



**HAL**  
open science

## **ANALYSE GENOMIQUE DE LA COMPOSITION DU MICROBIOTE CAECAL DU LAPIN**

Quentin Le Graverand, Yann Labrune, Patrick Aymard, Hervé Chapuis, Julie Demars, Virginie Helies, Juliette Riquet, Julien Ruesche, Olivier Zemb, Hélène Gilbert, et al.

► **To cite this version:**

Quentin Le Graverand, Yann Labrune, Patrick Aymard, Hervé Chapuis, Julie Demars, et al.. ANALYSE GENOMIQUE DE LA COMPOSITION DU MICROBIOTE CAECAL DU LAPIN. Journées de la Recherche Cunicole, ITAVI; INRAE, Mar 2023, Le Mans (72000), France. pp.146-150. hal-04646187

**HAL Id: hal-04646187**

**<https://hal.inrae.fr/hal-04646187v1>**

Submitted on 12 Jul 2024

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

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

Public Domain

# ANALYSE GENOMIQUE DE LA COMPOSITION DU MICROBIOTE CAECAL DU LAPIN

**Le Graverand Q.<sup>1</sup>, Labrune Y.<sup>1</sup>, Aymard P.<sup>1</sup>, Chapuis H.<sup>1</sup>, Demars J.<sup>1</sup>, Helies V.<sup>1</sup>,  
Riquet J.<sup>1</sup>, Ruesche J.<sup>1</sup>, Zemb O.<sup>1</sup>, Gilbert H.<sup>1</sup>, Garreau H.<sup>1</sup>**

<sup>1</sup> GenPhySE, Université de Toulouse, INRAE, 31326 Castanet-Tolosan, France

\*Auteur correspondant : herve.garreau@inrae.fr

**Résumé** – Le but de cette étude est d'identifier des régions chromosomiques (QTL) associées à la composition du microbiote caecal du lapin pour mieux connaître le déterminisme génétique de ce caractère. L'ADN ribosomique 16S de la communauté bactérienne du caecum a été extrait et séquencé chez 525 lapins, issus d'une lignée sélectionnée sur l'efficacité alimentaire et d'une lignée témoin non sélectionnée. Les séquences, après raréfaction et filtration, ont permis de retenir 859 OTUs (operational taxonomic unit). Les lapins ont été génotypés avec la puce Affimetrix Axiom Rabbit 200K. Après contrôle qualité 514 lapins et 103 072 SNPs ont été retenus pour l'analyse d'association. Quarante-deux QTLs ont été identifiés pour 69 OTUs, avec un total de 492 SNPs significativement associés à leur abondance. La plupart des QTLs étaient associés à l'abondance d'une seule OTU. Huit QTL étaient associés à 2 ou 3 OTUs. Des gènes candidats liés à la réponse immunitaire ou au métabolisme des acides biliaires localisés dans ces régions sont discutés.

**Abstract – Genome-wide association study of the rabbit caecal microbiota composition.** The aim of our study was to identify genomic regions influencing the rabbit caecal microbiota composition for a better knowledge of the genetic determinism of this trait. The ribosomal 16S DNA was extracted from the microbial community of the caecum and then sequenced in 525 rabbits originated either from a line selected on feed efficiency or a control non-selected line. After rarefaction and filtration, the sequences were used to identify 859 OTUs (operational taxonomic units) or phlotypes. The rabbits were genotyped with the Affimetrix Axiom Rabbit 200K chip. After quality control 514 rabbits and 103,072 SNPs were retained for association study analyses. Eighty-one QTLs were identified for 69 OTUs' abundance with a total of 492 significant SNPs ( $P < 0.05$  genome-wide). Most QTLs were associated with the abundance of a single OTU. Eight QTLs were associated with 2 or 3 OTUs. Candidate genes related to immune response or bile acid metabolism in these regions are discussed.

