chromosome of 1 753 809 bp, with a GC content of

39.12%, and one plasmid of 3369 bp. According to the automatic annotation provided by the AGMIAL platform (Institut National de la Recherche Agronomique, Jouy-en-Josas, France), the genome contains 1786 predicted coding sequences (CDS) (1769 on the chromosome and 17 on the plasmid), three copies of ribosomal RNA genes and 53 tRNA-encoding genes. Fifteen putative pseudogenes representing less than 1% of the total predicted CDS were also identified. This result is in agreement with the values usually reported for most prokaryotes, which range from 1 to 5% [27]. Only two types of insertion sequence (IS) elements were found: one truncated IS from the IS30 family and seven intact IS from the IS3 family. Most of them are located at the replication terminus (Fig. 1). According to the ISfinder database [28] (<https://www-is.biotoul.fr/>), both the IS3 and IS30 families are also found in the genomes of *Lactobacillus* subsp. and *Lactococcus* subsp.

The sequences of the assembled genomes from the *L. citreum* NRRL B-1299, KM20 and NRRL B-742 strains, and those of the draft genomes from the *L. citreum* C10, C11 and E16 strains, were compared [29–31]. Genome alignment revealed a global synteny between all assembled *L. citreum* chromosomes (Fig.

S1). Only a few regions specific to *L. citreum* NRRL B-1299 were highlighted, representing less than 1% of the chromosome genetic content (93 kb) (Fig. 1). The most striking region, from 1 038 301 to 1 079 702 bp, corresponds to a prophage sequence that was not detected in the other genomes despite the high conservation observed at the species level.

Leuconostoc species are known to be hosts for phages [32–34], and phages from *L. mesenteroides* and *L. pseudomesenteroides* are of particular interest due to their negative effect on fermentation processes in the dairy industry [35]. However, little is known about *L. citreum* phages. In the NRRL B-1299 chromosome, the entire prophage of 42 kb (flanked by attachment sites *attL* at the 1 038 301 bp position and *attR* at the 1 079 702 bp position) was integrated at a tRNA (Ser) gene site (Fig. 1). Prophage annotation using the PHAST server [36] (<http://phast.wishartlab.com/>) revealed 40 predicted CDS, most of which display homologies with CDS of Gram-positive bacteria. Genes that are involved in the main steps of the phage lifecycle were identified. A treatment with mitomycin C confirmed that this phage was able to switch towards a lytic cycle (data not shown). To our knowledge, this

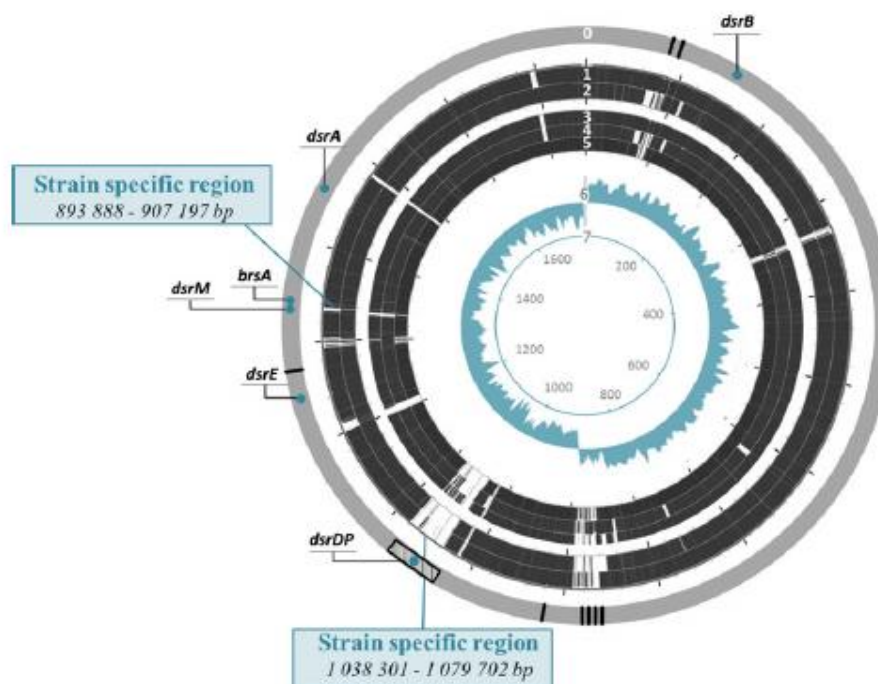


Fig. 1. Mapping of putative GH70 enzyme-encoding genes and mobile genetic elements on the *L. citreum* NRRL B-1299 chromosome, and comparison with available chromosomes of other *L. citreum* strains. Outer gray ring (0): location of GH70 enzyme-encoding genes (blue circles), insertion sequences (black lines) and prophage (hatched area) in the *L. citreum* NRRL B-1299 chromosome. Black rings (1–5): comparison of *L. citreum* NRRL B-1299 chromosome with the chromosomes of (1) KM20, (2) NRRL B-742, (3) E16, (4) C10 and (5) C11. The analysis was performed using the CGview server (http://stothard.afns.ualberta.ca/cgview_server/) with a BLASTn algorithm (cut-off 60% identity). Conserved regions are indicated in black. Regions specific to *L. citreum* NRRL B-1299 are shown in white. Regions containing *dsrM*, *bsrA* and *dsrDP* (within the prophage sequence) are indicated by blue rectangles 'Strain specific region'. The inner rings show GC skew (6) and scale marks in kbp (7).

prophage sequence has not been previously identified in its entire form in any other lactic acid bacteria.

Next, we investigated the presence of GH70-encoding genes. Six genes putatively encoding GH70 enzymes were detected in the genomic sequence. Among them, three genes (*dsrA*, *dsrB* and *dsrE*) have been previously functionally characterized [22–24], and three new genes were discovered: *dsrDP* (3837 bp) in the prophage sequence, and *brsA* (5634 bp) and *dsrM* (6198 bp).

The six genes are located on the reverse strand, and are distributed along the chromosome (Fig. 1). Only *brsA* and *dsrM* are arranged in tandem. A putative promoter was identified upstream of each gene. Except for *dsrDP*, sequence analysis further revealed that all of these genes are preceded by a putative ribosome binding site, localized at a maximum of four bases upstream of the initiation codon. Putative -10 (TATA-AT) and -35 (TTGACA) boxes, separated by a 17 bp spacer region, were also detected (data not shown), indicating that all of these genes could be expressed in *L. citreum* NRRL B-1299.

The sequences of *dsrA* and *dsrB* revealed some differences from the previously published sequences [22,23] (Fig. S2). As deep sequencing was performed (29 x depth), the observed discrepancies were attributed to mutations introduced by PCR reactions or sequencing errors that may have occurred in the previous studies. The sequence of *dsrE* is identical to the sequence previously reported [24].

DSR-DP, BRS-A and DSR-M enzyme sequence analyses

Based on their deduced amino acid sequences, DSR-DP, branching sucrose A (BRS-A) and DSR-M comprise 1278, 1877 and 2065 amino acids, and have calculated molecular masses of 145, 209 and 229 kDa, respectively. BRS-A and DSR-M may be extracellular enzymes, as revealed by the presence of a putative signal peptide. No signal peptide is predicted for DSR-DP, which is the case for approximately 19% (6/32) of the putative GH70 enzymes of *L. citreum* strains.

Based on alignment with the sequences of GTF-180 glucansucrase and ΔN_{123} -GBD-CD2 α -(1 \rightarrow 2) branching sucrose, for which 3D structures are available [8,10], DSR-DP, BRS-A and DSR-M are predicted to adopt an organization involving five distinct structural domains (A, B, C, IV and V), as previously described for the GH70 family enzymes (Fig. 2).

blastp analyses against the non-redundant protein sequence database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) revealed that DSR-DP shares the highest level of amino acid identity (44%) with a putative membrane-associated glycoside hydrolase of *L. mesenteroides* KFRI-MG (AHF19735.1). The four motifs (I, II, III and IV), considered as signature motifs in both the GH70 and GH13 families [13], were easily identified in the putative domain A of DSR-DP, and aligned with the corresponding sequences of GH70 enzymes of

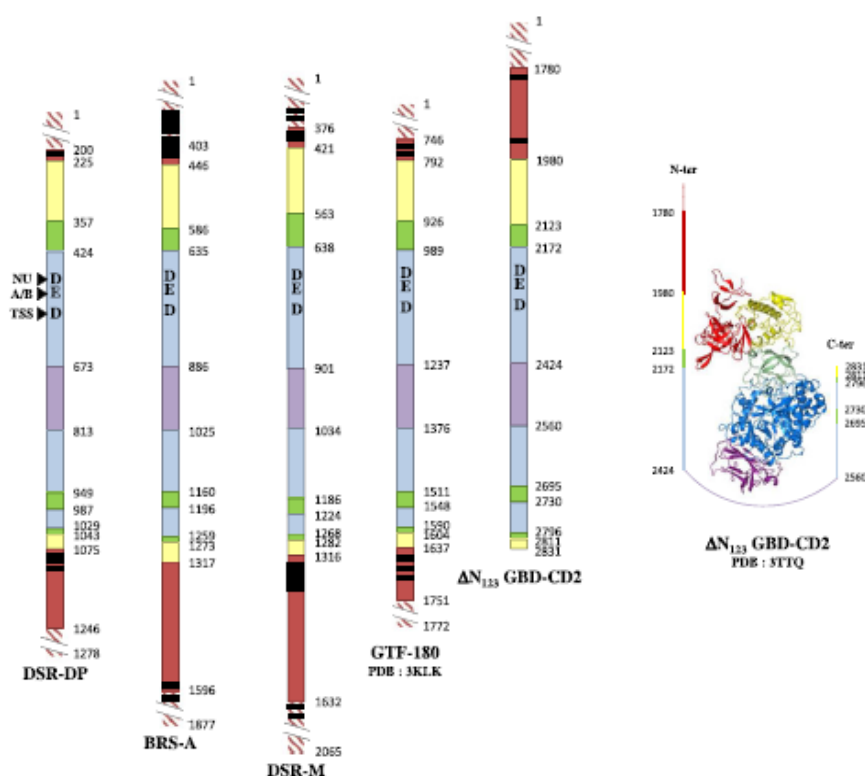


Fig. 2. Schematic representation of the structural organization of DSR-DP, BRS-A and DSR-M. The sequences were aligned with those of GTF-180 and ΔN_{123} -GBD-CD2 using ClustalW2. ΔN_{123} -GBD-CD2 is a truncated form of DSR-E, and its numbering refers to the DSR-E sequence [10]. Five structural domains were identified: (i) domain V, the glucan binding domain (red), (ii) domain IV (yellow), (iii) domain B (green), (iv) catalytic domain A (blue), and (v) domain C (purple). YG repeats are represented by black rectangles. Catalytically important residues [the nucleophile (NU), the acid/base catalyst (A/B) and the transition-state stabilizer (TSS)] are indicated using bold black letters. Numbers refer to amino acid numbers at the domain borders.

