UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO BIOLOGIA VEGETAL

Metabolic-derived functions during adaptation of *E. coli* to the mammalian gut

Miguel Filipe Marques Pedro

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Prof. Dr. Francisco Dionísio
Prof. Dr. Karina Xavier

2016

2016

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO BIOLOGIA VEGETAL

Metabolic-derived functions during adaptation of *E. coli* to the mammalian gut

Miguel Filipe Marques Pedro

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Prof. Dr. Francisco Dionísio
Prof. Dr. Karina Xavier

2016

Instituto Gulbenkian de Ciência

Metabolic-derived functions during adaptation of *E. coli* to the mammalian gut

Miguel Filipe Marques Pedro

This thesis was fully performed at Instituto Gulbenkian de Ciência (Oeiras, Portugal) under the direct supervision of Dr. Karina Xavier in collaboration with Dr. Isabel Gordo in the scope of the Master in Molecular Biology and Genetics of the Faculty of Sciences of the University of Lisbon. Prof. Dr. Francisco Dionísio was the internal designated supervisor in the scope of the Master in Molecular Biology and Genetics of the Faculty of Sciences of the University of Lisbon.

2016

Acknowledgements

Thank you to the whole IGC community for creating such an interesting setting.
Thank you to the BAS group for nestling me. A special thank you, with a lot of love, to Joana Dias. You always had everything on your mind, yet managed to help me every single time I needed, even if you had no idea what I was trying to say. Thank you for teaching me everything you could at a time when I didn't even know what to ask.

Thank you to

My parents for building the house;
Dionísio for showing me the door;
Isabel for opening it;
Karina for welcoming me;

Abstract

Mammals harbour a cosmos of microbes in their gastrointestinal tract throughout their life. This dense polymicrobial network is dominated by commensal bacteria that live mostly in harmony with their host. Disruption of this homeostasis is detrimental to nutrient acquisition and overall health of the host. Colonisation is the first step in persistence and/or infection in the GI tract, it is thus essential to understand how *in vivo* bacteria obtain energy, carbon and nitrogen to sustain their growth and resist elimination. This work focused on a particular species of enteric bacteria, by studying the best characterized *Escherichia coli* strain. Use of this commensal strain (K-12 MG1655) has been the most prevalent in recent studies, it is well characterised at the genetic level and has been studied in the intestinal environment. This worked was founded on the hypothesis that the establishment of a metabolic landscape compatible with the state and composition of the intestinal environment is an important pressure for *E. coli*, highlighting from an evolutionary scale the fundamental importance of metabolism for intestinal colonisation. Through the characterization of the genetic basis and the evolutionary and physiological effects of the predominant beneficial mutations during adaptation to the intestinal environment of mice, this work presents mechanistic evidence and novel insights on how bacteria can establish and persist in specific ecological settings of this complex ecosystem.

Keywords: Mammalian; Microbiota; *E. coli*; Evolution; Metabolism.

Resumo

O trato gastrointestinal de mamíferos incorpora uma comunidade abundante e diversa de microrganismos, vulgarmente designada microbiota. Além de seres eucariotas e archaea, as bactérias constituem a vasta maioria da microbiota com números que se aproximam do total de células do próprio hospedeiro. Estes seres são transmitidos verticalmente da progenitora para os descendentes, acompanhando-os desde os primeiros dias de desenvolvimento. Esta fase constituída por uma microbiota intestinal mais simples assume, portanto, um papel importante no subsequente desenvolvimento de uma microbiota mais complexa, mas também na aprendizagem nutricional e imunitária de animais. Devido à abundância de nutrientes disponibilizados e à maior densidade energética da alimentação pós-desmame, o ecosistema intestinal, praticamente despopulado à nascença é extremamente propenso à colonisação por enterobactérias que aí se multiplicam e persistem em enormes densidades populacionais. A alteração da dieta à base de oligosacáridos derivados do leite para carbohidratos de origem vegetal, causam uma alteração abrupta nas bactérias presentes no intestino à medida que estas são substituídas por espécies mais aptas a utilizar os nutrientes presentes. Daí em diante, um vasto número de espécies de duas linhas filogenéticas em particular, colonisam praticamente todo este nicho. Estas bactérias estabelecem redes de interações que se admite serem de extrema complexidade devido à especificidade de cada espécie, compreendendo interações negativas como a produção de compostos antimicrobianos que inibem outras espécies ou interações positivas em que os produtos metabólicos finais de uma espécie servem de alimento a outras, por exemplo. De acordo com a hipótese nicho-nutriente, são principalmente as interações metabólicas entre cada espécie e o perfil nutricional do intestino que dita a composição da microbiota. Especificamente, o sucesso de uma espécie está dependente da capacidade de divisão ser superior à taxa a que é eliminada, utilizando para isso um ou um número baixo de nutrientes limitantes no intestino. Um resultado importante advém deste fenômeno. Qualquer espécie invasora tem necessariamente de ultrapassar esta barreira de forma a atingir uma população de dimensões suficientes para causar efeitos adversos ao hospedeiro, no caso de estirpes patogénicas. Para tal, estirpes invasoras ocupam nichos disponíveis ou competem com espécies comensais com preferências nutricionais coincidentes. Alternativamente, várias espécies patogénicas ou oportunistas disrompem a estabilidade da microbiota comensal através da indução de inflamação ou tiram vantagem do uso de antibióticos que abrem novos nichos previamente ocupados, de forma a colonizarem o seu hospedeiro. É, portanto, importante compreender como certas espécies subsistem no ambiente intestinal, não só estirpes que prejudicam a saúde do seu hospedeiro, mas também aquelas que, competindo por nichos semelhantes, o previnem. Ainda assim, as bactérias são organismos extremamente adaptáveis ao seu ambiente. Através da evolução experimental é possível monitorizar *in vivo* a sua adaptação ao ambiente intestinal e identificar e medir a importância relativa de certos mecanismos fisiológicos, a longo prazo.

Este trabalho, incidiu sobre a adaptação de *Escherichia coli* ao trato intestinal de ratos. A estirpe K-12 **MG1655** tem um genótipo bem caracterizado e constitui uma espécie com um metabolismo particularmente versátil, com vários exemplos de estirpes tanto patogénicas como comensais, sendo, portanto, um modelo de interesse neste ambiente. Após um mês de colonização *in vivo,* a base genética de mutações que presumivelmente teriam tornado a estirpe ancestral mais apta neste ambiente, foi identificada por sequenciação genómica. Neste trabalho, estendemos a caracterização das bases fisiológicas nos mutantes evoluídos e estabelecemos uma comparação entre a adaptação *in vivo* na presença e ausência de microbiota, de forma a explorar a importância relativa de certos mecanismos especificamente dependentes do ambiente intestinal a colonizar.

Caracterizámos a adaptação desta estirpe de *E. coli* na ausência de outros membros da microbiota como tendo seguido um trajeto evolutivo no sentido de aumentar a capacidade de utilização de aminoácidos. Dos vários mutantes que suportam esta hipótese, selecionámos a mutação mais abundante e mais vezes identificada entre as várias experiências independentes, a disrupção do gene ***lrp***. Através da comparação com uma estirpe mutante de referência mostrámos que inserções neste gene em mutantes evoluídos levou predominantemente à abolição da sua função, o que propusemos ter levado à sua vantagem seletiva: estas mutações levaram à perda de capacidade para crescer num meio em que a degradação de glicina, para o qual Lrp é ativador, seria necessária, visto a estirpe referência com *lrp* deletado também não crescer, ao contrário da estirpe ancestral. A partir deste fenótipo, desenvolvemos aqui um método *high-throughput* para a identificação de mutações no *lrp.* Através deste método, mostrámos que a cinética de adaptação desta mutação ocorreu posteriormente ao primeiro passo adaptativo já conhecido, com efeitos seletivos mais baixos e estabilizando a níveis intermédios na população. O padrão adaptativo que elucidamos é reminiscente do efeito de interferência clonal, em que mutantes *lrp* competem com outras mutações benéficas, limitando assim a fixação da população mutante *lrp,* embora seja importante caracterizar outros alvos adaptativos. A hipótese alternativa que propomos para explicar a estabilização a níveis intermédios da população mutante para o *lrp* seria que o nicho ocupado por estes é esgotado pela sua expansão, limitando assim a fixação deste alelo. Em concordância, observámos que quando uma estirpe *a priori* com *lrp* deletado foi introduzida *in vivo*, só foi selecionada após dez dias, o que coincide temporalmente com a expansão *de novo* e sugere que inicialmente a composição metabólica intestinal é alterada pela colonização, que posteriormente favorece mutantes com perda de função do *lrp.* Este gene está extensamente caracterizado no seu papel de regulador transcricional, particularmente para metabolismo de aminoácidos. De modo a encontrar um meio em que os mutantes *lrp* tivessem vantagem em relação à estirpe ancestral determinou-se o crescimento em vários meios. Através da análise das redes metabólicas reguladas pelo Lrp, previmos que especificamente para crescimento com os aminoácidos Alanina, Serina e Treonina, a perda de função do *lrp* fosse vantajosa. De facto, mostramos que na presença destes 3 substratos os mutantes *lrp* evoluídos têm vantagem seletiva,

embora não num perfil de aminoácidos mais diverso, apontando para a importância do perfil nutricional especifico, principalmente no que diz respeito às diferentes proporções de aminoacidos.