## Introduction

L'appareil digestif du lapin héberge une communauté microbienne complexe et diversifiée qui joue un rôle clé dans de nombreuses fonctions physiologiques telles que la digestion ou l'immunité (Combes et al., 2011). La composition de la flore microbienne est affectée par des facteurs environnementaux et nutritionnels mais aussi par la génétique de l'hôte. La connaissance du déterminisme génétique de la composition du microbiote est un préalable indispensable à la possibilité d'intégrer cette information dans les schémas de sélection. Cette étude a pour but de rechercher les régions chromosomiques (QTL) associées à la composition du microbiote, par analyses d'association. Des gènes candidats positionnels et fonctionnels potentiellement en lien avec la composition du microbiote, sont discutés.

## 1. Matériel et méthodes

### 1.1. Animaux

Les animaux étaient issus de la lignée INRA 1001 (Larzul et De Rochambeau, 2005) et ont été élevés à l'installation expérimentale INRAE de l'UMR GenPhySE (Castanet-Tolosan, France), conformément à la réglementation relative à l'expérimentation animale (agrément APAFiS numéro #18416). Le détail du protocole est décrit par Garreau et al. (2019). Brièvement, deux lignées ont été utilisées dans cette étude : la lignée G10, sélectionnée pendant 10 générations pour la réduction de la consommation résiduelle et la lignée G0 produite à partir d'embryons congelés de la population ancestrale de la lignée sélectionnée. Un total de 490 lapins G10 et 410 lapins G0 a été testé, en 3 bandes séparées de 42 jours. Chaque lapereau a été adopté 48 heures après la naissance. Dans chaque bande, la moitié des lapins de chaque lignée a été élevée par des lapines G0 et la seconde moitié par des lapines G10.

Au sevrage (32 jours), 456 lapins ont été placés en cage individuelle d'élevage et 144 en cage individuelle de digestibilité. Tous les animaux ont été nourris ad libitum avec le même aliment granulé commercial jusqu'à la fin de la période d'engraissement (63 jours).

### 1.2. Détermination de la composition du microbiote

Les échantillons de contenu caecal ont été collectés sur 451 lapins élevés en cage individuelle d'élevage et 139 lapins élevés en cage individuelle de digestibilité, juste après l'abattage à 63 jours, puis conservés à -80°C jusqu'à leur traitement. Le kit ZR Soil Microbe DNA MiniPrepTM (ZymoResearch, Freiburg, Allemagne) a été utilisé pour l'extraction de l'ADN microbien. L'ADNr 16S microbien a été caractérisé par séquençage à l'aide d'un séquenceur MiSeq (Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Espagne). Les séquences présentant un degré de similarité supérieur à 97 % entre elles ont été regroupées en OTU (operational taxonomic unit). La

table d'abondance des OTUs comprenait 11 323 OTUs pour 525 animaux pour lesquels nous disposions des séquences 16S, des performances d'ingéré et des généalogies complètes. Pour corriger les hétérogénéités de profondeur de séquençage entre échantillons, la table d'abondance a été raréfiée à 12 228 lectures (profondeur de séquençage minimale des échantillons) à l'aide du package phyloseq (McMurdie and Holmes, 2013) implémenté sur la version 4.0.2 du logiciel R. Les OTUs ont été filtrées selon deux critères. Les OTUs rares, absentes dans plus de 10% des échantillons, ou ayant une abondance relative moyenne (somme des abondances de l'OTU dans les échantillons divisée par le nombre total de lectures du jeu de données) inférieure à 0,01%, ont été éliminés. Au total, 859 OTUs ont finalement été retenues. Une transformation de Box-Cox a été appliquée pour normaliser les distributions des abondances des OTUs.

Le logiciel d'affiliation FROGS (Galaxy version 3.2.2, Escudie et al., 2018) et la base de données SILVA 16S pintail100 (version 138.1, Quast et al., 2013) ont été utilisés pour prédire la taxonomie des 859 OTUs filtrés. Les affiliations taxonomiques ont été effectuées avec l'algorithme blastn+ (Camacho et al., 2009). Si le pourcentage d'identité était inférieur à 90 %, l'affiliation était jugée trop peu fiable et les groupes phylogénétiques étaient alors considérés comme inconnus.