known specificity (Table 1). The singularity of this protein sequence was further highlighted by the presence of unique amino acids, such as the K492 and A500W501 of motif III and the A571 and F577S578N579 sequence of motif IV, which have not been reported previously in other characterized GH70 enzymes (Table 1). With regard to BRS-A, this enzyme shares 58% amino acid identity with the whole sequence of GBD-CD2, a truncated form of DSR-E [37]. This percentage increased up to 65% when comparing the A, B and C domains of the two proteins [8–11]. It is noteworthy that specific sequences are conserved near the catalytically important residues, including the GLDA (amino acids 712 to 715, BRS-A numbering) and KVGA (amino acids 790 to 793, BRS-A numbering) sequences downstream of the acid/base residue (motif III) and the transition-state stabilizer (motif IV), respectively (Table 1). Finally, DSR-M is highly similar to the recently identified GSE16-5 glucansucrase from *L. citreum* LBAE-E16 [38], and shows 94% sequence identity with this enzyme. Motifs I–IV of DSR-M are very similar to the motifs found in dextranases specific for α -(1→6) linkage synthesis, as revealed by the SEVQT sequence (amino acids 791 to 795, DSR-M numbering) of motif IV (Table 1).

In addition, as shown in Table 1, the residues at subsite -1, which interact with the glucosyl ring of sucrose in the inactive mutant GTF-180- Δ N in complex with sucrose (PDB ID 3HZ3), are all conserved in DSR-DP, BRS-A and DSR-M. With regard to subsite +1, all the residues interacting with the fructosyl ring are conserved, except for residues N1029 and W1065 (GTF-180 numbering), which are replaced by a phenylalanine and a glycine, respectively, in BRS-A [8]. These findings suggest that sucrose binding and cleavage may proceed through the same mechanism, particularly for DSR-DP and DSR-M. The two divergent residues in BRS-A, which are also encountered in GBD-CD2, may play a role in the specificity of this enzyme (Table 1) [10].

The C-terminal domain V of BRS-A and DSR-M, which has been implicated in glucan binding, is much longer than domain V of DSR-DP and GTF-180- Δ N. Ten, eleven and four YG repeats, which have been reported to play a key role in glucan binding, were identified in domain V of BRS-A, DSR-M and DSR-DP, respectively [15]. Interestingly, the YG repeats of BRS-A are mainly located at the N-terminus, similar to the engineered GBD-CD2 and Δ N₁₂₃-GBD-CD2

Table 1. Sequence alignment of functional conserved motifs I–IV of GH70 enzymes. The GenBank accession numbers of the glucansucrases from *Leuconostoc citreum* NRRL B-1299 correspond to their locus tag on the genome (PRJEB5537). The numbering order of the motifs (II, III, IV and I) refers to the motifs originally defined in the family GH13. The two catalytic residues are indicated in red for the nucleophile residue (motif II) and the acid/base catalyst (motif III), and the transition-state stabilizer residue is shown in blue. The residues in green and orange correspond to residues of subsites -1 and +1, respectively, as originally identified in GTF180- Δ N [8]. Grey shading represents enzymes encoded by the genome of *L. citreum* NRRL B-1299.

GenBank accession number	Enzymes	Motif II	Motif III	Motif IV	Motif I	Specificity
AAC63063.1	GTF-I	449 SIRVDAVDNVD	486 HVSIIVEAWSDN	559 FARAHDSEVQDLIRD	931 ADWVPDQ	α -(1→3)
BAA02976.1	GTF-I	443 SIRVDAVDNVD	480 HVSIIVEAWSDN	553 FARAHDSEVQDIIRD	925 ADWVPDQ	
AAA88588.1	GTF-B	1011 SIRVDAVDNVD	1048 HLSILEAWSDN	1120 FIRAHSEVQDLIAD	1488 ADWVPDQ	
AAU08015.1	GTF-A	1020 SVRVDPDNDID	1056 HINILEDWNHA	1128 FVRAHDNNSQDQIQN	1508 ADWVPDQ	α -(1→4)/
AAY86923.1	GTF-O	1020 SVRVDPDNDID	1056 HINILEDWNSS	1128 FIRAHNNSQDQIQN	1508 ADWVPDQ	α -(1→6)
CAB65910.2	ASR	631 GIRVDAVDNVD	668 HLSILEDWNGK	762 FVRAHDYDAQDPIRK	1168 ADWVPDQ	α -(1→6)/ α -(1→3)
ABQ83597.1	GTF-W	748 GFRVDAADNID	785 HLVDYNEGYHSG	568 FVTNEDQR-KNVINQ	1216 EDLVMNQ	α -(4→6)
AAU08003.2	GTF-ML4	1012 GFRVDAADNID	1049 HLSYNEGYHSG	1121 FVTNEDQR-KNLINR	1479 EDIVMNQ	
ABF85832.1	DSR-CB4	526 GIRVDAVDNVD	563 HLSILEDWWSHN	636 FVRAHDSEVQTVIAQ	1001 ADWVPDQ	α -(1→6)
CAB76565.1	DSR-C	498 GIRVDAVDNVD	535 HLSILEDWWSHN	608 FVRAHDSEVQTVIAQ	973 ADWVPDQ	
AAD10952.1	DSR-S	547 GIRVDAVDNVD	584 HLSILEDWWSHN	657 FVRAHDSEVQTVIAQ	1023 ADWVPDQ	
AAU08001.1	GTF-180	1021 GIRVDAVDNVD	1058 HINILEDWGWD	1131 FVRAHDSNAQDQIRQ	1503 ADWVPDQ	
BN964_01466	DSR-A	361 GYRVDAVDNVD	399 HLSILEDWGDE	472 FIRAHSEVQTIIAD	844 NDWVPDQ	α -(1→6)
BN964_00137	DSR-B	401 GIRVDAVDNVD	449 HLSILEDWWSHN	512 FVRAHDSEVQTVIAQ	877 ADWVPDQ	α -(1→6)
BN964_01272	DSR-E (CD1)	520 GYRVDAVDNVD	557 HSIILEDWDNN	630 FIRAHSEVQTVIAQ	1010 NDWVPDQ	α -(1→6)
BN964_01272	DSR-E (CD2)	2206 SIRIDAVDFIH	2243 HILSLVEAQLDA	2317 IIAHAEKGVQEKVGA	2688 ADVVDNQ	α -(1→2)
BN964_01092	DSR-DP	455 GIRVDAVDNVD	492 KISILEDWAWG	565 FIRAHDAESQDIFSN	940 ADYVPDQ	α -(1→6)
BN964_01348	BRS-A	668 SIRIDAVDFVS	705 HLSLVEAQLDA	779 IIAHAEKDIQDKVGA	1151 ADVVANQ	α -(1→2)
BN964_01347	DSR-M	673 SIRIDAVDNVD	710 HIHILEDWSPN	785 FIRAHSEVQTIIAK	1177 ADFVPDQ	α -(1→6)

branching sucrases [25]. In addition, DSR-M and BRS-A display highly conserved fragments of sequences in domain V; the first 151 N-terminal amino acids and the last 515 C-terminal amino acids share more than 96% identity. Importantly, both C-terminal sequences contain APY repeats with as yet unknown functions, and exhibit more than 98% identity with the C-terminal region of *L. citreum* NRRL B-1355 alternansucrase and *L. citreum* CW28 inulosucrase [39,40]. Like GSE16-5 glucansucrase, DSR-M exhibits a dextransucrase-like catalytic core and an alternansucrase-like C-terminal region [38].

Phylogenetic analyses were performed on the characterized GH70 enzymes indexed in the CAZy database [4] and the three new enzymes DSR-DP, BRS-A and DSR-M. BRS-A clusters together with the α -(1 \rightarrow 2) branching sucrose GBD-CD2, whereas DSR-DP and DSR-M are located in separate branches (Fig. 3).

Biochemical characterization of DSR-DP, BRS-A and DSR-M

The genes encoding these enzymes were cloned in-frame with N- and C-terminal tags and expressed in various *E. coli* strains to find the best system for expression of each enzyme (Table S1). The highest enzyme production yields were 750, 5000 and 2880 U per liter of culture for DSR-DP, BRS-A and DSR-M, respectively. Western blot analyses showed that DSR-DP and BRS-A were produced with their N- and C-terminal tags, whereas DSR-M was slightly degraded at both the N- and C-termini, probably because of its larger molar mass (229 000 g \cdot mol $^{-1}$) (data not shown). The specific activities of the DSR-DP, BRS-A and DSR-M proteins were 17, 12 and 3 U \cdot mg $^{-1}$, respectively. These values are lower than that reported for DSR-S vardele Δ 4N from *L. mesenteroides* NRRL 512F, whose specific activity is approximately 400 U \cdot mg $^{-1}$, but of the same order of magnitude as the specific activity of Δ N $_{123}$ -GBD-CD2, the α -(1 \rightarrow 2) branching sucrose from *L. citreum* NRRLB-1299 [10,41].