Na presença da microbiota, a análise funcional das mutações benéficas aponta para um trajeto adaptativo praticamente distinto do da mono colonização. Neste caso a maior parte das mutações apontam para um aumento da capacidade de utilização de açucares ou compostos semelhantes, o que pode advir da presença da microbiota. Na presença de uma microbiota abundante que também é capaz de consumir aminoácidos, é possível que as funções degradativas de outros membros da microbiota levem neste caso à disponibilidade de carbohidratos simples que *E. coli* consegue utilizar. A dependência pelo fornecimento de nutrientes por outras espécies, terá tornado o uso destes de uma forma energeticamente eficiente, crucial para o fitness de *E. coli.* Com esta hipótese, utilizámos duas mutações benéficas identificadas como correspondendo a genes funcionalmente relacionados com respiração anaeróbia para estudar a hipótese que o seu fitness seria mais elevado em condições anaeróbias. Dois mutantes com inserções na região regulatória dos operões ***dcuBfumB*** **e** ***focApflB***, relacionados com a produção e transporte de fumarato e formato, respetivamente, foram competidos *in vitro* contra a estirpe ancestral, utilizando um meio que se aproximasse da composição de carbohidratos preferenciais *in vivo.* Estava já descrito que estes mutantes teriam uma maior expressão transcricional de cada operão. Mostrámos que nestas condições, fenotipicamente, os mutantes evoluídos apresentam uma vantagem seletiva, que terá resultado de uma maior produção e respiração de fumarato e formato, respetivamente. Um aceitador de eletrões alternativo foi também testado com a hipótese de que teria efeitos que diferenciassem os dois mutantes.

Em suma, o presente trabalho reporta evidências de vários tipos de *landscapes* metabólicos que *E. coli* assume durante adaptação ao seu ambiente nativo. Quando é a primeira a colonizar, a ausência de degradação de fontes de carbono complexas favorece a utilização de aminoácidos. Na presença de microbiota, aminoácidos tornam-se limitantes e o fornecimento cruzado fornece carbohidratos simples, cuja respiração anaeróbia se torna essencial.

Acima de tudo, os estudos aqui referidos demonstram um grande paralelismo na adaptação da *E. coli* ao intestino. A reprodutibilidade evolutiva no ambiente complexo do intestino mostra que a evolução paralela não está restrita aos ambientes de laboratório, mas pode também ser comum em ambientes “naturais”, de vasta complexidade. Por si só, este facto é de salientar.

Keywords: Mamíferos; Microbiota; *E. coli*; Evolução; Metabolismo.

Tabela de conteúdos

1. Introduction 1

1.1. Mammalian Host – GI tract 1

1.1.1. Microbiota development 1

1.1.2. Colonisation resistance 2

1.1.3. Pathology and Enterobacteria 3

1.1.4. Pathogens and Metabolism 3

1.1.5. Metabolism of *E. coli* in the mammalian GI tract 4

1.1.6. Health promotion - control and opportunities of Microbes 6

1.2. Mechanisms of evolution 6

1.2.1. Adaptive patterns 7

1.3. Experimental evolution in the mouse GI tract 7

1.3.1. Two adaptive classes of targets in *E. coli* in a polymicrobial microbiota 8

1.3.2. One predominant class of targets in *E. coli* during mono colonisation 8

1.3.3. Lrp - the Leucine-responsive Regulatory Protein 9

1.3.4. *FocApflB* & *dcuBfumB* – Formate and Fumarate Respiration 10

1.3.5. *Gat* Operon & Galactitol utilisation 11

2. Protocols and Methodology 12

2.1. Media and growth conditions 12

2.2. Bacterial strains 12

2.3. Strain construction 12

2.4. Lrp phenotypic assay 12

2.5. Fitness measurements 13

2.6. Metabolic network analysis of Lrp 13

3. Results 14

3.1. A Simplified Community: Mono colonised mice 14

3.1.1. Characterization of *lrp* mutants isolated from mice mono-colonised with *E. coli* 14

3.1.2. Evolutionary dynamics of the two major targets in *E. coli* mono-colonised mice 16

3.1.3. Identification of physiological conditions conferring advantage to *lrp* mutants 18

3.1.4. Physiological basis of fitness advantage of *lrp* 20

3.2. Polymicrobial community: Antibiotic-treated Microbiota 22

3.2.1. *F*itness of *dcuBfumB* and *focApflB* mutants evolved in polymicrobial conditions 23

3.2.2. Physiological basis of fitness advantage of *dcuBfumB* & *focApflB* 24

4. Discussion 26

5. Bibliography 31

6. Annexes 34

6.1. Strains, plasmids and primers used 34

6.2. Supplementary Figures 35

Tabela de figuras

Figure 3.1. Construction of in-frame *lrp* deletion strain 14

Figure 3.2. Growth in glyMM media 16

Figure 3.3. Evolutionary dynamics in mono colonised mice 18

Figure 3.4. *In vivo* competition between *lrp* mutant and ancestral 19

Figure 3.5. Fitness of *lrp* mutants in the presence of aminoacids 21

Figure 3.6. Fitness of *dcuBfumB* and *focApflB* mutants evolved in a polymicrobial Microbiota 24

Figure 3.7. Effect of formate or terminal electron acceptors on fitness of evolved clones 25

SuppTable 6.1. Strains, plasmid and primer sequences used in this work 34

SuppFig. 6.1. Adaptive targets during adaptation to the mouse GI tract 35

SuppFig. 6.2. Genotype of *lrp*::IS isolated mutants 35

SuppFig. 6.3. Selective strength of *lrp* and *gat* mutations 35

SuppFig. 6.4. Metabolomic identification of aminoacid profiles in vivo 36

SuppFig. 6.5. Metabolomic identification of carbohydrate profiles in vivo 36

SuppDiagram.1 Digital (CD)

SuppDiagram.2 Digital (CD)

1. Introduction

1.1. Mammalian Host – GI tract

The gastro-intestinal tract harbors a complex microbiota, encompassing Archaea, Bacteria and Eukarya. Bacteria have been estimated to reach numbers up to 1000 billion individuals, outweighing host eukaryotic cells by a factor of 10:1 [1]. These estimates were based on the assumption that the more accessible samples analyzed represented the entirety of the GI tract. More recent estimates put the microbiota of a "reference human" around a 1:1 ratio, in relation to bacterial numbers[2]. Nonetheless, faster growth rates and lower individual spatial occupation, place bacteria as the predominant fraction of the microbiota. Trace levels of bacteria are already present in the placenta[3], and faecal contents of new-borns[4]. Still, gestational organs provide a remarkably isolated environment, ensuring a practically sterile individual, during gestation. Henceforth, a rise in microbiota abundance accompanies the development of the host towards an *adult* microbiota. Certain physiological characteristics and constraints shape the following dynamics of the microbiota towards identifiable structures.

1.1.1. Microbiota development

On the host side, immunoglobulins provided in the progenitor's milk can target known microbial molecular patterns, substituting the underdeveloped immune system of the host and before weaning, ingestion of the progenitor's milk serves as a continuous inoculation of *staphylococci*, *streptococci*, lactic acid bacteria and *bifidobacteria*, among others[5]. Additionally, increased permeability of early epithelial cells lining the GI tract permits an important diffusion of oxygen into the intestinal lumen. In agreement, aerobes and facultative anaerobes have been found to predominate in initial stages of development. The transition from a diet of milk to plant-derived nutrients during weaning represents an abrupt stepping stone that reflects on the microbiota, suggesting a deterministic nature of microbiota development. In humans[6], pigs[7] and mice[8] similar shifts in microbiota composition have been observed at this point, even though weaning occurs at different life times. At the functional level the bacterial metagenome becomes enriched for plant-derived carbohydrate degradation genes, in detriment of milk-derived oligosaccharides[7].

When an adult microbiota is established, *bacteroidetes* and *firmicutes* usually become the majority of the microbiota, while *proteobacteria* the phylum to which *E. coli* belongs, persist at lower levels[9,10], with core communities persisting for years[11,12]. At this stage, nutrient gradients and host immune activity are known to play a role in microbiota composition.

In the small intestine, nutritional availability is high in simple carbon sources like sugars[13] because of active host-secreted degradative enzymes and degradation in the stomach. Competition

with host absorption and low pH, limit bacterial numbers and metabolic styles, reflecting an enrichment for pathways for the import of simple sugars, compared to faecal samples[13].

In the caecum and large intestine, the highest bacterial densities are achieved due to lower transit time, in addition to overall more favorable conditions from decreased acidity and antimicrobials. Communities establish largely in the intestinal crypts and folds, thus escaping the flow of digesta. The mucus layer that lines the intestinal wall becomes crucial. This, network-like layer, largely consisting of mucin glycoproteins secreted by intestinal goblet cells, protects the mucosa and is regarded as key in the establishment of these communities[14]. At this point, luminal contents are predominantly in the form of complex carbon sources that resist upstream degradation (polysaccharide, oligosaccharide, proteins, peptide and glycoprotein[15]). Escaping expulsion through dense habitats established in the mucus, the microbiota thrives in the abundance of such compounds. Thus, mucus allows the physical retention of the microbiota, enabling slow fermentation of nutrients for which mammals lack the enzymatic repertoire. Reciprocally, the products of this activity, mainly short-chain fatty acids (SCFAs), become available for use by the host and link this symbiosis to energy homeostasis. Colonocytes derive ~60% of energy requirements from these fermentation products. In germ-free mice, maintained in isolated environments, the absence of microbes and thus SCFA production, leads to colonocyte starvation[16]. As a result, increased levels of several aminoacids, specially branch-chained aminoacids, are detected in these cells due to autophagy (degradation of cellular proteins as energy source). Addition of butyrate (a product of fermentation) is sufficient to recover homeostasis as in mice with a conventional microbiota, as is the introduction of a single butyrate-producing species[17].