### 1.3. Génotypage des animaux

L'ADN des animaux a été extrait à partir de biopsies d'oreille préalablement collectées et stockées à -20°C. Les lapins ont été génotypés avec la puce Affimetrix Axiom Rabbit 200K (Santa Clara, CA, USA) au laboratoire Centro Nacional de Genotipado (CeGen, Santiago de Compostela, Spain). L'assemblage OryCun2.0 du génome du lapin (Carneiro et al., 2014) a servi de référence pour les positions des SNPs sur le génome.

Un contrôle de qualité a été effectué sur les génotypes avec le logiciel PLINK (version 1.07, Purcell et al., 2007). Tous les SNP dont la fréquence d'un des allèles était inférieure à 5% et dont le pourcentage de génotypage était inférieur à 5% ont été supprimés. Les animaux avec moins de 95% de génotypes validés ont été retirés de l'analyse. En outre, les compatibilités de transmission parent-descendant ont été vérifiées. Au total, 73 830 SNP et 11 animaux ont été écartés. De plus, sur la base des écarts par rapport à l'équilibre de Hardy-Weinberg ( $p < 4.10^{-7}$ ), 2 773 SNP ont été éliminés. Les SNP restants qui n'ont pas été cartographiés sur les autosomes (20 017) n'ont pas non plus été pris en compte dans les analyses. Finalement, sur 525 animaux et 199 692 marqueurs, 514 lapins et 103 072 SNP ont été conservés.

#### 1.4. Analyse d'association génomique

Une analyse de variance a été appliquée pour tester les effets fixes affectant les valeurs d'abondance raréfiées et transformées de chaque OTU. Le même jeu d'effets fixes a été retenu pour toutes les OTUs. Les effets ont été retenus s'ils étaient significatifs ( $p < 0,05$ ) pour au moins 10% des OTUs : effet combiné de la bande et de la plaque de séquençage (72% des OTUs), effet de la mère adoptive (37 %), du type de cage (31 %), et de la taille de portée au sevrage (10 %). L'étude d'association génomique a été réalisée avec le logiciel GEMMA (version 0.94.1, Zhou and Stephens, 2012) pour chacune des OTUs sur les données d'abondances raréfiées, transformées et corrigées pour les quatre effets fixes retenus. Le modèle animal mixte univarié suivant a été appliqué pour estimer les effets des marqueurs SNP pour chacune des 859 OTUs:

$$y^* = \beta_0 + Wa + u + \epsilon$$

$$\text{avec } u \sim N(\mathbf{0}, \sigma_u^2 \mathbf{G}) \text{ et } \epsilon \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$$

ou  $y^*$  est le vecteur des abondances ajustées de l'OTU,  $\beta_0$  est l'ordonnée à l'origine,  $W$  est le vecteur des génotypes des marqueurs,  $a$  est l'effet du marqueur,  $u$  est l'effet aléatoire polygénique additif et  $\epsilon$  est l'effet aléatoire résiduel.  $\sigma_u^2$  et  $\sigma_e^2$  sont les variances polygéniques et résiduelles.  $\mathbf{G}$  est la matrice de parenté génomique centrée calculée avec le logiciel GEMMA (Zhou and Stephens, 2012).

Afin de prendre en compte les tests multiples pour 103 072 marqueurs SNP, le seuil de significativité des  $p$ -values a été corrigé par la méthode de Bonferroni en utilisant le logiciel SimpleM (Gao et al., 2008, 2010). En prenant en compte la matrice de corrélation des génotypes SNP, le nombre total de tests indépendants a été établi à 20 551. Le seuil corrigé pour une  $p$ -value=0,05 est alors  $-\log_{10}\left(\frac{0,05}{20\,551}\right) = 5,61$

Une région QTL a été définie par un ou plusieurs SNP significatifs distants de moins de 1 Mb, plus la région flanquante de 1 Mb de part et d'autre. Des gènes candidats ont été recherchés dans la base de données Ensembl (release 103, Yates et al., 2020) dans ces régions génomiques. Parmi les 1 943 gènes identifiés à proximité des SNPs significatifs (< 1Mb), seuls les gènes candidats associés aux OTUs qui avaient plus de 20 SNPs significatifs ou étaient associés aux signaux d'association les plus forts ( $-\log_{10}(p\text{-value}) > 8$ ) sont discutés.