DSR-DP

High-performance size-exclusion chromatography (HPSEC) of the products synthesized by DSR-DP after total sucrose depletion showed that this enzyme synthesizes a highly polydisperse polymer. Indeed, five populations, with weight-average molar masses estimated at (i) 580 g \cdot mol $^{-1}$, (ii) 2000 g \cdot mol $^{-1}$, (iii) 4240 g \cdot mol $^{-1}$, (iv) 9300 g \cdot mol $^{-1}$ and (v) greater than 2×10^6 g \cdot mol $^{-1}$, are distinguishable on the HPSEC profile (Fig. 4). Eighty-nine percent of the glucosyl residues from sucrose were

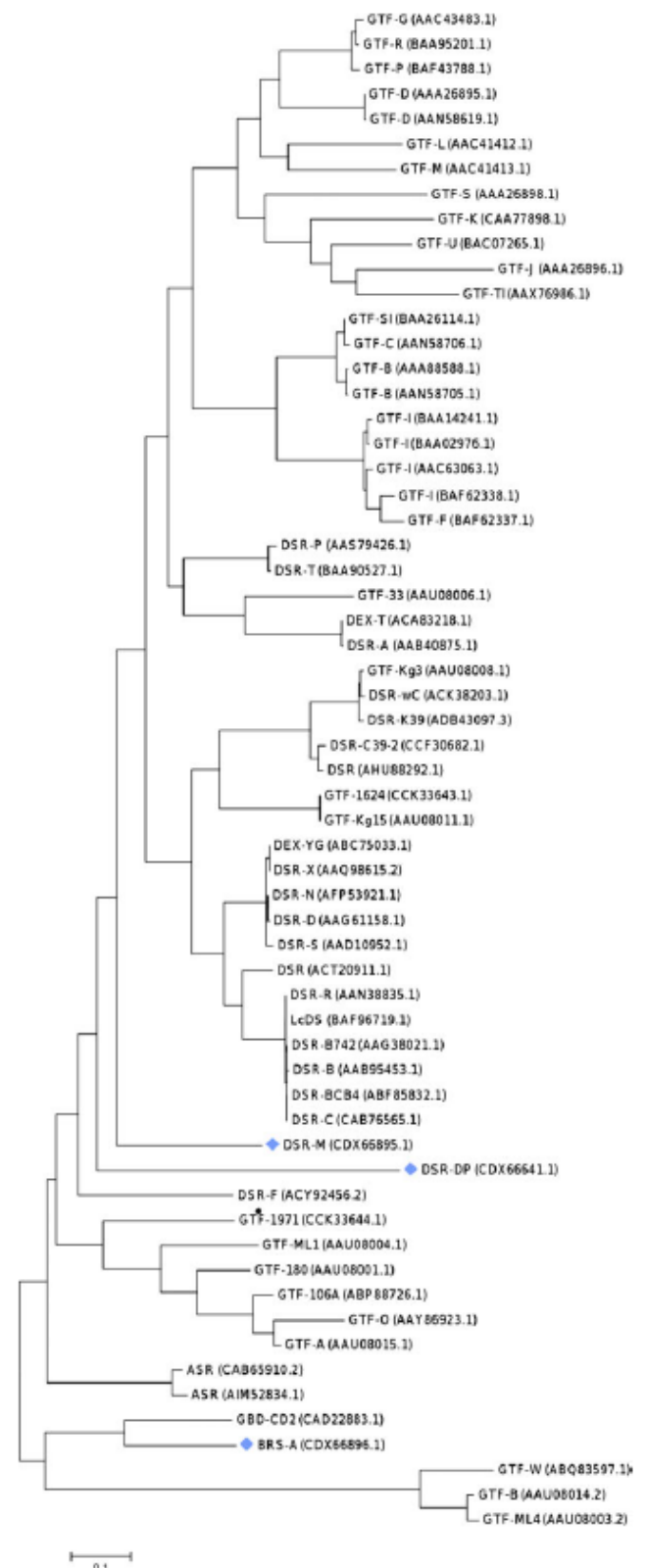


Fig. 3. Phylogenetic tree of the GH70 enzymes, built using the maximum-likelihood method based on the JTT matrix-based model using MEGA6 [59]. The tree with the highest log likelihood is shown. The tree is drawn to scale, with branch lengths measured as the number of substitutions per site. The analysis involved 61 amino acid sequences. Each protein is labeled with its GenBank accession number. BRS-A, DSR-DP and DSR-M are indicated by blue diamonds.

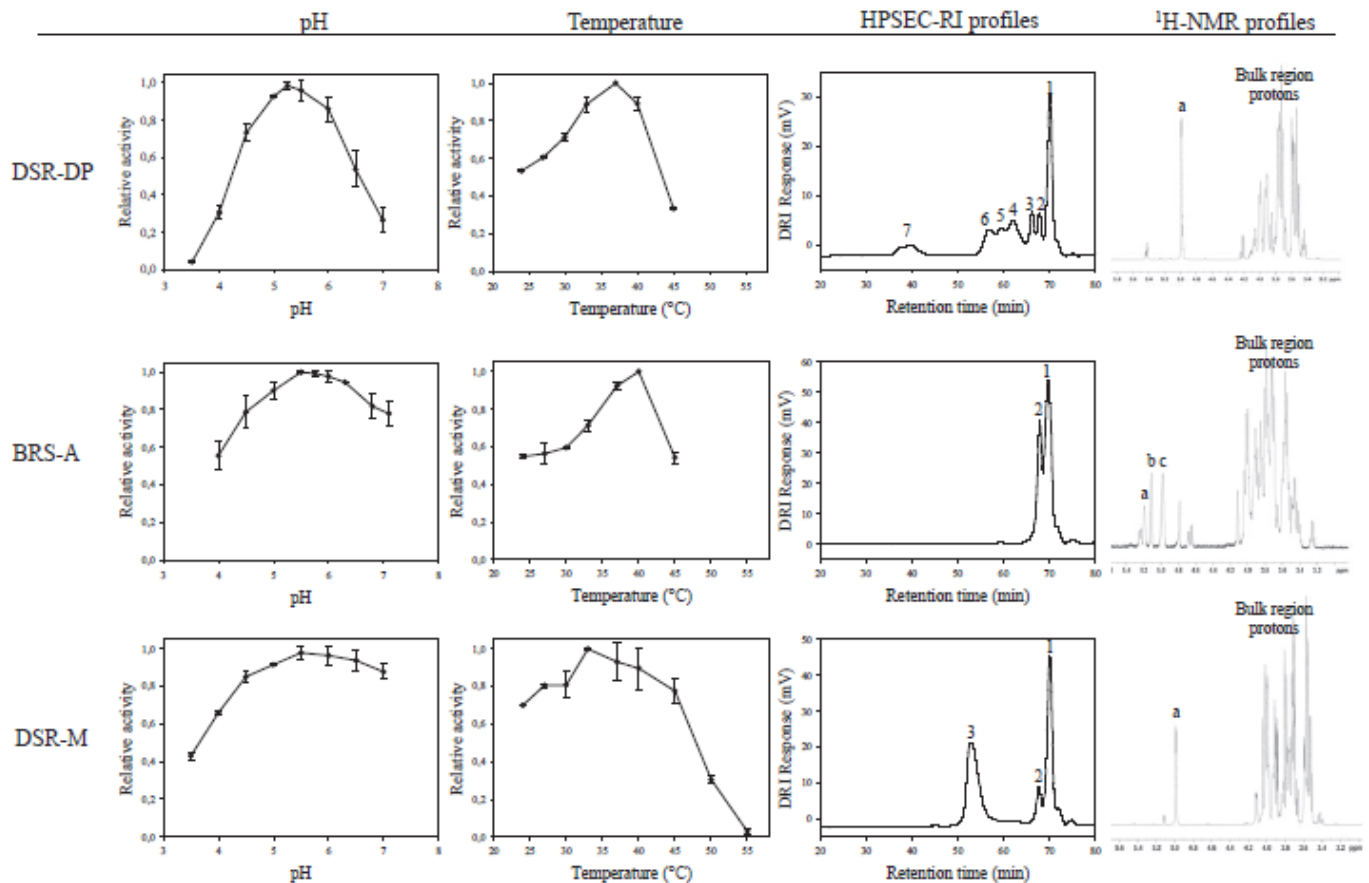


Fig. 4. Biochemical properties of DSR-DP, BRS-A and DSR-M. Columns 1 and 2: pH and temperature profiles. Values are means \pm standard deviation calculated from two independent experiments. Column 3: HPSEC profiles of the products synthesized from sucrose at a concentration of 292 mM at 30 °C. Products are indexed according to their weight-average molar mass. For DSR-DP: (1) 180 g·mol⁻¹, corresponding to a mixture of fructose and glucose (degree of polymerization of 1); (2) 342 g·mol⁻¹, corresponding to leucrose (degree of polymerization of 2); (3) 580 g·mol⁻¹; (4) 2000 g·mol⁻¹; (5) 4240 g·mol⁻¹; (6) 9300 g·mol⁻¹; (7) more than 2×10^6 g·mol⁻¹. For BRS-A: (1) 180 g·mol⁻¹, corresponding to a mixture of glucose and fructose; (2) 342 g·mol⁻¹, corresponding to leucrose. For DSR-M: (3) 27 000 g·mol⁻¹ dextran. Column 4: ¹H-NMR spectra of the products. Peak attribution is described in Fig. S3. For DSR-DP, the spectrum of dextrans is shown from 580–9300 g·mol⁻¹ (peaks 3–6 of the HPSEC profile) with 'a' representing the anomeric proton of the α -(1 \rightarrow 6)-linked Glcp unit. For BRS-A, the spectrum of products synthesized from sucrose at a concentration of 439 mM and 1500 g·mol⁻¹ dextran at a concentration of 33 mM is shown, with 'a' indicating the signal of the anomeric proton of a 2,6-di-*O*-substituted Glcp unit, 'b' indicating the anomeric protons of the α -(1 \rightarrow 2)-linked Glcp, and 'c' indicating the anomeric proton of the α -(1 \rightarrow 6)-linked Glcp unit. For DSR-M, the spectrum of dextran synthesized from sucrose sucrose at a concentration of 292 mM is shown, with 'a' indicating the anomeric proton of the α -(1 \rightarrow 6)-linked Glcp unit.

incorporated into these glucans (77% in low-molar-mass glucans, ranging from 580–9300 g·mol⁻¹, and 12% in high-molar-mass glucan). Only 3% and 8% of the glucosyl units were transferred onto water molecules and fructose to produce glucose and leucrose (5-*O*- α -D-glucopyranosyl-D-Fructose), respectively. Proton NMR analyses of the low- and high-molar-mass polymers showed that they mainly comprised α -(1 \rightarrow 6) glucosidic linkages (Fig. 4 and Fig. S3a,b).