In mammals, the presence of anaerobic fermenting bacteria are thus believed to be essential to host health but also for the establishment of stable ecological networks where the degradation products of metabolism can feed other species that lack the ability for complex degradation of mucus and fibers[18].

1.1.2. **Colonisation resistance**

Even though the structure and function of the microbiota is remarkably different between developmental stages of the host but also in host niche diversity[18], these interactions are assumed to be strongly dependent on past states. That is, the resident community is believed to constrain the developmental path and contribute to a stable adult microbiota. The indefinite persistence of a particular population without reintroduction implies replication. Any particular assembly of organisms or ecological function that prevails, has the potential of continuously adapting to this environment, shaping not only the resident microbiota but also invading bacteria and novel ecological interactions. The contribution of the indigenous intestinal microbiota in resisting colonisation was initially

recognized in the 1950s when Bohnhoff et al. demonstrated that disruption of the established ecological networks by antibiotic treatment lowered the dose for *Salmonella* infection by 100 000-fold[19].

1.1.3. **Pathology and Enterobacteria**

Although the microbiota of the GI tract is generally referred to as commensal, studies show that composition and diversity is crucial to both host health and disease. Evidence for the importance of the microbiota in many diseases of enteric infection, include: immune-mediated and metabolic diseases such as inflammatory bowel disease[20], obesity[21], asthma and atopy[22]. Three major enteric pathogens include *Escherichia coli, Salmonella Typhimurium,* and *Clostridium difficile. E. coli* is a common commensal inhabitant of the mammalian intestinal tract that can develop pathogenic behaviors by genetically acquiring pathogenicity islands and plasmids[23]; *S. enterica* is a group of facultative intracellular pathogens that, in addition to an intestinal coloniser, can also invade host tissues and cause disease ranging from mild gastroenteritis to acute systemic infection (typhoid fever); *Clostridium difficile*, unlike *S. enterica* or pathogenic *E. coli*, is often present in the microbiota of healthy individuals. At high colonisation levels, however, it leads to a condition known as C. difficile–associated diarrhea (CDAD) and is the main cause of pseudomembranous colitis, a form of antibiotic associated diarrhea[23].

Because of this functional heterogeneity, commensal and pathogenic subpopulations have likely co-existed for a long time, leading to strategies that oppose each of the behaviors. If it is assumed that the large majority of the microbiota is at any point in time under a state of commensalism then it becomes clear why pathogens have developed such complex mechanisms. These uncommon behaviors become necessary in order to invade niches that are already occupied by overwhelming numbers of adapted resident bacteria. These residents have developed efficient ways to utilise available nutrients as well as active mechanisms to protect their environment against competing bacterial species[24].

1.1.4. **Pathogens and Metabolism**

To overcome resistance to colonisation, a similar process is used by many mucosal pathogens. As in the case of pathogenic *E. coli,* a multi-step process of pathogenesis consists of colonisation, evasion of host defenses, replication and host damage[25]. This can be achieved by triggering inflammation to disrupt the stability of commensal communities, through the use of *traditional* virulence factors (toxins, adhesins, and invasins) or by taking advantage of antibiotic use which opens niche space[19]. Thus, it is important to understand colonisation for its basic principles allowing the first step of this process of pathogenesis, as well as of directly competing, commensal populations. Linking

commensal control on pathogenesis, emerging studies[26] show the relationship between metabolism and the *traditional* virulence factors that directly cause disease, by proposing that the ability to grow in and thus colonise the GI tract is a basal requirement. This ability then establishes the grounds for the development of pathogenicity, in populations with such genetic potential.

Facultative anaerobes like *E. coli* are known to scavenge oxygen, an important trait revealed by the colonisation defects of an *E. coli* mutant for the high-affinity oxygen reductase but not for a mutant of the low-affinity homolog, which is better used under ample oxygen availability[27]. While a cooperative trait from the perspective of oxygen-sensitive anaerobes (the majority of the microbiota), this is also a level of resistance that commensal strains exert on pathogens, as the latter are mostly aerobes or facultative anaerobes. Still, *Salmonella* is a good example of a species that overcomes this, supplementing its ability to respire by creating a second respiratory electron acceptor[28]. When *Salmonella* induces inflammation, the host generates ROS that oxidises thiosulfate, a product of other bacteria[29], to tetrathionate, which *Salmonella* can respire. This pro-inflammatory trait succeeds at specifically increasing *Salmonella*'s competitiveness, as the ability to use tetrathionate is likely restricted in bacteria that lack this gene cluster (*ttr*) and have evolved to thrive in conditions where ROS production is not triggered. The crucial observation is that metabolic and inflammatory functions, as in this example, are often carried in pathogenicity islands of the same mobile genetic element[28], recently acquired in evolution.

Other times it is the loss of genes that mediates adaptation to pathological environments. For example, *Pseudomonas aeruginosa*, the predominant microbe in the cystic fibrosis lung, loses *lasR*, this species' major quorum-sensing transcriptional regulator of both virulence and metabolism. *LasR* mutants can grow better on aminoacids, presumably accessible in the mucus, but have decreased virulence[30]. It is possible that when a pathogen establishes itself into a niche where nutritional competence brings more benefits than virulence, pathogenic behaviors may no longer be of selective advantage. Some previous work has suggested a similar scenario during adaptation of *E. coli*[31].

Examples like these blur the line between pathogenic- and commensal-specific behaviors and support the understanding of metabolic traits as important fitness factors in colonisation and/or infection, as important as classic virulence genes.

1.1.5. **Metabolism of *E. coli* in the mammalian GI tract**

E. coli is a ubiquitous member of the intestinal microbiota as the most abundant facultative anaerobe, it is arguably the best understood of all model species and covers a wide spectrum of lifestyles. *E. coli* is thus an interesting model to study the intestinal environment. This bacterial group is likely one of the most ancient members of the microbiota of mammals as the earliest colonisers of the GI tract. These populations likely cause strong founder effects on the microbiome and perhaps

other immune, developmental or metabolic aspects affecting susceptibility to disease, regardless of their pathogenicity.

An early ecological framework, now postulated as *Freter's nutrient-niche hypothesis*, states that competition for nutrients is the major determinant of success in the GI tract, commonly stated as: *For a species to be successful in the intestine, it's growth rate must outweigh the washout rate supported by the use of one or a small number of growth-limiting nutrients, better than all other competitors*[32,33]. *In vitro* studies have shown that metabolism of *E. coli* is limited to monosaccharides, di- and tri- saccharides and maltodextrins. There are several glycoside hydrolases in the genome of *E. coli* MG1655 (Commensal) and *E. coli* EDL933 (Pathogen), none of which are secreted nor hydrolyse complex polysaccharides[34,35]. However, these simple dietary sugars from host intake are limiting in the lower intestine due to absorption by the host. To grow in the lower intestine, *E. coli* must acquire usable carbon sources, other than directly from host sugar intake. It has been suggested that close to the intestinal mucosa, bacteria live in communities of biofilms. The mucus layer is in itself structurally similar to a biofilm and many species that create biofilms *in vitro* have been shown to adhere to mucus ligands. Indeed, community profiles have been distinguished between mucosal and faecal samples[36], indicating that some groups have the ability to persist by increasing retention in the large intestine. *E. coli* is found in these mixed communities in the mucus layer[37], whereas its growth is limited in the lumen[38]. Therefore, to colonise the large intestine, it is generally assumed that *E. coli* depends on anaerobes that can degrade oligo- and polysaccharides (celluloses, hemicelluloses or starch) to provide the mono- and di-saccharides and maltodextrins it needs for growth.

An extension to Freter's framework is the *Restaurant hypothesis*, later proposed to account for this increasing evidence that the GI tract is far from the idealized, continuously flowing, perfectly mixed conditions assumed in Freter's models, where all microrganisms were physically exposed to an average nutrient profile. Whereas, *in vitro,* both commensal and pathogenic strains of *E. coli* metabolise several sugars in the same order[39], *in vivo,* only a smaller set of those sugars appears necessary for maximum fitness and moreover, this preference is different between the commensal and pathogen[39]. Together with experiments where pathogen clearence can be achieved when the commensal and pathogenic *E. coli* strains have the same substrate preference *in vivo,* the nutrient-niche hypothesis continues to be supported, being that the better use of limiting nutrients not used by other microbes can support the colonisation of both strains of the same species. Furthermore, pathogenic and commensal *E. coli* strains appear to have a similar genetic composition of carbon-substrate hydrolases[40] and *in vitro* catabolic potential[39,41,42]. Interestingly, both exhibit distinct feeding profiles only *in vivo,* suggesting that the regulatory architecture and/or the metabolic environments that each bacterium can inhabit largely control the phenotypic behaviors ranging from commensality to pathogenicity.