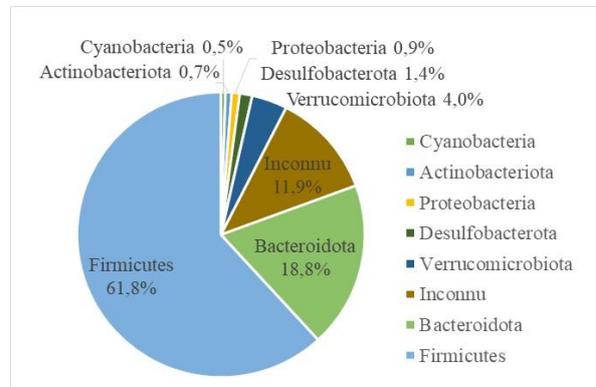
## 2. Résultats et discussion

### 2.1. Composition taxonomique du microbiote

Après raréfaction, 6 036 496 lectures ont été retenues. L'affiliation taxonomique des lectures montre que les phyla *Firmicutes* et *Bacteroidota* sont les plus abondants (61,8% et 18,8% des lectures, respectivement). La prédominance de ces deux phyla chez le lapin a été précédemment rapportée (Combes et

al., 2011; Velasco-Galilea et al., 2018). Un total de sept phyla différents a été identifié, et 11,9% des lectures n'ont pas pu être affiliées (Figure 1).

Figure 1 : Distribution des lectures par phylum



### 2.2. QTL associés avec la composition du microbiote

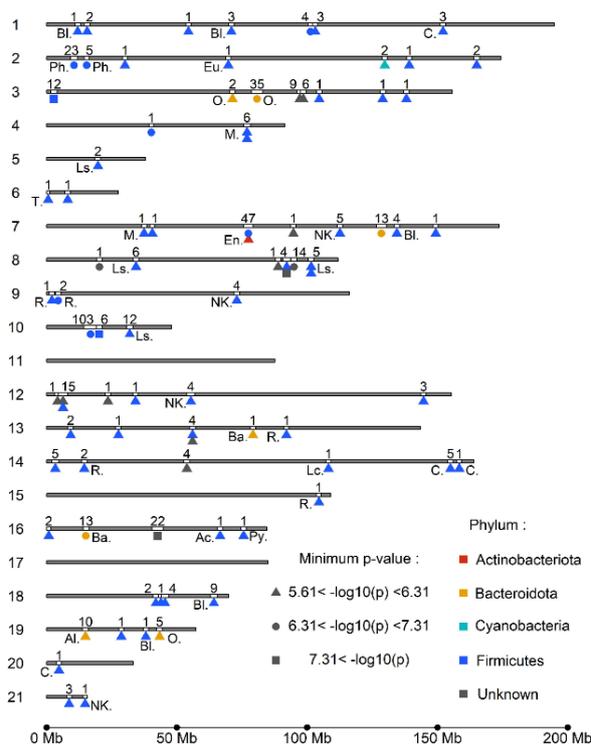
L'analyse d'association génomique met en évidence 492 SNPs significatifs ( $-\log_{10}(p\text{-value}) > 5,61$ ) pour 69 OTUs. Parmi celles-ci, 31 OTUs étaient affiliées à 16 genres bactériens connus et distincts. Les genres les plus fréquemment associés à des SNPs significatifs étaient *Ruminococcus* (4 OTUs), *Blautia* (4 OTUs) et le groupe *Lachnospiraceae NK4A136* (4 OTUs). Les SNPs significatifs étaient répartis sur 19 des 21 autosomes (Figure 2). Aucun n'a été détecté sur les chromosomes *Oryctolagus cuniculus 11* et *17* (*OCU11* et *OCU17*). Les signaux significatifs étaient distribués dans 81 régions génomiques avec un nombre inégal de SNPs : 34 régions comprenaient un seul SNP significatif, tandis qu'une région sur OCU10 comprenait 103 marqueurs significatifs. De même, le nombre de SNPs significatifs par OTU variait : 19 OTUs n'avaient qu'un seul SNP significatif, tandis qu'un maximum de 107 SNPs était associé de manière significative à un seul OTU appartenant à l'embranchement des *Firmicutes* (NR4209).