BRS-A

With sucrose as the sole substrate, 56%, 31% and 13% of the glucosyl residues were incorporated into

glucose, leucrose and other oligosaccharides, respectively. No polymer was formed (Fig. 4). An acceptor reaction was then performed using sucrose (146 mM) and 1500 g·mol⁻¹ dextran (33 mM), a linear α -(1 \rightarrow 6) glucan with a mean degree of polymerization of 9. Analyses of the reaction products showed that the linear oligosaccharides were used as acceptors, leading to formation of new compounds clearly observable on HPAEC-PAD chromatograms (Fig. 5). A ¹H-NMR spectrum of the reaction mix at the end of the reaction showed that the new products contained 34% α -(1 \rightarrow 2) linkages (Fig. 4 and Fig. S3c). The chromatographic profiles and ¹H-NMR spectra were identical to those obtained with the engineered α -(1 \rightarrow 2) branching

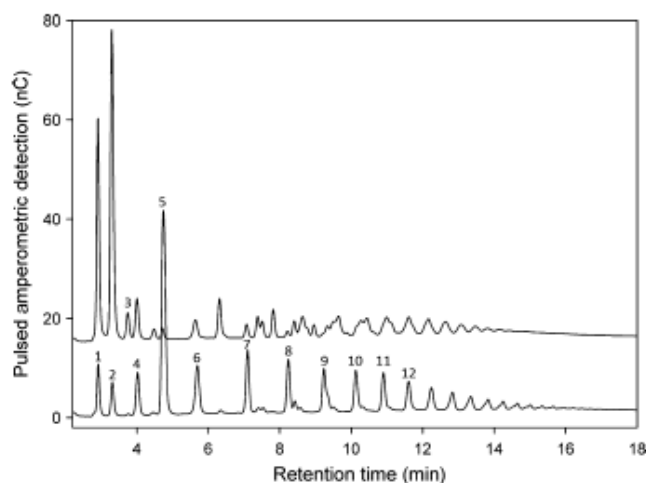


Fig. 5. HPAEC-PAD chromatograms of the acceptor reaction using sucrose (146 mM) and $1500 \text{ g}\cdot\text{mol}^{-1}$ dextran (33 mM) performed by BRS-A at 30°C and pH 5.75. The bottom line represents the reaction at the initial time (1 min reaction time). The top line represents the reaction after 8 h, indicating α -(1 \rightarrow 2)-branched products formed from the dextran acceptor, as well as fructose and leucrose by-products: (1) glucose, (2) fructose, (3) leucrose, (4) isomaltose, (5) sucrose, and (6–12) isomalto-oligosaccharides with degrees of polymerization of 3–9, respectively.

enzyme, ΔN_{123} -GBD-CD2, under similar reaction conditions [10]. Several dextran acceptors, varying in terms of their weight-average molar mass (68.4×10^3 , 503×10^3 and $2 \times 10^6 \text{ g}\cdot\text{mol}^{-1}$) were tested at a sucrose/acceptor ratio of 3. All the dextrans were branched by BRS-A and showed 37% α -(1 \rightarrow 2) linkages.

DSR-M

The enzyme incorporated 81% of the glucosyl units of sucrose into a dextran polymer, 15% into leucrose and 4% as free glucose. According to the $^1\text{H-NMR}$ spectrum and HPSEC profile, the dextran polymer exclusively comprised α -(1 \rightarrow 6) linkages and had a low weight-average molar mass estimated at $27\,000 \text{ g}\cdot\text{mol}^{-1}$ (Fig. 4).

The effects of pH and temperature on the enzyme activities are shown in Fig. 4. DSR-M and BRS-A showed a broad pH and temperature tolerance. The optimum temperatures of DSR-DP, BRS-A and DSR-M at pH 5.75 were 37, 40 and 33°C , respectively.

Comparison of the GH70 enzyme diversity in *Leuconostoc citreum* genomes

The diversity of genes putatively encoding GH70 enzymes of various *L. citreum* strains is shown in Fig. 6. Genes homologous to *dsrA*, *dsrB* and *dsrE*

were found at the same chromosomal location in other *L. citreum* genomes. Some of these genes contain from two to five stop codons and probably encode truncated or inactive proteins. The genetic environment of *dsrA*-like, *dsrB*-like and *dsrE*-like genes is conserved in all strains (Fig. 6 and Fig. S4), and include genes involved in metabolism and carbohydrate transport, energy metabolism, cell envelope biogenesis or regulatory functions. A gene homologous to *brsA* was identified in the C10 strain (contigs 31/40), and genes homologous to *dsmM* were found in the C10 and E16 strains (contigs 36 and 40 in the C10 strain, and GSE16-5 in the E16 strain). *L. citreum* is clearly a species enriched in GH70 α -transglucosylases, particularly when compared to other *Leuconostoc* subsp. strains. Indeed, according to BLAST analyses, the genome of *Leuconostoc kimchii* C2 contains no GH70 CDS, and the number of putative GH70-encoding genes found in the *L. mesenteroides* ATCC 8293, *L. mesenteroides* J18 and *L. carnosum* JB16 genomes is limited to three [42–45].

In addition, the *brsA* and *dsmM* tandem is located from 893 888 to 907 197 between two genes, the *hisC* gene and a gene coding for extracellular protein. These two genes are conserved in all other sequenced *L. citreum* genomes, and delineate a region that varies in size from 6 to 13 kb depending on the strain, which always contain one or two genes encoding GH70 enzymes with various linkage specificities (such as alternansucrase or dextransucrase instance). Finally, *dsmDP* is the sole GH70-encoding gene that is found only in the NRRL B-1299 strain.

Discussion

Since the 1950s, *L. citreum* NRRL B-1299 has received particular attention due to its ability to synthesize an unusual dextran containing up to 35% α -(1 \rightarrow 2) linkages [26,46–49]. Here we provide the assembled genome for this strain. Furthermore, we identified three new genes, *brsA*, *dsmM* and *dsmDP*, which encode three unusual α -transglucosylases from the GH70 family.

One of the most striking results was identification of the first natural α -(1 \rightarrow 2) branching sucrose (named BRS-A, for branching sucrose A). This enzyme belongs to the GH70 family, is not a polymerase, and catalyzes the transfer of glucosyl residues from sucrose to α -(1 \rightarrow 6) dextran specifically through formation of α -(1 \rightarrow 2) glucosidic linkages. To date, only the second catalytic domain (CD2) of DSR-E has been shown to display this level of specificity [24]. The sequence identity between the CD2 of DSR-E and BRS-A in the region downstream of the acid/base catalyst (Table 1)