1.1.6. **Health promotion - control and opportunities of Microbes**

For a long time, microbes have been generally regarded as harmful and undesirable. From this idea stemmed the widespread use of antibiotics. With recent advances in the forces behind the extremely proliferative capacity of bacteria, the possibility of favoring particular communities[43,44] has created a novel paradigm for the way health is promoted. In one hand, specific treatments can be tailored by targeting common mechanisms (serotypes, nutritional changes) prevalent in specific medical conditions, which can avoid unwanted effects of antibiotic use (breaking colonisation resistance, opening alternative niches, resistance to antibiotics). Even in situations where pathogens are not the principal etiology, their management and/or use as therapeutics could be advantageous[45]. Often, the identification of traits essential for colonisation or infection are inconsistent, due to differences in experimental conditions[46] and underline that not all conclusions may prevail, in light of evolution. More recent approaches seek to ask the relative long-term importance of these traits by designing experiments to observe the role that evolutionary forces play in the phenotypic landscapes of the microbiome[37,47,48]. These create the possibility to find and target prevalent mechanisms that are stable through time and not strain specific, supporting the likelihood that any unknown population in the GI tract of a mammal, with the potential to exist constantly evolving, would follow such behaviors. Key traits could then be disfavored or, vice versa, favored to capitalize on the known ability of native communities to inhibit pathogenic behaviors of strains with such genetic potential[49,50].

1.2. **Mechanisms of evolution**

It is established by theoretical and *in vitro* studies that to endure in a particular environment, the frequency at which new mutations arise in natural populations and their effects on fitness are crucial. There are several well-known evolutionary mechanisms that can explain adaptation[47,51,52]: Mechanisms like mutations, can occur in several ways: single nucleotide changes, deletions or insertions, by duplications or translocations, or by the movement of transposable elements, such as insertion sequences. The latter are scattered in bacterial genomes and have the capacity to mobilize within a bacteria's chromosome. Although common, these genetic elements often encode nothing more than their own mobility but can have large effects on adjacent genes, depending on which region these insert in (loss/gain/decrease in expression). Transposable elements are thus thought to be major players in evolution and were indeed predominant in experiments described in this work. Another important mechanism is genetic drift. It is characterized as the random sampling of alleles from one generation to the next, which influences the likelihood of selection of a particular mutation. The exchange of genetic information between different strains (recombination) and natural selection are also important drivers in evolution.

1.2.1. **Adaptive patterns**

Selection can drive several canonical signatures of evolutionary dynamics in asexual organisms, as in bacteria used here[47,53]. In strong selection (hard sweeps), when no overlap occurs in the expansion of mutations, a strong mutation can reach genetic fixation. If such pattern is maintained a second strong mutation would, in turn emerge and overtake its ancestral. In periodic selection, such mutation is not as strong (depending on rate of appearance, fitness effects and population size) but will nonetheless, outcompete all other beneficial mutations segregating, effectively eliminating other alleles. In clonal interference however, a pattern in which this mutation is not particularly strong and the rate of beneficial mutations is high, other mutations have the room to expand from another lineage. In asexual populations this means that such mutants will compete between them, thus decreasing the probability of either one fixating (soft sweeps). These dynamics can be driven by within and between species ecological interactions like the ones that can be anticipated to occur in the GI tract. These interactions can lead to negative frequency dependent selection. Negative frequency-dependent selection translates in high fitness advantage of mutants when at low frequency, but lower or even deleterious when at high frequencies, which can favor maintenance of genetic diversity[54].

The evolutionary mechanisms presented above have been mostly described for evolutionary experiments *in vitro* and are less characterized in the GI tract.

1.3. **Experimental evolution in the mouse GI tract**

Recent work used a genomic approach to explore how *E.coli* adapts to its native environment, the GI tract[47]. Our work is built upon such studies, detailed below. The streptomycin-treated mouse model was used as the closest intestinal environment in which adaptation can easily be studied. Treatment with this antibiotic (streptomycin) clears a physical niche that was occupied by other microbes, eliminating competitors of *E. coli* and effectively breaking resistance to colonisation[47]. The streptomycin-treated mouse intestine is thus a complex environment where *E. coli* is subjected to multiple selective pressures, including the polymicrobial community. In this model, a genetically homogeneous population of *E. coli*, except for chromosomally encoded fluorescence markers, was used to trace the occurrence of different adaptive mutations. The population introduced was composed of equal amounts of two subpopulations expressing either a yellow (YFP) or cyan (CFP) fluorescent protein, to follow adaptation through the divergence of marker frequencies. To then identify a broad range of putative beneficial mutations, large *E. coli* populations were isolated from each mouse, from faecal samples after 23 days of colonisation as at this day, marker divergence had stabilised indicating absence of further adaptive events. Genome sequencing of these sampled populations enables a view of the genetic composition of the evolving population. This approach does not characterise how mutations segregate in time nor explores their linkage, but can identify mutations segregating at high frequency, and thus, likely to be beneficial. A criteria of mutations present both at >5% and in more

than one mouse was generally used to define parallel mutations. Parallel mutations are indicative of selection and can be subsequently targeted. The parallel targets observed in 15 independently evolving populations are reproduced in SuppFig.6.1[47]. A similar experiment was performed in previously germ-free mice, to study the adaptation of the same ancestral *E. coli* strain, in the same host, though in microbial isolation (mono colonisation). These mice are maintained in sterile conditions from birth and throughout the experiment only have contact with *E. coli*. The parallel targets observed in 8 of the independently evolving populations in monocolonisation are shown in (SuppFig.6.1, Dias, J unpublished). Having identified the mutational profile during adaptation, specific genotypes can be characterized. It is possible to measure their fitness effects and compare the selective advantage depending on the environment on which they emerge and comparing adaptation between mono colonisation and in a polymicrobial microbiota (streptomycin-treated mouse model), to determine if beneficial mutations are microbiota-independent, emerging in the presence and absence of other members of the microbiota, or, if they are microbiota-dependent, that is, adaptive targets selected on either the absence or presence of other species.

1.3.1. **Two adaptive classes of targets in *E. coli* in a polymicrobial microbiota**

Adaptation of *E. coli* in polymicrobial conditions yielded a total of 11 genes classified as parallel targets[47] (SuppFig.6.1). Analysis of their ontology indicates two main classes of metabolic functions: carbon source utilization and anaerobic respiration. In Carbon Source utilization: the ***gat*** operon (codes for enzymatic transport and degradation of the sugar-alcohol Galatitol), the ***srlR*** gene (the repressor of the operon for the transport and degradation of the sugar-alcohol Sorbitol), the ***kdgR*** gene (the repressor of the operon for the transport and degradation of the sugar Gluconate). In Anaerobic Respiration: the ***dcuBfumB*** operon (codes for enzymatic transport and production of the terminal electron acceptor Fumarate) and the ***focApflB*** operon (codes for enzymatic transport and production of the electron donor Formate).

1.3.2. **One predominant class of targets in *E. coli* during mono colonisation**

During adaptation of *E. coli* in mono colonised mice, 6 targets were identified as parallel, at the same time point as above (SuppFig.6.1). Similarly, by analyzing the ontology of these genes/operons, it is possible to find one predominant class of mutations, in terms of metabolic function. A unifying theme is again genes related to Carbon source utilisation: the ***gat*** operon (codes for enzymatic transport and degradation of the sugar-alcohol Galatitol), the gene ***alaA*** (codes for the main glutamate-pyruvate aminotransferase), the ***dtpB*** gene (codes for a dipeptide/tripeptide:H+ symporter), the ***frlR*** gene (codes for the repressor of the operon for fructose-lysine transport and degradation) and the ***lrp*** gene (codes for a Master regulator of aminoacid metabolism).

By comparison, the only target that is common to both conditions was the *gat* operon, indicating that this mutation is specific to colonisation of the GI tract (microbiota-independent). Also, the mutational targets observed are very context-dependent. The functional nature of these events, suggest that these two environments are quite distinct and that *E. coli* adjust accordingly: the majority of adaptive events are related to the acquisition and utilisation of carbon sources but this was microbiota-dependent. Targets specific to a polymicrobial environment and related to carbon utilisation were limited to sugar or sugar-like metabolites, whereas in mono colonised conditions, to aminoacid compounds. Additionally, targets linked to respiratory functions (***dcuBfumB*** and ***focApflB***), was only observed in the presence of a microbiota and not when *E. coli* is the sole coloniser.

From these, 3 targets were chosen for further characterization: ***lrp, dcuBfumB*** and ***focApflB***.

1.3.3. **Lrp - the Leucine-responsive Regulatory Protein**

In SuppFig.6.1 can be seen that the gene coding for Leucine-responsive regulatory protein (*lrp*) was targeted in adapting populations of *E. coli*. All 10 mice mono-colonised had mutations in *lrp* at >5%, predominantly with insertions in the coding and the upstream intergenic regions.