Cinquante OTUs n'étaient associées qu'à un seul QTL, 18 OTUs à deux QTL et une OTU à quatre QTL. L'OTU 317315 (groupe *Clostridia vadinBB60*) était associée au plus grand nombre de QTL : deux QTL situés sur OCU2 (29,0-31,0 et 163,9-166,2 Mb) et deux autres sur OCU3 (128,0-130,0 et 137,1-139,1 Mb). Sept régions génomiques étaient associées à l'abondance de deux. Ces régions étaient situées sur 7 chromosomes : OCU4, OCU7, OCU8, OCU10, OCU12, OCU13 et, OCU18 (Tableau 1). Une huitième région sur OCU7 était associée à l'abondance de trois OTUs appartenant au genre *Enterorhabdus* (NR4488) ou à des genres inconnus (NR7819, NR6518).

**Tableau 1 : Phyla des OTUs partageant une même région QTL**

OCU	Position (Mb)	OTUs	Phylum
4	75.5-78.2	299902	<i>Firmicutes</i>
		NR2948	<i>Firmicutes</i>
		NR7819	<i>Firmicutes</i>
7	75.6-79.3	NR4488	<i>Actinobacteriota</i>
		NR6518	<i>Firmicutes</i>
8	90.6-93.7	207532	<i>Firmicutes</i>
		NR2381	unknown
8	100.4-102.6	NR4583	<i>Firmicutes</i>
		NR7047	<i>Firmicutes</i>
10	19.1-21.3	NR4209	<i>Firmicutes</i>
		NR3368	<i>Firmicutes</i>
12	4.4-8.1	NR3758	unknown
		NR761	<i>Firmicutes</i>
13	54.9-57.1	NR229	<i>Firmicutes</i>
		325750	unknown
18	40.8-42.9	NR7011	<i>Firmicutes</i>
		NR7433	<i>Firmicutes</i>

**Figure 2 : Localisation des QTL identifiés pour la composition du microbiote caecal**



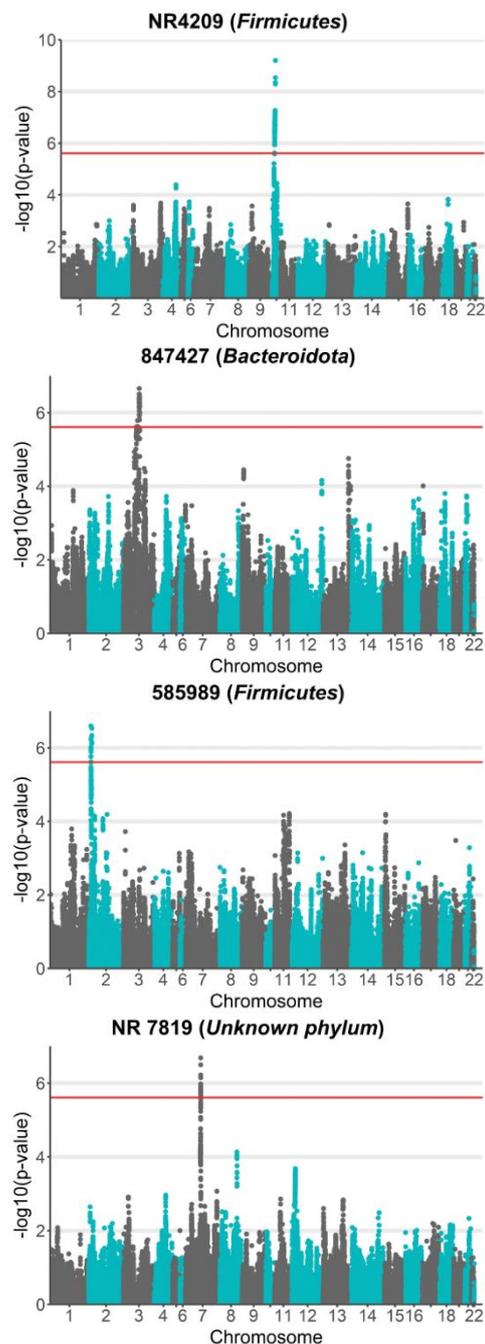
Chiffre au-dessus du QTL = nombre de SNPs significatifs.