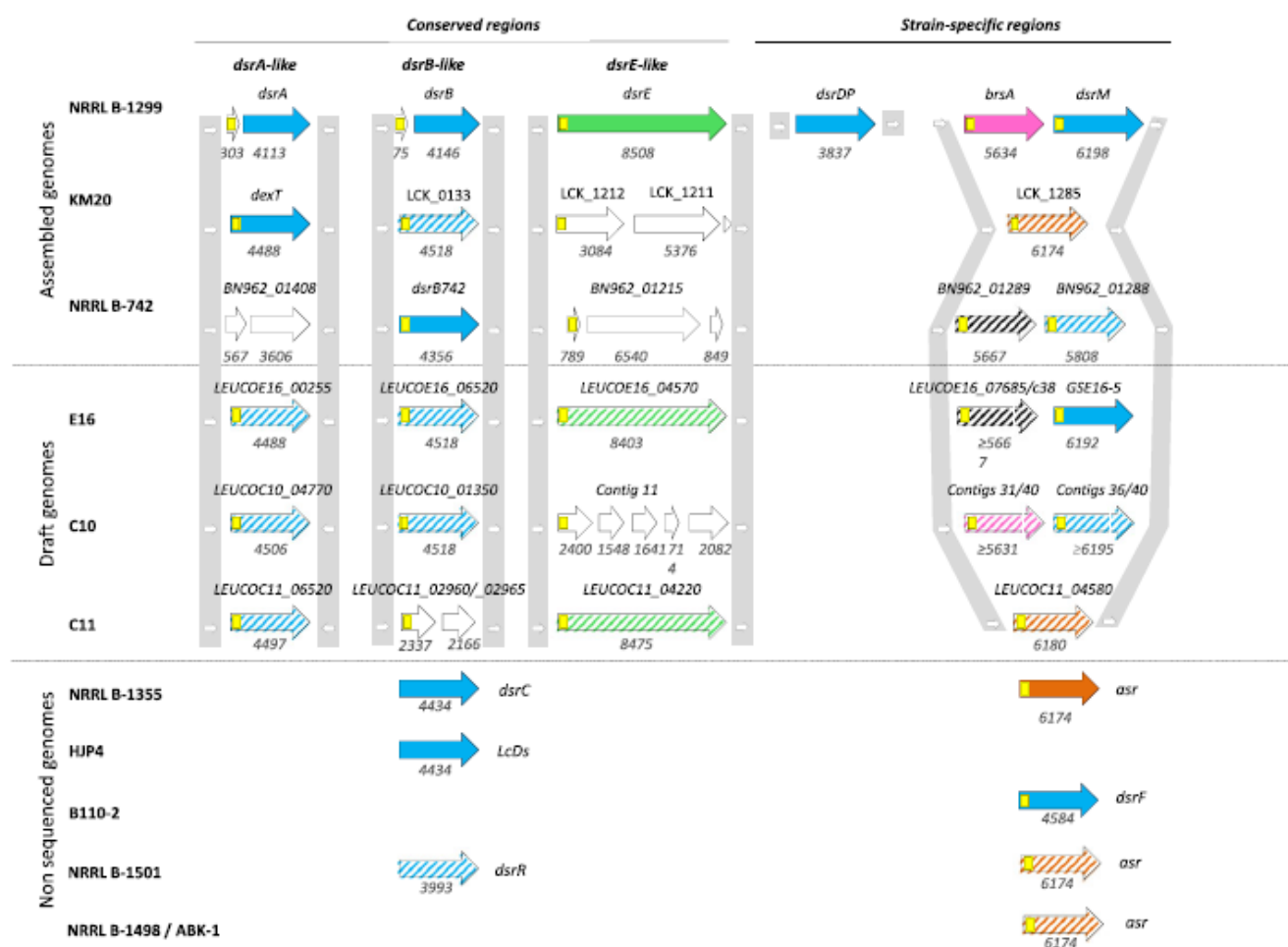


Fig. 6. Location and diversity of genes encoding putative GH70 enzymes in *Leuconostoc citreum*. Fully colored arrows, genes encoding characterized GH70 enzymes; hatched colored arrows, genes encoding putative GH70 enzymes (uncharacterized); white arrows, disrupted genes. Based on the sequence conservation, putative linkage specificities are represented by a specific color: blue, α -(1→6); green, α -(1→6)/ α -(1→2); pink, α -(1→2); orange, α -(1→6)/ α -(1→3); black, unknown. Yellow squares represent signal peptide sequences. The gene size (bp) is indicated below each CDS. The genome or gene accession numbers are NC_010471.1 (*L. citreum* KM20), CCNG00000000 (NRRL B-742), NZ_CAGG00000000.1 (E16), NZ_CAGE00000000.1 (C10), AJ250172.1 (*dsrC* from NRRL B-1355), AJ250173.2 (*asr* from NRRL B-1355), AB362781.1 (*LcDs* from HJP4), FJ844434.2 (*dsrF* from B-110-2), AY142210.1 (*dsrR* from NRRL B-1501), KF360258.1 (*asr* from NRRL B-1501) and KF360257.1 (*asr* from NRRL B-1498).

suggested that these two enzymes may share the same specificity, and this hypothesis was confirmed by biochemical data. The sequences GLDA and KxxxxKVGA, found in conserved motifs III and IV, respectively, of domain A may represent signatures for α -(1→2) branching linkage specificity. Due to the presence of a signal peptide in the protein sequence, BRS-A is predicted to be secreted by *L. citreum* NRRL B-1299. Hence, we assume that this enzyme is involved in the synthesis of the α -(1→2) glucosidic branches found in the native dextran produced by this strain. This assumption is supported by the fact that recombinant DSR-E produces a dextran with only 5% α -(1→2) branched linkages, and cannot account alone for the large number of branches in the native dextran of

L. citreum NRRL B-1299 (35%) [25]. This hypothesis is also supported by comparison between the GH70 enzyme content of the *L. citreum* NRRL B-1299 and C10 strains. The C10 strain was isolated from sourdough, and produces a dextran with 21% α -(1→2) linkages [21]. Like *L. citreum* NRRL B-1299, the *L. citreum* C10 genome carries genes homologous to *dsrE* and *brsA* (Fig. 6). However, the *dsrE*-like gene of the C10 strain contains several stop codons and is probably a pseudogene, which is not the case for the *brsA*-like gene, which shows 99% identity with *brsA*. Consequently, in strain C10, α -(1→2) linkage synthesis is probably catalyzed solely by the BRS-A-like enzyme. However, the possibility cannot be excluded that some degraded forms from DSR-E may also be

involved in the α -(1→2) branching of *L. citreum* NRRLB-1299 dextran. As most of the GH70 enzymes produced by *L. citreum* NRRLB-1299 (80%) are tightly bound to the cell wall, thereby preventing their isolation and purification, gene deletion combined with phenotypic characterization is required to provide more definite conclusions.

The second novel enzyme, DSR-M, is similar to the GSE16-5 glucansucrase (94% identity) that was recently identified in *L. citreum* LBAE-E16 [38]. DSR-M displays an α -(1→6) linkage specificity that may be predicted from the sequence of conserved motifs I–IV of domain A, which are highly similar to those found in DSR-S, the dextransucrase of *L. mesenteroides* NRRL B-512F. Compared to DSR-S, the most unusual characteristic of this enzyme is its ability to synthesize a linear dextran of low molar mass (27 000 g·mol⁻¹). This may be attributed to a reduced processivity, although DSR-M exhibits a large number of YG repeats, which were previously proposed to play a key role in the enzyme polymerization process [50,51]. Investigation of the relationship between structure and function in DSR-M will undoubtedly improve understanding of the polymerization mechanism of GH70 GSs.

Finally, DSR-DP is probably the most unusual GS reported in this study with regard to its primary structure (particularly the sequence of the conserved motifs), its product pattern, and its location on the chromosome. DSR-DP is mainly specific for α -(1→6) linkage synthesis, although the sequence downstream from the second aspartic acid of the catalytic triad (in motif IV) is highly divergent from the consensus SEVQT that is generally recognized as a signature of α -(1→6) linkage specificity [52]. The specificity of DSR-DP could not have been anticipated based on sequence alignment. In addition, DSR-DP mainly produces heterogeneous dextrans of low molar mass (580–9300 g·mol⁻¹). Finally, this enzyme exhibits a low number of YG repeats, which may explain the lack of processivity, but further studies are required to confirm this. The *dsrDP* gene was only identified in the *L. citreum* NRRL B-1299 genome, and, to date, no orthologous version has been found either in this species or in any other genomes of lactic acid bacteria. As it is located in a prophage that was shown to switch to a lytic cycle after induction, we suggest that this gene was acquired by horizontal transfer involving a bacteriophage, and probably came from other lactic acid bacteria. Indeed, many CDS of this bacteriophage are similar to those of *Lactobacillus* temperate bacteriophages (data not shown). Our results provide the first evidence of α -transglucosylase gene acquisition by transduction.

Horizontal gene transfer mediated by transposons has been suggested previously for GS-encoding genes of several *Streptococcus* and *Lactobacillus* strains [53–55]. Of the *L. citreum* strains analyzed here, only the truncated genes LCK_1211/LCK_1212 of strain KM20 (putatively encoding GS) were close to insertion sequences, showing that acquisition of GH70-encoding genes by horizontal gene transfer is not frequently observed in these strains.

The comparison of six genomes from various *L. citreum* strains further demonstrated that this species is enriched in various versions of GH70 enzyme-encoding genes. Most of these genes encode signal peptides, indicating that GH70 enzymes may be secreted and may play a role in strain adhesion, colonization or protection against diverse environmental stresses. All the strains possess *dsrA*-like, *dsrB*-like and *dsrE*-like genes. In each strain, at least one of them is predicted to be functional. In addition, comparative analysis also revealed the presence of a 6–13 kb genomic platform that is always located at the same position in all *L. citreum* genomes and contains α -transglucosylase-encoding genes such as *brsA*, *dsrM* or *asr* (coding for an alternansucrase) (Fig. 6). These three genes share 51% nucleotide identity throughout the entire sequence, and the identity values reach 98% when considering only the last 1550 bp. These genes share a common ancestor, and probably arose from gene duplication followed by speciation. These evolutionary processes may have occurred at the onset of GS diversification, as proposed for the glucosyltransferases of *Streptococcus mutans* [8,54]. Moreover, this genomic region may be considered as an indicator of *L. citreum* enzyme diversity, and may be used to rapidly screen new variants in *L. citreum* strains without having to sequence the entire genome.

In conclusion, our work has enabled the discovery and characterization of three new and unusual GH70 enzymes of prime interest, and provided new insights into the structure–function relationships within this protein family. This research will also facilitate the development of new processes for the production of α -glucan, gluco-oligosaccharides or glucoconjugates that vary in terms of structure, size, organization and the type of osidic linkages. Future investigations are planned with the BRS-A and DSR-M enzymes, which tolerate a broad range of pH values and temperatures and may become the enzymes of choice for controlled synthesis of low-molar-mass α -glucan from sucrose. This study also sheds new light on the evolutionary model of GH70 α -transglucosylases. In particular, we present arguments in favor of gene duplications and horizontal transfer mediated by a bacteriophage in *Leuconostoc citreum*. An in-depth phylogenetic analysis

is required to further investigate past events, explain the dissemination of GH70 sequences in the species, and obtain new insights into the functional role of these enzymes.

Experimental procedures

Isolation of genomic DNA

L. citreum NRRL B-1299 (isolated by Jeanes *et al.* in 1954 [16]) was grown in 1.5 mL MRS medium (Biokar Diagnostics, Pantin, France) at 30 °C overnight. Genomic DNA was isolated using a Wizard[®] genomic DNA purification kit (Promega, Madison, WI, USA), according to the manufacturer's instructions for Gram-positive bacteria.

NRRL B-1299 genome sequencing, assembly and annotation

The whole genome of strain NRRL B-1299 was obtained by combining shotgun and 8 kb paired-end library sequencing using the Roche 454 GS FLX Titanium pyrosequencing system (performed by MWG Eurofins, Ebersberg, Germany). Automated genome assembly was performed using Newbler 2.6 (Roche Diagnostics, Basel, Switzerland). The genome consisted of 49 contigs (29 contigs > 1 kb) distributed on two scaffolds of 1 757 178 bp and 3369 bp. The calculated mean depth of the shotgun sequencing was more than 29 x. The two scaffolds were used for a PCR-based approach for gap closure. Ten primer pairs (Table S2) were designed to close the gaps by sequencing the PCR products (GATC Biotech, Constance, Germany). The complete sequence was submitted to the AGMIAL annotation platform (Institut National de la Recherche Agronomique, Jouy-en-Josas, France) [56]. The automatic annotation provided was manually checked for regions containing GS-encoding genes. Use of web servers, such as PFAST [36] (<http://phast.wishartlab.com/>) for phages or ISfinder [28] (<https://www-is.biotoul.fr/>) for insertion sequences, enabled improvement of the automatic annotation.