Lrp is one of 7 Master regulators that together directly control 50% of all genes in *E. coli* [55]. This DNA-binding protein activates transcription of genes involved in aminoacid synthesis and inhibits transcription of genes involved in degradation of aminoacids, in general. Thus, this regulatory structure allows the levels of Lrp within a cell to act as a controller for the allocation of resources between endogenous production, relative to dependence on exogenous aminoacids. Conceptually, an increase in Lrp increases inhibition of degradation while favoring the cells own machinery to meet its aminoacid requirements, by increasing activation synthesis. Control of Lrp levels occurs by means of autoregulation by direct repression of its own transcription, transcriptional repression by H-NS, NsrR and activation by GadE and post-transcriptionally by the sRNAs MicF and GcvB[56]. The result dynamics of the cellular levels of Lrp are such that during steady-state exponential growth, as cells are actively dividing and so, aminoacid requirements for protein synthesis is high, active levels of Lrp are high[56], promoting aminoacid availability through synthesis. As growth slows down, cell division is no longer being supported. In these conditions, Lrp levels decrease, decreasing synthesis of aminoacids, no longer at high demand, while relieving inhibition of uptake and degradative functions that are sufficient to meet this lower demand or acting as an alternative nutrient source, when available[56]. These responses by *E. coli* can be seen as gradual changes in transcriptional regulation, as Lrp levels change. In contrast, the activity of Lrp on a subset of genes can also be modulated by quick responses, through the effect of the co-factor Leucine. By binding Lrp, Leucine can potentiate, lower or be neutral to the activity of Lrp on a subset of genes. This creates an additional layer of possible regulatory outcomes with effects on a shorter time-scale. For example, in the case of an upshift in nutritional availability that also contains aminoacids, the activity of Lrp, which would otherwise

promote synthesis over their degradation, can be quickly inverted by Leucine towards using aminoacids as an energy source. In combination, the levels of Lrp and the signalling effect of Leucine assist in modulation of nutritional behaviour according to environmental conditions and the partitioning of aminoacids between energy source and protein synthesis[57].

Other genes regulated by Lrp include, for example the inhibition of oligopeptide permease (*opp* operon) which can transport short, non-specific peptides; the outer-membrane porins that mediate control of osmolarity: *OmpC* is favoured in high osmolarity and is inhibited by *lrp* whereas *ompF* is favoured under low osmolarity and indirectly controlled by *lrp* through the sRNA *micF*; and the *liv* class of transporters specific to branch-chained aminoacids (isoleucine valine and leucine) transcription of which is inhibited by *lrp*. Many other genes are directly or indirectly regulated by *lrp* and have been characterised to different extents[58].

1.3.4. *FocApflB* & *dcuBfumB* – Formate and Fumarate Respiration

In the cases of the *dcuBfumB* and *focApflB* targets, in 7 and 4 mice respectively, mutations were identified at >5%. Interestingly, these targets were never seen together in the same genotype and when observed, both were rarely selected in the same mouse, simultaneously (1 out of 9), indicating these are antagonistically related. The structure, ontology, and function of the two targets are seemingly very similar. Both mutations were caused by insertions in only the intergenic region; both consist of a transporter upstream of a biosynthetic gene[59–61]; and both are known to be positively controlled by the regulator Fnr[62]. The *dcuBfumB* operon, codes for a C4-dicarboxylate transporter (*dcuB*) and a fumarate-producing enzyme (*fumB*). The other, in the *focApflB* operon codes for a formate transporter (*focA*) and a formate-producing enzyme (*pflB*). Both are known to be predominantly expressed in the absence of oxygen[62] where they assist in carrying out respiratory functions.

Extensive work *in vitro* has revealed the metabolic and regulatory functions that *E. coli* coordinates in order to adjust to a wide variety of environments, coordinating energy production with respiration. A wide variety of substrates can be degraded to release energy to electron-carriers like NAD+ and be stored in their reduced forms (NADH, for example). In respiration, this first step of exchange can often be mediated by a multitude of dehydrogenases with specificity for individual, or groups of substrates and, without it, growth would not be possible. In the case of formate, an electron-rich molecule, three different dehydrogenases can be utilised. Fdh-H, Fdh-N and Fdh-O. Fdh-O is somewhat constitutive, Fdh-H is used in the absence of electron acceptors and Fdh-N is used when Nitrate is available[63].

Fumarate on the other hand, is more oxidised and serves as an electron sink in a later step. Some dehydrogenases may directly develop the conditions necessary for energy production by producing a net increase in protons extracellularly, forming the necessary gradient for the activity of ATP synthase. A second step can take place, coupling the electrons stored from the first step (NADH

for example) to an electron sink, which produces no energy, or to a second proton gradient-forming reaction. Importantly, this final step can only occur with transfer of electrons and so, is dependent on availability of a suitable final electron acceptor molecule. If the final step does not occur, not only is ATP production less efficient but also intermediaries like NADH accumulate, while NAD+ is depleted, which limits growth. This clearly demands the regulatory prowess to not only adjust to fluctuating nutritional availability but also to compute optimal electron donor/acceptor arrangements. Oxygen is prefered[27] but in the intestine a quick shift towards different respiratory functions should be crucial to maintain competitiveness[64]. Such alternatives include fumarate and nitrate respiration.

1.3.5. *Gat* Operon & Galactitol utilisation

Galactitol utilisation operon (*gat*YZABCDR) is systematically knocked out as the initial step of adaptation to mice[47]. This is evidenced by the temporal dynamics followed via population phenotypes, as gat-negative mutants can easily be identified by the use of MacConkey medium supplemented with galactitol which shows white colonies, instead of red, when galactitol is not consumed[47]. The galactitol utilisation operon codes for catabolism of one of the three naturally occurring hexitols: galactitol[65], a product of the reduction of the sugar galactose[66]. An insertion in *gatR*, the repressor for the entire operon, results in its unregulation with respect to galactitol availability. This known polymorphism is present in the strain used here, partially explaining this mutation. Galactitol is likely not available at conducive levels[65]. Furthermore, following the activity of *gat*YZD on internalized galactitol, the end-product tagatose-6-phosphate still requires several extra enzymatic steps to enter glycolysis as D-glyceraldehde-3-phosphate and dihydroxy-acetone-phosphate. This relative inefficiency likely adds to the demotion of galactitol, when compared to the remaining hexitols, sorbitol and mannitol, which are readily converted to the glycolytic intermediates D-fructose 6-phosphate and β-D-fructofuranose 6-phosphate, respectively[67]. Although the mechanisms causing the observed adaptive processes around this complex trait were not further developed here, besides comparison of the initial adaptive step in mono-colonised mice, it is clearly a strong model to study the ecological and evolutionary pressures E.coli faces in the mammalian gut[47,68].

The present work aimed to characterize the unknown molecular basis of adaptive steps providing an advantage to evolved *E. coli* mutants; and, by mutual exclusion, the environmental differences causing their expansion in the GI tract of mice colonised with a single population of *E. coli* or in the presence of polymicrobial communities. Specifically, this work focused on 3 targets: *lrp*, *dcuBfumB* and *focApflB*, chosen for being the most frequent and parallel mutations that could represent good candidates of high fitness peaks of *E. coli* in the environment of its highest abundance. A molecular explanation for the expansion of *lrp, dcuBfumB* and *focApflB* was explored here.

2. Protocols and Methodology

2.1. Media and growth conditions

Lysogeny Broth was used for general strain handling & construction unless stated otherwise. Antibiotics were included at the following concentrations, when necessary: Streptomycin µg/mL, Kanamycin µg/mL, Chloramphenicol µg/mL. Composition and preparation of M9 glucose minimal medium, differential MacConkey + sugar medium as described before[47] and selective glyMM medium is detailed in 5.2. Supplemented aminoacids and sugars were sterilized by filtration (Milipore® 0.22µm filters) and included as indicated. Anaerobic media always included 5mg/L $MnSO_4.4H_2O$, 0.125mg/L $FeSO_4.7H_2O$, 1µM NaSeO, 1 µM $NaMoO$[69] and 4µM Resazurin. To test for sugar utilisation, MacConkey plates were always incubated at 30°C.

2.2. Bacterial strains

Bacterial strains used in the present study are listed in SuppTable 6.1. All strains used were derived from Escherichia coli K-12, strain MG1655[47]. The original K-12 was obtained from a faecal sample of a recovering diphtheria patient in Palo Alto, CA, in 1922. Strains DM08-YFP and DM09-CFP (MG1655, galK::YFP/CFP amp^R, str^R (rpsl150), Δ*lacIZYA*) were used in the first evolution experiment[47] and germ-free evolution experiment. These strains contain the yellow (YFP) or cyan (CFP) fluorescent genes linked to amp^R in the *galK* locus under the control of a *lac* promoter and were obtained by P1 transduction from previously constructed strains[70].

2.3. Strain construction

The allelic-exchange method for homologous recombination of linear DNA fragments, described by Datsenko and Wanner[71] was used to construct the *lrp* deletion strains. Briefly, amplification of pKD3's chloramphenicol cassette by PCR with primers with homology for FRT region of pkD3 and the extremities of the *lrp* open reading frame with primer TCAGACAGGAGTAGGGAAGGAATACAGAGAGACAATAATAATGGTAGATAGTGT AGGCTGGAGCTGCTTC (forward) and TGTAATCAAAATACGCCGATTTTGCACCTGTTCCGTGTTAGC GCGTCTTACATATGAATATCCTCCTTAGT (reverse) yielded a linear fragment, which was purified by digestion of methylated bacterial DNA with DpnI restriction enzyme. This fragment was introduced in the chromosome of a clean, pKD46-bearing MG1655 (KX1086) by electroporation. Increased recombination events with linear DNA fragments is promoted through λ Red induction with arabinose and resultant, recombined transformant selection in LB chl plates (MG1655 *lrp*::cm^R). After loss of the temperature-sensitive KD46 plasmid by growth at 42°C and confirmation by PCR (fig.3.1), this allele (*lrp*::cm^R) was transduced to the ancestrals of the evolution experiment (DM08, DM09) by bacteriophage P1 transduction. The cassette was then excised after electroporation and selection of pTL17 in LB kan. Finally, ptL17 was lost by short growth in M9+glucose and strain construction confirmed by absence of growth on LB kan nor LB chl and PCR (DM08/09 Δ*lrp*). The remaining constructs originate from the Keio collection of mutants. Briefly, a lysate of each mutant was generated by infection with a lysate of a clean MG1655 and the resulting P1 bacteriophage particles were in turn used to infect a culture of the recipient strain. After selection of transformants with only the kanamycin resistance cassette and phage loss by successive growth in LB kan, the resistance marker is removed by electroporation with pCP20 and growth at 30°C in LB chl. Finally, pCP20 was eliminated by growth at 42°C, confirmed by PCR and absence of growth on LB chl nor LB kan.