Affiliation au Genre: abréviations au-dessous du QTL: Ac. Acetatifactor; Al. Alistipes; Ba. Bacteroides; Bl. Blautia; C. Christensenellaceae R-7 group; En. Enterorhabdus; Eu. Eubacterium ruminantium group; Lc. Lachnoclostridium; Ls. Lachnospiraceae NK4A136 group; M.

Marvinbryantia; NK. NK4A214 group; O. Odoribacter; Ph. Phascolarctobacterium; Py. Pygmaibacter; R. Ruminococcus

Quatre OTUs présentaient chacune plus de 23 SNP significatifs, représentant presque la moitié des marqueurs significatifs (196 SNPs) (Figure 3). Ces OTUs étaient principalement assignées aux phyla prédominants : 3 aux *Firmicutes* et 1 aux *Bacteroidota*. Seulement deux de ces OTU sont assignées à des genres connus : *Odoribacter* et *Phascolarctobacterium*.

**Figure 3. Manhatten plots pour les OTUs avec plus de 23 SNPs significatifs**



### 2.3. Gènes candidats

Un total de 1 359 gènes codant des protéines a été identifié dans les régions QTL détectées. La recherche de candidats a été limitée aux OTU associés à plus de 20 SNPs significatifs. Des candidats ont pu être identifiés pour 3 OTUs.

**OTU 847427 (*Bacteroidota, Odoribacter*, figure 4A):** deux QTL ont été identifiés sur OCU3. Pour la première région située entre 70,3 et 72,4 Mb, le gène *LYN* (proto-oncogène, tyrosine kinase de la famille Src ; 71,4-71,6 Mb) joue un rôle dans la croissance et la régulation des cellules hématopoïétiques, dans la régulation des lymphocytes B, dans les signaux liés à la transduction des cytokines et à la phagocytose des pathogènes (Bishop et al., 2014).

**OTU NR4209 (*Firmicutes*) :** 2 régions QTL ont été identifiées. Trois gènes proches de la première région QTL (OCU10, 14,2-19,4 Mb) sont des candidats potentiels : *NOD1* (Nucleotide Binding Oligomerization Domain Containing 1) est un récepteur protéique qui peut activer la genèse de follicules lymphoïdes isolés après la reconnaissance de certains peptidoglycane bactériens. *NOD1* a été décrit précédemment comme étant exprimé dans le caecum du lapin (Guo et al., 2017). Un autre candidat est *ENSOCUG00000012951*, orthologue du gène du poulet *TMSB4X* (84% d'identité) codant pour une protéine séquestrant l'actine. Il a été observé une plus faible expression de *TMSB4X* chez les poulets à faible consommation résiduelle alimentaire (Yang et al., 2017). Pour le second QTL (OCU10, 19,0-21,1 Mb), un candidat putatif est identifié. Le SNP le plus significatif de l'étude (AX-147162305 ; position : 20,1 Mb ; valeur  $p = 6,25 \cdot 10^{-10}$ ) est situé dans le locus AOA1, un gène codant pour l'acyloxyacyl hydrolase. Cette hydrolase contribue à la désacylation des lipopolysaccharides de la paroi cellulaire des bactéries et son absence peut retarder la guérison d'une infection par des bactéries Gram-négatives chez la souris (Lu et al., 2008).