In silico analysis

DNA sequence comparison of NRRL B-1299 and the C10, C11 and E16 draft genomes or the KM20 and NRRL B-742 assembled genomes was performed using a BLASTn algorithm on the CGView server [57]. BLAST hits were considered positive when they showed over 60% DNA similarity. Multiple chromosome alignments between NRRL B-1299, NRRL B-742 and KM20 genomic sequences were determined using the progressiveMauve algorithm of the Mauve software [58] using the default options. Mauve produces an output file containing orthologous genes (70%

DNA identity with 80% overlapping length). This file was used to calculate the total size of specific genes in the NRRL B-1299 strain. Comparison of *L. citreum* chromosomes was performed using the CGview server (http://stohard.afns.ualberta.ca/cgview_server) and the BLASTn algorithm (60% identity cut-off) with the NRRL B-1299 genome as a reference. Phylogenetic analyses were performed using MEGA6 [59] with a total of 61 amino acid sequences. These sequences were curated, and the poorly conserved regions were eliminated using GBLOCKS (version 0.91b) [60]. Putative -10 and -35 promoter boxes were identified using the PePPER webserver [61] or from consensus sequences proposed by McCracken *et al.* [62] for Gram-positive bacteria. Signal peptide sequences were identified using the SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP/>). Protein sequence comparison was performed using blastp against the non-redundant protein sequence database for local alignments and ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) for global alignments.

Phage manipulation

The lytic phase of the NRRL B-1299 prophage was induced by addition of 20 µg·mL⁻¹ of mitomycin C (Sigma-Aldrich, St Louis, MO) during the early growth phase (attenuance at 600 nm of 0.2). After 5 h culture, the excised phages were precipitated and then extracted as previously described for lactococcal phages [63].

Bacterial strains and growth medium

E. coli TOP10 competent cells (Life Technologies, Carlsbad, CA, USA) were used for the cloning experiments. Recombinant DSR-M and DSR-DP enzymes were produced in *E. coli* BL21 Star DE3 cells, and BRS-A was produced in *E. coli* BL21-AI cells (Life Technologies).

Modified ZYM5052 medium [64] was used for the production of recombinant GSs as described in Table S3.

Cloning of glucansucrase genes

Glucansucrase genes were first amplified by PCR from a *L. citreum* NRRL B-1299 genomic DNA template using the primers listed in Table S4. Using the Gateway cloning system, the genes of interest were first inserted into the pENTR/D-TOPO vector (Life Technologies), and then recombined with destination vectors pET-53-DEST, pET-55-DEST (both VWR, Radnor, PA, USA) and pBAD-DEST49 (Life Technologies) for *dsrM*, *dsrDP* and *brsA* expression, respectively. Positive clones were selected on LB plates supplemented with 100 µg·mL⁻¹ ampicillin. The absence of mutation was confirmed by gene sequencing (GATC Biotech).

Production of α -transglucosylases

Starter cultures

Thirty milliliters of LB medium, supplemented with ampicillin ($100 \mu\text{g}\cdot\text{mL}^{-1}$), was inoculated with $200 \mu\text{L}$ of freshly transformed *E. coli* BL21 Star DE3 cells (for production of recombinant DSR-DP and DSR-M) or BL21-AI DE3 cells (for production of recombinant BRS-A), and incubated overnight at 37°C under agitation (200 rpm).

Erlenmeyer flask cultures

One liter of modified ZYM5052 medium (Table S3), supplemented with ampicillin ($100 \mu\text{g}\cdot\text{mL}^{-1}$), was inoculated with the starter culture at an attenuation at 600 nm of 0.05. The cultures were incubated at 21°C under agitation (150 rpm). After 26 h incubation, the cells were harvested by centrifugation ($5\,000 \text{ g}$, 15 min , 4°C), resuspended in 50 mM sodium acetate buffer ($\text{pH } 5.75$) at a final attenuation at 600 nm of 80, and disrupted by sonication. The recombinant enzymes were recovered after centrifugation of the crude cell extract ($15\,000 \text{ g}$, 30 min , 4°C).

Purification of recombinant enzymes

Purification of DSR-DP by Strep-tag II affinity chromatography

Purification was performed using the AktaXpress system (GE Healthcare, Little Chalfont, Buckinghamshire, UK) in a cold room. Enzymes in the soluble fraction (recovered as described previously) were pre-equilibrated with 20 mM imidazole and 280 mM NaCl at $\text{pH } 7.4$, and filtered through a $0.22 \mu\text{m}$ cartridge. The proteins were then injected at $1 \text{ mL}\cdot\text{min}^{-1}$ into a 5 mL StrepTrap HP™ column (pre-packed with StrepTactin™ Sepharose high performance chromatography medium, GE Healthcare), previously equilibrated with buffer A ($1 \times \text{PBS}$, 280 mM NaCl, $\text{pH } 7.4$). After 1 h binding, the Strep-tagged enzymes were eluted at $4 \text{ mL}\cdot\text{min}^{-1}$ using a gradient of D-desthio-biotin (from 0 – 2.5 mM) in buffer A over 20 column volumes. Fractions (3 mL) of the eluate were desalted by loading onto a 10-DG column (Bio-Rad, Hercules, CA, USA) pre-equilibrated in 50 mM sodium acetate buffer, $\text{pH } 5.75$, with 0.45 mM CaCl_2 and 100 mM NaCl.

Purification of BRS-A and DSR-M on Sephacryl® S-300 HR gel

One volume of Sephacryl® S-300 HR resin (GE Healthcare) was washed with 10 volumes of buffer B (50 mM sodium acetate buffer, $\text{pH } 5.75$, 0.45 mM CaCl_2 , 100 mM NaCl). Ten volumes of enzymatic extracts were added to

the resin. After 2 h incubation at 4°C , the unbound proteins were removed by washing the resin four times with 10 volumes of buffer B. The proteins were then eluted using 7×2 volumes of buffer B containing $0.25 \text{ g}\cdot\text{L}^{-1}$ dextran $1500 \text{ g}\cdot\text{mol}^{-1}$ (Sigma Aldrich) and 6×2 volumes of buffer B containing of $50 \text{ g}\cdot\text{L}^{-1}$ dextran $1500 \text{ g}\cdot\text{mol}^{-1}$. To remove the dextran, purified enzymatic fractions were desalted onto 10-DG columns (Bio-Rad) pre-equilibrated in 50 mM sodium acetate buffer, $\text{pH } 5.75$, with 0.45 mM CaCl_2 and 100 mM NaCl.

Protein purification analyses

The protein concentration was determined by spectroscopy at 280 nm on a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) using the protein theoretical molar extinction coefficient and molecular mass calculated using the ExpASY ProtParam tool (<http://web.expasy.org/protparam/>).

NuPAGE® Novex® 3–8% Tris-acetate protein gels, 1.5 mm thick (Life Technologies), were used to analyze the purified protein extracts. The samples were diluted in Laemmli sample buffer (Bio-Rad), and $15 \mu\text{L}$ of each sample was loaded onto a precast gel. The Precision Plus Protein All Blue standard (Bio-Rad) was used as the protein ladder. The separation was performed in NuPAGE® Tris-acetate SDS running buffer (Life Technologies) for 1 h at 150 V constant voltage. The proteins were then stained using PAGE BLUE protein staining solution (Thermo Scientific).

Enzymatic assay using the dinitrosalicylic acid method

Enzymatic assays were performed at 30°C in 50 mM sodium acetate at $\text{pH } 5.75$. The enzyme activities were determined by measuring the initial amount of released fructose for an initial sucrose concentration of 292 mM . The fructose concentrations were determined using the dinitrosalicylic acid method [65]. One unit was defined as the amount of enzyme that catalyzes the production of $1 \mu\text{mol}$ fructose per min under the assay conditions.

Determination of the optimal pH and temperature

Determination of the optimal pH and temperature was performed in the presence of 292 mM sucrose. The optimal pH for enzyme activity was determined in duplicate in 50 mM citrate phosphate buffer, varying from 3.5 to 7 pH units (increments of 0.5 units) at 30°C . The optimal temperature was determined in 50 mM acetate sodium buffer at $\text{pH } 5.75$. The assays were performed at temperatures of 23 , 26 , 30 , 33 , 37 , 40 and 45°C in duplicate.

Enzymatic reaction and product characterization

Reaction with sucrose as the sole substrate

Enzymatic reactions were performed at 30 °C using 1 U·mL⁻¹ enzyme in 50 mM sodium acetate buffer at pH 5.75 supplemented with 292 mM of sucrose. Samples were taken at regular intervals until total sucrose depletion. The reaction was stopped by 5 min incubation at 95 °C, and samples were stored at -20 °C until further analysis.

Acceptor reactions

Acceptor reactions with sucrose and various α-(1→6) dextrans ranging from 1500 to 2 × 10⁶ g·mol⁻¹ (Sigma-Aldrich) and at various concentrations were performed using the BRS-A enzyme (1 U·mL⁻¹) at 30 °C in 50 mM sodium acetate buffer at pH 5.75. The reactions were incubated for 8–16 h. The total sucrose depletion was checked by HPAEC-PAD analysis (see below). Samples were taken at various times, and the reaction was stopped by 5 min incubation at 95 °C. Samples were stored at -20 °C until further analyses.