2.4. *Lrp* phenotypic assay

For high throughput *lrp* mutant screening, developed and optimised in this work, an aliquot from each frozen faecal sample was diluted and plated for single colonies on non-selective LB agar plates to ensure growth. 96 colonies per sample were then each picked into individual wells in 96-well plates (Corning® Costar®), with M9 medium (no carbon or nitrogen source) and subsequently picked onto glyMM agar plates (same composition of

M9 glucose except for Glycine 200µg/mL in place of NH_4Cl) and MacConkey+Galactitol(1%) agar plates, being a rich medium, also served as a positive growth control, in addition to identifying the *gat* phenotype. Each plate also contained the ancestral (DM09), Δ*lrp* (MFP11), *lrp*::IS_{COD} (MFP88) and *lrp*::IS_{INT} (MFP89) strains as reference phenotypes (fig.3.2). Both sets of petri plates were then incubated for 48 and 24 hours, at 42°C and 30°C, respectively, at which point colony size (glyMM) and indicator dye color (MacConkey) were assessed. Colonies in the glyMM plates with, at least WT size were scored as ***lrp***$^+$, no growth scored as ***lrp***$^-$ (loss-of-function) and smaller colonies than WT (as (3) in fig.3.2b) were always confirmed by PCR of *lrp* and only when ISs were found in the intergenic region, scored as *lrp*::IS_{INT}.

2.5. Fitness measurements

For *in vitro* co-cultures, cultures were grown overnight in M9+Glycerol 0.5%, harvested and washed at least twice with M9 with no carbon or nitrogen source, diluted and mixed (1:1, ~5E5 cells/mL) into 200µL of media in 96-well plates (Corning® Costar®) incubated in an anaerobic chamber under a N/CO_2 atmosphere, at 37°C with no shaking. Fitness calculated as selective coefficient (per generation)[72]. For monoculture fitness measurements, the same preparations were followed as above. 150µL cultures (~1M cells) were incubated in a Bioscreen® microplate reader (Fast&High shaking settings) and growth at 37°C was followed by measuring optical density at 600nm, registered at 30' intervals (*Bioscreen*®). Maximum growth rate during steady-state growth was determined (*GrowthRates*®) and relative fitness is shown as maximum growth rate of each *lrp* mutant strain relative to that of the ancestral. Mouse handling in the competitive experiments done here, were carried out as before[47].

2.6. Metabolic network analysis of Lrp

All specific functional metabolic categories gene IDs belonging to the Lrp regulon (as accessed on *Ecocyc.org*® up to 30/09/16) were introduced into a *smartable* in *Ecocyc*® followed by enrichment of this gene set for pathways, by querying if any subset of this set of genes are any more likely to belong to any specific pathway than would be expected to occur by chance (Fisher exact with no correction, $p<0.05$). The resulting Pathway IDs have at least one gene from the *lrp* regulon and are mostly related to aminoacid use, as expected[57]. To then visualise what possible differential regulatory effects occur between WT and *lrp* mutants, all complete pathways annotated under either the class *Aminoacid Degradation* or *Aminoacid Biosynthesis* were joined by connecting common metabolites and subsequently joined with the Tricarboxylic acid cycle. Then, reactions whose genes are regulated by Lrp were coloured according to manually predicted qualitative differences between strains: Green identifies reactions that are likely to be increased, red indicates reactions decreased in an *lrp* mutant and black shows reactions not directly regulated by Lrp (it was assumed that changes in transcription would affect the flux through the pathways). These pathways can depend, at least quantitatively, on the presence of the regulatory effects of Leucine. The metabolic networks shown, assume its presence because it is identified in mice, by NMR analysis (SuppFig.6.4). SuppDiagrams 1 and 2 are annexed (CD) as .graph files for visualization in Ecocyc®, as follows:

1. Access https://ecocyc.org/cytoscape-js/ovsubset.html?empty=T or alternatively https://ecocyc.org/, followed by Pathway collages under the term Metabolism and selecting "Invoke pathway collage viewer";
2. Load a .graph file by selecting "File" -> "Load collage from file" -> "Choose file" and access either the SuppDiagram1 or SuppDiagram2 supplied in digital format (CD);
3. Navigate to "View" -> "Customize global properties" -> "Pathways" and Uncheck "Show pathway background boxes" to facilitate navigation by dragging the background.

3. Results

3.1. A Simplified Community: Mono colonised mice

3.1.1. Characterization of *lrp* mutants isolated from mice mono-colonised with *E. coli*

During the adaptation of *E. coli* to the GI tract of mono-colonised mice, insertions in the *lrp* gene or in the intergenic region upstream of this gene were present at frequencies higher than 5% in all populations sequenced. To understand the functional role of the *lrp* insertion mutants isolated during evolution experiments, an *lrp* deletion mutant (Δ*lrp*) was constructed by homologous recombination[71]. All construction steps were confirmed by PCR using extragenic primers flanking the *lrp* gene (fig.3.1). The size of the PCR fragments obtained was compared with the ancestral strains (fig.3.1. MG1655 (1), DM08 (3), DM09 (4)), which have an intact *lrp* gene. The resulting strains containing the in-frame *lrp* deletion after excision of the resistance cassette (lanes 5 and 6) were named MFP08 (YFP marker) and MFP11 (CFP marker) (fig.3.1 lanes 7 and 8, respectively).

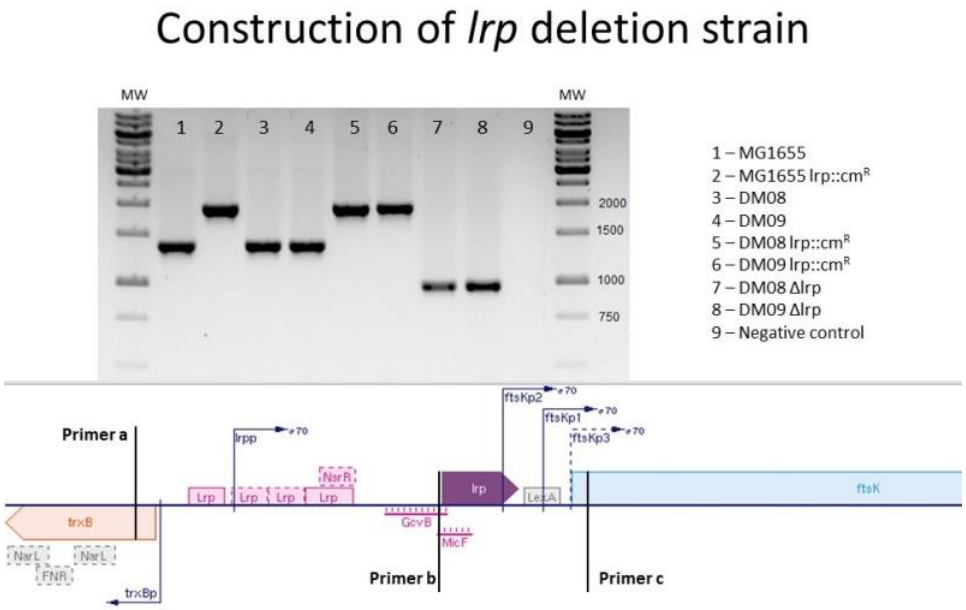


Figure 3.1. Construction of in-frame *lrp* deletion strain. Confirmation of in-frame deletion of *lrp* (purple), by *Wanner* method[71]. Chloramphenicol resistance cassette was recombined with the open reading frame of *lrp* (lane 2), compared with ancestral (lane 1) and transduced to the fluorescently marked DM08-YFP and DM09-CFP ancestral strains (lanes 3 and 4) by P1 transduction. The resulting Δ*lrp*::cm^R alleles on the ancestral background (lanes 5 and 6) were subsequently excised using the pTL17 plasmid (lanes 7 and 8). Total amplification with primer ***a*** (forward) and primer ***c*** (reverse). Lane 9 is the same PCR without template DNA as negative control.

The Δ*lrp* MFP08 and MFP11 strains were used to determine the phenotype of the *lrp* deletion in comparison with the ancestral strain and two evolved clones with insertion elements in either the coding region of *lrp* (MFP88 *lrp*::IS_{COD},) or the upstream intergenic region (MFP89 *lrp*::IS_{INT}). These were isolated from the evolution experiment at early time points, to lower the chances of carrying other mutations. The genotype, regarding the *lrp* gene was determined by PCR. Using primer pair (a) and (c) (within the genes flanking the *lrp* gene) in combination with primer (b) (reverse primer for the start codon of *lrp*), it is possible to confirm the region of insertion (SuppFig.6.2). These strains were used for further characterization, as representatives of the two different types of insertions identified in the population sequence results that could potentially affect Lrp function.