**NR5836 (phylum inconnu) :** le gène candidat MALRD1, localisé dans la région mise en évidence sur OCU16 entre 40,4 et 44,7 Mb, contribue à la régulation de la synthèse des acides biliaires (Li et Chiang, 2015).

### Conclusions

L'identification de 81 QTL permet de documenter le déterminisme génétique du microbiote cœcal du lapin. Le lien entre ces régions génomiques et les OTUs associées reste à élucider. Il est également nécessaire d'évaluer si cette information génomique est pertinente pour la sélection.

### Remerciements

Nous remercions le personnel de l'installation expérimentale de GenPhySE ainsi que la plateforme bioinformatique genotoul Toulouse Occitanie (Bioinfo

Genotoul, doi : 10.15454/) et le groupe Sigenae pour l'aide ainsi que les ressources de calcul et de stockage. Cette étude fait partie du projet Feed-a-Gene, financé par le programme H2020 de l'Union européenne sous la convention de subvention n° 633531.

### Références

- Bishop, J. L., Roberts, M. E., Beer, J. L., Huang, M., Chehal, M. K., Fan, X., et al. (2014). Lyn activity protects mice from DSS colitis and regulates the production of IL-22 from innate lymphoid cells. *Mucosal Immunol.* 7, 405–416. doi: 10.1038/mi.2013.60
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics* 10, 1–9. doi: 10.1186/1471-2105-10-421.
- Carneiro, M., Rubin, C.-J., Di Palma, F., Albert, F. W., Alföldi, J., Barrio, A. M., et al. (2014). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science (80-. )*. 345, 1074 LP – 1079. doi: 10.1126/science.1253714.
- Combes, S., Michelland, R. J., Monteils, V., Cauquil, L., Soulié, V., Tran, N. U., et al., 2011. Postnatal development of the rabbit caecal microbiota composition and activity. *FEMS Microbiol. Ecol.* 77, 680–689. doi: 10.1111/j.1574-6941.2011.01148.x.
- Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., et al. (2018). FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics* 34, 1287–1294. doi: 10.1093/bioinformatics/btx791.
- Garreau, H., Ruesche, J., Gilbert, H., Balmisse, E., Benitez, F., Richard, F., et al. (2019). Estimating direct genetic and maternal effects affecting rabbit growth and feed efficiency with a factorial design. *J. Anim. Breed. Genet.* 136, 168–173. doi: 10.1111/jbg.12380.
- Guo, M., Wu, F., Hao, G., Qi, Q., Li, R., Li, N., et al. (2017). *Bacillus subtilis* improves immunity and disease resistance in rabbits. *Front. Immunol.* 8. doi: 10.3389/fimmu.2017.00354.
- Larzul, C., and De Rochambeau, H. (2005). Selection for residual feed consumption in the rabbit. *Livest. Prod. Sci.* 95, 67–72. doi: 10.1016/j.livprodsci.2004.12.007.
- Li, T., and Chiang, J. Y. L. (2015). Bile acids as metabolic regulators. *Curr. Opin. Gastroenterol.* 31, 159–165. doi: 10.1097/MOG.0000000000000156.
- Lu, M., Varley, A. W., Ohta, S., Hardwick, J., and Munford, R. S. (2008). Host Inactivation of Bacterial Lipopolysaccharide Prevents Prolonged Tolerance Following Gram-Negative Bacterial Infection. *Cell Host Microbe* 4, 293–302. doi: 10.1016/j.chom.2008.06.009.
- McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8. doi: 10.1371/journal.pone.0061217.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, 590–596. doi: 10.1093/nar/gks1219.
- Yang, L., Liu, S., Ding, J., Dai, R., He, C., Xu, K., et al. (2017). Gut microbiota co-microevolution with selection for host humoral immunity. *Front. Microbiol.* 8, 1–11. doi: 10.3389/fmicb.2017.01243.
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824. doi: 10.1038/ng.2310