HPAEC-PAD analysis

HPAEC-PAD analyses were performed using a CarboPacTM PA100 analytical column (4 mm × 250 mm) coupled with a CarboPacTM PA100 guard (4 mm × 50 mm). Glucose, fructose, leucrose and sucrose were separated using a sodium acetate gradient (6–500 mM) in 150 mM NaOH over 36 min (1 mL·min⁻¹), and quantification was performed using standards of these sugars at 5, 10, 15 and 20 mg·kg⁻¹. Samples were prepared in ultra-pure water, and diluted to a final concentration of 10 mg·kg⁻¹ of total sugars. The percentages of glucosyl moieties derived from sucrose and incorporated into free glucose (%G_{glucose}) or leucrose (%G_{leucrose}) were calculated as follows:

$$\%G_{\text{glucose}} = \frac{([\text{Glucose}]_{t_f} - [\text{Glucose}]_{t_0}) \times 342}{([\text{Sucrose}]_{t_0} - [\text{Sucrose}]_{t_f}) \times 180} \times 100$$

$$\%G_{\text{leucrose}} = \frac{([\text{Leucrose}]_{t_f} - [\text{Leucrose}]_{t_0})}{([\text{Sucrose}]_{t_0} - [\text{Sucrose}]_{t_f})} \times 100$$

HPSEC analysis

HPSEC analyses were performed using Shodex OH-Pak SB-802.5 and SB-805 columns (Showa Denko, Minato-ki, Tokyo, Japan) in series coupled with a Shodex OH-Pak SB-G guard column, and placed in a 70 °C oven. The samples were diluted to a maximum of 10 g·kg⁻¹ of total sugars in the eluent (0.45 M NaNO₃ and 1% ethylene glycol). Elution was performed at a flow rate of 0.3 mL·min⁻¹.

The percentage of glucosyl units from sucrose incorporated into glucan (%G_{glucan}) was calculated as follows:

$$\%G_{\text{glucan}} = \frac{\text{Glucan final area} \times 342}{\text{Sucrose initial area} \times 162}$$

The percentage of glucosyl units from sucrose incorporated into oligosaccharides (%G_{oligosaccharides}) was calculated as follows:

$$\%G_{\text{oligosaccharides}} = 100\% - (\%G_{\text{glucan}} + \%G_{\text{glucose}} + \%G_{\text{leucrose}})$$

The weight-average molar masses of synthesized dextrans were determined using a calibration curve with standards of 10 g·kg⁻¹ of fructose, sucrose, maltoheptaose (Carbosynth, Berkshire, UK) and dextrans of 11 300, 39 100, 42 000, 68 400, 503 × 10³ and 2 × 10⁶ g·mol⁻¹ (Sigma-Aldrich).

¹H-NMR analysis

Freeze-dried polymer samples were resuspended in deuterated water at a final concentration of 20 mg·mL⁻¹. ¹H-NMR spectra were recorded on a Bruker Advance 500 MHz spectrometer (Bruker, Billerica, MA, USA), and the data were processed using TOPSPIN 3.0 software (Bruker). The spectra were acquired using a zgpr pulse sequence (with water suppression). The percentage of α-(1→2) linkages in the branched dextran was calculated as previously described [37]. All measurements were performed at 298 K, and the chemical shifts were compared to an internal reference (sodium 2,2,3,3-tetradeutero-3-trimethylsilylpropanoate).

Nucleotide sequence accession number

The draft genome of *L. citreum* NRRL B-1299 was deposited in the European Molecular Biology Laboratory/European Nucleotide Archive database under accession numbers CCNH01000001–CCNH01000010 for the chromosome sequence and LM651913 for the pB1299 plasmid sequence.

Acknowledgments

This work was supported by the French National Research Agency (ANR-10-ALIA-0003 Oenopolys 2011-2013, ANR-12-CDII-0005, Engel 2012-2015) and the Région Midi-Pyrénées (France). We are grateful to MetaSys, the Metabolomics & Fluxomics Center at the Laboratory for Engineering of Biological Systems & Processes (Toulouse, France), for performing NMR experiments. We thank the ICEO facility dedicated to enzyme screening and discovery that is part of the Integrated Screening Platform of Toulouse (PICT, IB-iSA) for providing access to HPLC equipment and protein purification system.

Author contributions

DP, MV, PM, CFF, CM and MRS planned experiments, DP, MV, LU, SM and VL performed the experiments, DP, MV, SM, CM and MRS analyzed the data, CFF contributed reagents or other essential material, and DP, MV, CM and MRS wrote the paper.

References

- Naessens M, Cerdobbel A, Soetaert W & Vandamme EJ (2005) *Leuconostoc* dextransucrase and dextran: production, properties and applications. *J Chem Technol Biotechnol* **80**, 845–860.
- Freitas F, Alves VD & Reis MAM (2011) Advances in bacterial exopolysaccharides: from production to biotechnological applications. *Trends Biotechnol* **29**, 388–398.
- Badel S, Bernardi T & Michaud P (2011) New perspectives for *Lactobacilli* exopolysaccharides. *Biotechnol Adv* **29**, 54–66.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM & Henrissat B (2013) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, 490–495.
- Leemhuis H, Pijning T, Dobruchowska JM, van Leeuwen SS, Kralj S, Dijkstra BW & Dijkhuizen L (2013) Glucansucrases: three-dimensional structures, reactions, mechanism, α -glucan analysis and their implications in biotechnology and food applications. *J Biotechnol* **163**, 250–272.
- Devulapalle KS, Goodman SD, Gao Q, Hemsley A & Mooser G (1997) Knowledge-based model of a glucosyltransferase from the oral bacterial group of mutans streptococci. *Protein Sci* **6**, 2489–2493.
- van Hijum SAFT, Kralj S, Ozimek LK, Dijkhuizen L & van Geel-Schutten IGH (2006) Structure–function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. *Microbiol Mol Biol Rev* **70**, 157–176.
- Vujičić-Žagar A, Pijning T, Kralj S, López CA, Eeuwema W, Dijkhuizen L & Dijkstra BW (2010) Crystal structure of a 117 kDa glucansucrase fragment provides insight into evolution and product specificity of GH70 enzymes. *Proc Natl Acad Sci USA* **107**, 21406–21411.
- Ito K, Ito S, Shimamura T, Weyand S, Kawarasaki Y, Misaka T, Abe K, Kobayashi T, Cameron AD & Iwata S (2011) crystal structure of glucansucrase from the dental caries pathogen *Streptococcus mutans*. *J Mol Biol* **408**, 177–186.
- Brison Y, Pijning T, Malbert Y, Fabre E, Mourey L, Morel S, Potocki-Veronese G, Monsan P, Tranier S, Remaud-Simeon M et al. (2012) Functional and structural characterization of α -1,2 branching sucrose derived from DSR-E glucansucrase. *J Biol Chem* **287**, 7915–7924.
- Pijning T, Vujičić-Žagar A, Kralj S, Dijkhuizen L & Dijkstra BW (2012) Structure of the α -1,6/ α -1,4-specific glucansucrase GTFA from *Lactobacillus reuteri* 121. *Acta Crystallogr* **68**, F1448–F1454.
- MacGregor EA, Jespersen HM & Svensson B (1996) A circularly permuted α -amylase-type α/β -barrel structure in glucan-synthesizing glucosyltransferases. *FEBS Lett* **378**, 263–266.
- Monchois V (1999) Glucansucrases: mechanism of action and structure–function relationships. *FEMS Microbiol Rev* **23**, 131–151.
- Moulis C, Joucla G, Harrison D, Fabre E, Potocki-Veronese G, Monsan P & Remaud-Simeon M (2006) Understanding the polymerization mechanism of glycoside-hydrolase family 70 glucansucrases. *J Biol Chem* **281**, 31254–31267.
- Giffard PM & Jacques NA (1994) Definition of a fundamental repeating unit in streptococcal glucosyltransferase glucan-binding regions and related sequences. *J Dent Res* **73**, 1133–1141.
- Jeanes A, Haynes WC, Wilham CA, Rankin JC, Melvin EH, Austin MJ, Cluskey JE, Fisher BE, Tsuchiya HM & Rist CE (1954) Characterization and classification of dextrans from ninety-six strains of bacteria 1b. *J Am Chem Soc* **76**, 5041–5052.
- Brooker BE (1977) Ultrastructural surface changes associated with dextran synthesis by *Leuconostoc mesenteroides*. *J Bacteriol* **131**, 288–292.
- Kobayashi M & Matsuda K (1977) Structural characteristics of dextrans synthesized by dextransucrases from *Leuconostoc mesenteroides* NRRL B-1299. *Agric Biol Chem* **41**, 1931–1937.
- Sarbini SR, Kolida S, Naeye T, Einerhand A, Brison Y, Remaud-Simeon M, Monsan P, Gibson GR & Rastall RA (2011) In vitro fermentation of linear and α -1,2-branched dextrans by the human fecal microbiota. *Appl Environ Microbiol* **77**, 5307–5315.
- Sarbini SR, Kolida S, Naeye T, Einerhand AW, Gibson GR & Rastall RA (2013) The prebiotic effect of α -1,2 branched, low molecular weight dextran in the batch and continuous faecal fermentation system. *J Funct Foods* **5**, 1938–1946.
- Bounaix M-S, Gabriel V, Robert H, Morel S, Remaud-Siméon M, Gabriel B & Fontagné-Faucher C (2010) Characterization of glucan-producing *Leuconostoc* strains isolated from sourdough. *Int J Food Microbiol* **144**, 1–9.
- Monchois V, Willemot R-M, Remaud-Simeon M, Croux C & Monsan P (1996) Cloning and sequencing of a gene coding for a novel dextransucrase from *Leuconostoc mesenteroides* NRRL B-1299 synthesizing only α (1–6) and α (1–3) linkages. *Gene* **182**, 23–32.