According to previous literature, *lrp* loss of function mutants should grow as well as the WT in rich medium (like LB) at 37°C but can be distinguished from the WT because the growth of an *lrp* mutant is impaired at 42°C in minimal media with glucose and NH3 as sole carbon and nitrogen sources[73]. Growth was tested with Δ*lrp* and ancestral, under those conditions. No growth defect was observed in LB at 37 and 42°C, and indeed the Δ*lrp* mutant had a lower growth yield in comparison to the ancestral in glucose M9 medium, particularly at 42°C (data not shown). However, the difference achieved was small and deemed insufficient for distinction.

To further characterize the phenotype of the *lrp* mutants and obtain a phenotypic assay where the difference between the WT and the reference Δ*lrp* mutant was more pronounced, we took advantage of the fact that Lrp is an important transcriptional regulator involved in regulating nitrogen metabolism and is known to be required under conditions of low nitrogen availability. This regulator is required for the activation of enzymes involved in the assimilation of NH3 at low concentrations. Specifically, *lrp* is required to activate transcription of the high affinity Glutamate synthase (*gltBD*) and is also known to be a transcriptional activator of the glycine cleavage operon (*gcvTHP*)[74]. The latter operon is associated with NH3 assimilation as it encodes for enzymes required for a cycle in which glycine is degraded, liberating NH3[74]. This led to the hypothesis that in an *lrp* mutant, where the expression of glycine cleavage operon is low, glycine would not be able to support growth as a sole nitrogen source.

Accordingly, in the absence of NH3 and presence of glycine as sole nitrogen source (glyMM medium) at 42°C, a large difference in growth yield was achieved between the Δ*lrp* strain and WT (fig.3.2a). This difference was also easily seen in solid glyMM. The WT formed large colonies, but no colonies were detected with the Δ*lrp* mutant after 48h of incubation at 42°C (fig.3.2b, panels 1 and 2).

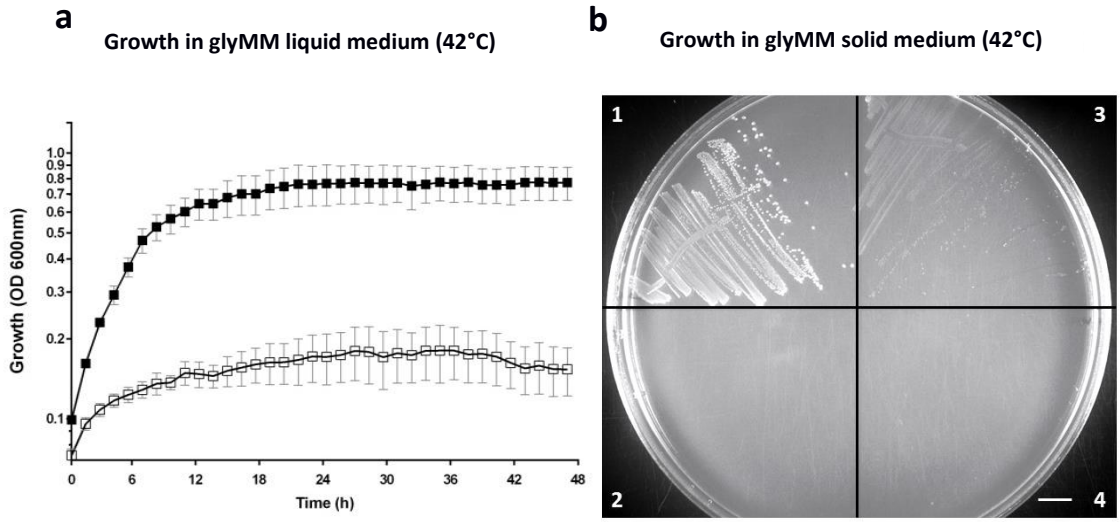


Figure 3.2. Growth in glyMM media. Panel a) Representative growth comparison of ancestral (DM08, black squares) and Δ*lrp* (DM08Δ*lrp* MFP08, white squares) in glyMM at 42°C. **Panel b)** Representative growth comparison on glyMM solid medium. Ancestral (DM08, (1)), Δ*lrp* (DM08Δ*lrp* MFP08, (2)), *lrp*::IS_{INT} (MFP89) (3) and *lrp*::IS_{COD} (MFP88) (4). Strains were streaked in a glyMM agar plate and incubated for 48h at 42˚C. Scale bar represents 1 cm.

Importantly, the phenotype of the Δ*lrp* and the *lrp*::IS_{COD} strains was the same in glyMM (fig.3.2b panels 2 and 4). On the other hand, the *lrp*::IS_{INT} had an intermediate phenotype, forming colonies smaller than the ancestral (fig.3.2b, panel 3). These results show that *lrp*::IS_{COD} mutants isolated from the samples collected during the evolutionary experiment are loss of *lrp* function, whereas the *lrp*::IS_{INT} mutation is likely a partial loss of function, resulting in lowered levels of Lrp. Alternatively, if disruption of the regulatory region caused by these insertion events led to overexpression of Lrp it would be expected to observe at least WT-size colonies due to unrestricted glycine catabolism and thus, NH3 assimilation. Therefore, we concluded that the genetic basis for the adaptation in the *lrp* locus is a loss of function.

3.1.2. **Evolutionary dynamics of the two major targets in *E. coli* mono-colonised mice**

During adaptation in mono-colonised mice, *E. coli* acquired mutations in two predominant loci. *Lrp* and the *gat* operon were the only two targets detected in all independent mice, in the sequencing data (8 out 8 mice tested). These results were obtained at one single time point, which is not sufficient however, to establish whether *lrp* mutants are well adapted to a steady-state germ-free environment, or if instead are the consequence of a transient pressure. The temporal dynamics of emergence of *lrp* mutants was determined in order to understand the type of selection *lrp* was subjected to. The distinctive capacity between WT and *lrp* mutants for growth in minimal medium with glucose and glycine (as in fig.3.2b) was adapted to score for the emergence and frequency of *lrp* phenotype in the populations from mice, during the course of the evolution experiment (see section 2.4). The phenotypes of *lrp* and *gat* were used as a proxy for the genotype of evolving populations.

Both traits were tested for each sampled clone to determine in which background *lrp* mutants emerge from, with respect to the first known adaptive step (*gat-negative*)[47].

Gat-negative mutants, forming white colonies on MacConkey plates with galactitol (gat-negative phenotype), were present in the populations collected from all mice from day 3 of colonisation (red lines in fig.3.3), reaching frequencies above 90%, by the end of the experiment. For mouse lines 4,5,6,7 the experiment was prolonged to day 40 and in these cases, such frequencies were still observed. Although gat-positive clones were still detected, a disrupted *gat* operon was clearly selected in this environment. This is in general agreement with previous experiments of *E. coli* adaptation to the GI tract of mice in polymicrobial conditions, where the first adaptive step of *E. coli* was also the selection for the gat-negative trait. In comparison, its emergence occurred later in mono-colonised mice, with the same ancestral *E. coli* strain (fig.3.3 and Batista et al.[47]). These results show that the emergence of the gat-negative trait is microbiota-independent, but indicate that the strength of the selection effects for this trait is microbiota-dependent.

In this sweeping gat-negative population, a second adaptive step was observed two weeks after colonisation, with the emergence of lrp-negative mutants (fig.3.3 blue and back lines). The majority of the clones with lrp-negative phenotype were scored with complete loss of function of *lrp*, as evidenced by their inability to grow on glyMM plates. These gat^-, lrp^- double mutants exhibited a strong increase in frequency in all mice (fig.3.3 blue lines). The temporal dynamics exhibit a clear pattern of selection for lrp-negative mutants with frequencies as high as 80% (in mouse 6), two weeks after emergence is observed (fig.3.3). Following this expansion, *lrp* mutants appear to stabilise in most populations. Particularly, in the mouse lines where experiments were prolonged up to day 40 (mice 4-7), where the lrp-negative population stabilised at frequencies ranging from 40-60% for 3 consecutive time points (day 28-35-40). These dynamics can be used to estimate the strength of the beneficial traits associated with mutations causing loss of function of *gat* and *lrp* (SuppFig.6.3). The selective advantage was estimated from the slope of the Ln((x/1-x)), where x is the frequency of lrp-negative (S_{lrp}) or gat-negative (S_{gat}) phenotype in time. These dynamics reveal an overall lower average selective strength for *lrp*, with S_{lrp}=0.019 (±0.005, 2se) in comparison with *gat* mutants S_{gat}=0.026 (±0.004, 2se).

Insertions in the intergenic region of *lrp* were also observed in all mice although highest frequencies in each mouse ranged from 1 to 30% (dotted black lines, fig.3.3). These mutant clones, identified by forming colonies smaller than the ancestral (as in fig3.2b (3)) with subsequent PCR confirmation, never reached the same levels as the lrp-negative mutants, indicating that insertions leading to the intermediate *lrp* phenotype provided a lower selective advantage than the lrp-negative mutants. Furthermore, the fact that in all the samples tested only two clones were detected to be *lrp*::IS_{INT}/gat-positive (Mouse2day16 and Mouse4day23) and none lrp::IS_{COD}/gat-positive, indicates that the possible gat^+/lrp^- population is rare.