- 23 Monchois V (1998) Cloning and sequencing of a gene coding for an extracellular dextransucrase (DSRB) from *Leuconostoc mesenteroides* NRRL B-1299 synthesizing only a $\alpha(1-6)$ glucan. *FEMS Microbiol Lett* **159**, 307–315.
- 24 Bozonnet S, Dols-Laffargue M, Fabre E, Pizzut S, Remaud-Simeon M, Monsan P & Willemot R-M (2002) Molecular characterization of DSR-E, an α -1,2 linkage-synthesizing dextransucrase with two catalytic domains. *J Bacteriol* **184**, 5753–5761.
- 25 Fabre E, Bozonnet S, Arcache A, Willemot R-M, Vignon M, Monsan P & Remaud-Simeon M (2005) Role of the two catalytic domains of DSR-E dextransucrase and their involvement in the formation of highly α -1,2 branched dextran. *J Bacteriol* **187**, 296–303.
- 26 Kim D & Robyt JF (1995) Dextransucrase constitutive mutants of *Leuconostoc mesenteroides* B-1299. *Enzyme Microb Technol* **17**, 1050–1056.
- 27 Liu Y, Harrison PM, Kunin V & Gerstein M (2004) Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol* **5**, R64.
- 28 Siguier P, Perochon J, Lestrade L, Mahillon J & Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**, D32–D36.
- 29 Kim JF, Jeong H, Lee J-S, Choi S-H, Ha M, Hur C-G, Kim J-S, Lee S, Park H-S, Park Y-H *et al.* (2008) Complete genome sequence of *Leuconostoc citreum* KM20. *J Bacteriol* **190**, 3093–3094.
- 30 Laguerre S, Amari M, Vuillemin M, Robert H, Loux V, Klopp C, Morel S, Gabriel B, Remaud-Siméon M, Gabriel V *et al.* (2012) Genome sequences of three *Leuconostoc citreum* strains, LBAE C10, LBAE C11, and LBAE E16, isolated from wheat sourdoughs. *J Bacteriol* **194**, 1610–1611.
- 31 Passerini D, Vuillemin M, Laguerre S, Amari M, Loux V, Gabriel V, Robert H, Morel S, Monsan P, Gabriel B *et al.* (2014) Complete genome sequence of *Leuconostoc citreum* strain NRRL B-742. *Genome Announc* **2**, e01179-14.
- 32 Mahony J, Ainsworth S, Stockdale S & van Sinderen D (2012) Phages of lactic acid bacteria: the role of genetics in understanding phage-host interactions and their co-evolutionary processes. *Virology* **434**, 143–150.
- 33 Lu Z, Altermann E, Breidt F & Kozyavkin S (2010) Sequence analysis of *Leuconostoc mesenteroides* bacteriophage I-A4 isolated from an industrial vegetable fermentation. *Appl Environ Microbiol* **76**, 1955–1966.
- 34 Kleppen HP, Nes IF & Holo H (2012) Characterization of a *Leuconostoc* bacteriophage infecting flavor producers of cheese starter cultures. *Appl Environ Microbiol* **78**, 6769–6772.
- 35 Ali Y, Kot W, Atamer Z, Hinrichs J, Vogensen FK, Heller KJ & Neve H (2013) Classification of lytic bacteriophages attacking dairy *Leuconostoc* starter strains. *Appl Environ Microbiol* **79**, 3628–3636.
- 36 Zhou Y, Liang Y, Lynch KH, Dennis JJ & Wishart DS (2011) PHAST: a fast phage search tool. *Nucleic Acids Res* **39**, W347–W352.
- 37 Brison Y, Fabre E, Moulis C, Portais J-C, Monsan P & Remaud-Siméon M (2009) Synthesis of dextrans with controlled amounts of α -1,2 linkages using the transglucosidase GBD-CD2. *Appl Microbiol Biotechnol* **86**, 545–554.
- 38 Amari M, Gabriel V, Robert H, Morel S, Moulis C, Gabriel B, Remaud-Siméon M & Fontagné-Faucher C (2014) Overview of the glucansucrase equipment of *Leuconostoc citreum* LBAE-E16 and LBAE-C11, two strains isolated from sourdough. *FEMS Microbiol Lett*, **362**, 1–8.
- 39 Argüello-Morales MA, Remaud-Simeon M, Pizzut S, Sarçabal P, Willemot R & Monsan P (2000) Sequence analysis of the gene encoding alternansucrase, a sucrose glucosyltransferase from *Leuconostoc mesenteroides* NRRL B-1355. *FEMS Microbiol Lett* **182**, 81–85.
- 40 Olivares-Illana V, López-Munguía A & Olvera C (2003) Molecular characterization of inulosucrase from *Leuconostoc citreum*: a fructosyltransferase within a glucosyltransferase. *J Bacteriol* **185**, 3606–3612.
- 41 Moulis C, Arcache A, Escalier P-C, Rinaudo M, Monsan P, Remaud-Simeon M & Potocki-Veronese G (2006) High-level production and purification of a fully active recombinant dextransucrase from *Leuconostoc mesenteroides* NRRL B-512F. *FEMS Microbiol Lett* **261**, 203–210.
- 42 Jung JY, Lee SH & Jeon CO (2012) Complete genome sequence of *Leuconostoc carnosum* strain JB16, isolated from kimchi. *J Bacteriol* **194**, 6672–6673.
- 43 Jung JY, Lee SH, Lee SH & Jeon CO (2012) Complete genome sequence of *Leuconostoc mesenteroides* subsp. *mesenteroides* strain J18, isolated from kimchi. *J Bacteriol* **194**, 730–731.
- 44 Lee SH, Jung JY, Lee SH & Jeon CO (2011) Complete genome sequence of *Leuconostoc kimchii* strain C2, isolated from kimchi. *J Bacteriol* **193**, 5548–5548.
- 45 Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N *et al.* (2006) Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* **103**, 15611–15616.
- 46 Bourne EJ, Sidebotham RL & Weigel H (1972) Studies on dextrans and dextranses: part X. Types and percentages of secondary linkages in the dextrans elaborated by *Leuconostoc mesenteroides* NRRL B-1299. *Carbohydr Res* **22**, 13–22.
- 47 Seymour FR, Slodki ME, Plattner RD & Jeanes A (1977) Six unusual dextrans: methylation structural

- analysis by combined g.l.c.—m.s. of per-O-acetyl-aldononitriles. *Carbohydr Res* **53**, 153–166.
- 48 Watanabe T, Shishido K, Kobayashi M & Matsuda K (1978) Acetolysis of *Leuconostoc mesenteroides* NRRL B-1299 dextran. Isolation and characterization of tetrasacharrides containing secondary linkages from the borate-insoluble fraction. *Carbohydr Res* **61**, 119–128.
- 49 Dols M, Remaud-Simeon M, Willemot RM, Vignon M & Monsan P (1998) Characterization of the different dextranase activities excreted in glucose, fructose, or sucrose medium by *Leuconostoc mesenteroides* NRRL B-1299. *Appl Environ Microbiol* **64**, 1298–1302.
- 50 Monchois V, Reverte A, Remaud-Simeon M, Monsan P & Willemot R-M (1998) Effect of *Leuconostoc mesenteroides* NRRL B-512F dextranase carboxy-terminal deletions on dextran and oligosaccharide synthesis. *Appl Environ Microbiol* **64**, 1644–1649.
- 51 Moulis C, Vaca Medina G, Suwannarangsee S, Monsan P, Remaud-Simeon M & Potocki-Veronese G (2008) One-step synthesis of isomalto-oligosaccharide syrups and dextrans of controlled size using engineered dextranase. In *Biocatalysis and biotransformation* pp. 141–151. Taylor & Francis, Abingdon, UK.
- 52 Kralj S, van Geel-Schutten IGH, Faber EJ, van der Maarel MJEC & Dijkhuizen L (2005) Rational transformation of *Lactobacillus reuteri* 121 reuteranase into a dextranase. *Biochemistry* **44**, 9206–9216.
- 53 Hoshino T, Fujiwara T & Kawabata S (2012) Evolution of cariogenic character in *Streptococcus mutans*: horizontal transmission of glycosyl hydrolase family 70 genes. *Sci Rep* **2**, 518.
- 54 Argimón S, Alekseyenko AV, DeSalle R & Caufield PW (2013) Phylogenetic analysis of glucosyltransferases and implications for the coevolution of mutans *Streptococci* with their mammalian hosts. *PLoS One* **8**, e56305.
- 55 Kralj S, van Geel-Schutten GH, Dondorf MMG, Kirsanovs S, van der Maarel MJEC & Dijkhuizen L (2004) Glucan synthesis in the genus *Lactobacillus*: isolation and characterization of glucanase genes, enzymes and glucan products from six different strains. *Microbiology* **150**, 3681–3690.
- 56 Bryson K, Loux V, Bossy R, Nicolas P, Chaillou S, van de Guchte M, Penaud S, Maguin E, Hoebeke M, Bessières P et al. (2006) AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system. *Nucleic Acids Res* **34**, 3533–3545.
- 57 Grant JR & Stothard P (2008) The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res* **36**, W181–W184.
- 58 Darling AE, Mau B & Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**, e11147.
- 59 Tamura K, Stecher G, Peterson D, Filipiński A & Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.
- 60 Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540–552.
- 61 De Jong A, Pietersma H, Cordes M, Kuipers OP & Kok J (2012) PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genom* **13**, 299.
- 62 McCracken A, Turner MS, Giffard P, Hafner LM & Timms P (2000) Analysis of promoter sequences from *Lactobacillus* and *Lactococcus* and their activity in several *Lactobacillus* species. *Arch Microbiol* **173**, 383–389.
- 63 Chopin A, Bolotin A, Sorokin A, Ehrlich SD & Chopin M-C (2001) Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res* **29**, 644–651.
- 64 Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* **41**, 207–234.
- 65 Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* **31**, 426–428.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Fig. S1. Whole-genome alignment of NRRL B-1299, NRRL B-742 and KM20 strains.

Fig. S2. Differences between newly sequenced *dsrA/dsrB* genes and previously published sequences.

Fig. S3. Detailed results for ¹H-NMR spectroscopy.

Fig. S4. Genetic context of NRRL B-1299 glucanase genes.

Table S1. Strains and vectors used for expression of recombinant glucanases and the level of production in optimized ZYM5052 medium.

Table S2. Primer pairs designed to close genome gaps.

Table S3. Final concentration of α-lactose, L-arabinose, glucose and glycerol in modified ZYM5052 medium for production of each glucanase.

Table S4. Primers used to amplify glucanase genes.