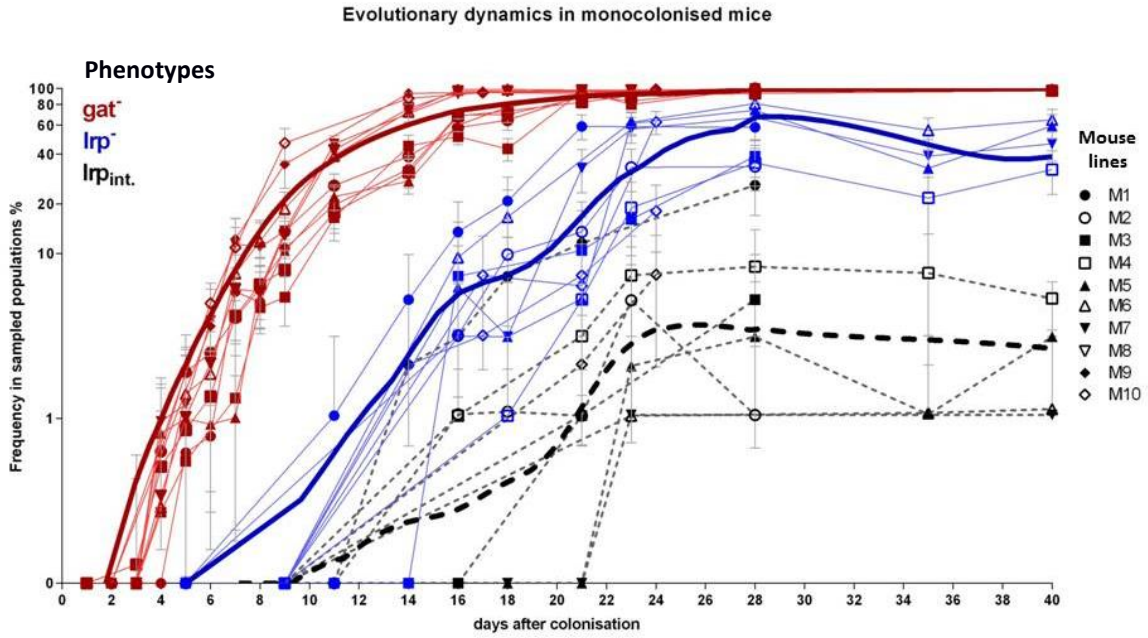


Figure 3.3. Evolutionary dynamics of *gat*⁻ and *lrp*⁻ phenotypes in mono colonised mice. Temporal evolutionary dynamics of the two first adaptive steps in mono colonised mice (gat-negative (red), *lrp*-negative (blue) and *lrp*::IS_{INT} (dotted black) were determined by following phenotype frequencies of 96 clones per time point, in glyMM plates (*lrp* phenotype) and MacConkey+galactitol (*gat* phenotype). Thick lines represent the average of individually fitted lines. n=10.

Overall, the high frequency of *lrp* mutants observed and their estimated strength reveal a clear advantage for *E. coli* upon loss of this master-regulator in the mono-colonised animals. However, unlike the *gat* mutants the populations of *lrp* mutants did not continue to increase and it stabilised at lower levels.

3.1.3. Identification of physiological conditions conferring advantage to *lrp* mutants

The results presented above, strongly support that when *E. coli* is the first coloniser of the GI tract, benefits from the loss of a functional Lrp. To show that indeed loss of the *lrp* gene is beneficial in mono-colonised mice, Δ*lrp* strain (MFP18) was competed with its ancestral (MFP54). Both strains were *a priori* deleted for the *gatZ* gene to avoid uncontrolled emergence of gat-negative mutants. Due to the fact that *lrp* mutants stay at intermediate levels (fig.3.3 blue), the advantage of the Δ*lrp* strain was tested with 2 different ratios (1:1 or 1:9) of Δ*lrp*Δ*gatZ* to Δ*gatZ* (fig.3.4a).

Unexpectedly, the *lrp* population declined in frequency in all mice from the first day of colonisation and no evidence of frequency-dependence was found, as a decrease in *lrp* frequency was observed at both initial ratios. However, four days after colonisation, the frequency of *lrp* mutants stabilized to levels between 2-20%. These results indicate a disadvantage of the *lrp* mutant during the

first days of colonisation, which would be against the hypothesis supported by the dynamics in the evolution experiment. However, we noticed that during these two weeks after colonisation, *lrp* clones never dropped under detection limits and thus, the *lrp* population was never eliminated. Furthermore, following the stable period, after day 10, an upshift in *lrp* frequency was observed in all mice, for three consecutive time points. Maximum levels in faecal samples reached 30% and did not appear to stabilise at the time that the experiment was terminated (day 15). Therefore, it is possible that loss of *lrp* is not beneficial at initial stages of colonisation and has advantage only at later stages when the GI tract has been altered by initial exposure to *E. coli*.

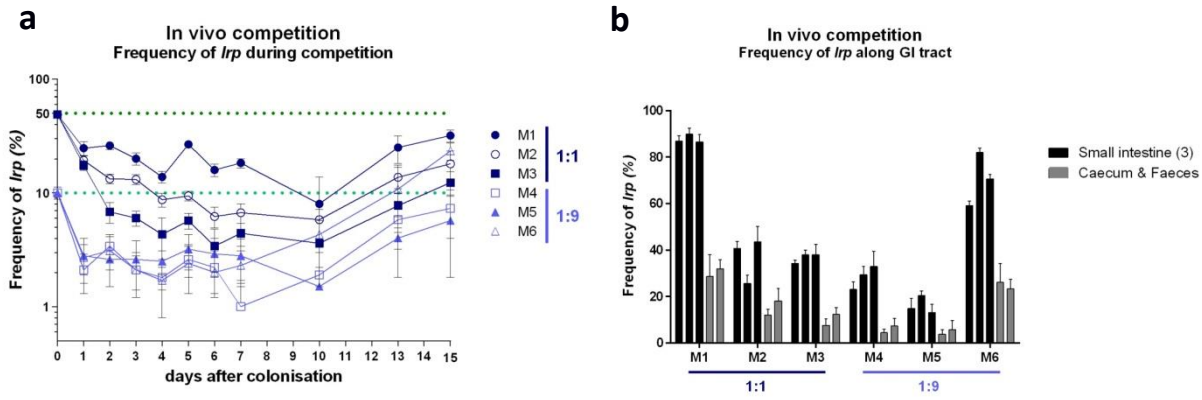


Figure 3.4. *In vivo* competition between *lrp* mutant and ancestral. Panel a) Frequency of *lrp* mutants in competition between MFP18-YFP (Δ*lrp*Δ*gatZ*) and MFP54-CFP (Δ*gatZ*) at two initial strain ratios (1:1 and 1:9 of MFP18 to MFP21) in germ-free mice. Dotted lines indicate initial strain ratios. (n=3). **Panel b)** Niche preference of *lrp*. Frequency of *lrp* (MFP18-YFP) in 5 sections of the GI tract determined by sectioning the whole GI tract and plating contents, from the last day of competition (day 15 of fig.3.4a). For each mouse, frequencies were grouped (from left to right) as: Small intestine (3 sections): proximal, medial and distal (black) and Caecum and Faeces (grey). Strain dynamics were followed by plating faecal samples and counting colony fluorescences. n=3.

To determine if the advantage of the *lrp* mutant could be different along different niches of the GI tract, at the last day of the experiment, the whole GI tract was sectioned, strain frequencies were determined and results were compared with those of faecal samples (fig.3.4a day 15). Interestingly, there was a significant heterogeneity in the frequency of the strains (fig.3.4b). In all mice, the abundance of *lrp* was higher in the 3 sections of the small intestine, compared to the caecum and faecal samples (fig.3.4b). These results suggest the existence of niche preference for *lrp* mutants within the mono-colonised GI tract. Additionally, these results also show a clear advantage for *lrp* mutants in the small intestine.

3.1.4. **Physiological basis of fitness advantage of *lrp***

Lrp has been reported to directly regulate over 100 genes and likely indirectly bring about changes in many more[58]. To decipher the source of advantage upon loss of this master-regulator, an appropriate approach seemed to be one that would integrate all genes up-to-date annotated as part of the Lrp regulon and narrow down on specific categories of biological functions likely to be relevant *in vivo*. By means of the pathway analysis toolbox of the *E. coli Encyclopedia* (*Ecocyc*®)[75], pathways of the *lrp* regulon were grouped and the metabolic networks of aminoacid synthesis and degradation were connected to visualise and paint according to manual prediction of possible differential regulatory effects between WT and *lrp* mutants (see section 2.6).

As can be seen in SuppDiagram.1 (biosynthesis), synthesis of most aminoacids that include genes affected directly by Lrp are overall diminished in an *lrp* mutant. Particularly several genes for the synthesis of branch-chained aminoacids (Isoleucine, Valine and Leucine) are downregulated in an *lrp* mutant. Availability of these aminoacids may have played a role in selecting for the loss of *lrp*. This was tested on the fitness (relative growth rate) of Δ*lrp* and the two isolated *lrp*::IS_{COD} and *lrp*::IS_{INT} strains relative to the ancestral (fig.3.5a). As sole carbon sources, branch-chained aminoacids (BCAA) did not support growth of any strain. With glucose the three *lrp* strains exhibit lower relative fitness compared to the ancestral. With glucose and BCAA the *lrp* mutants still have a lower fitness but the difference between ancestral and *lrp* isolate strains is reduced. Importantly, these aminoacids are identified *in vivo* by NMR (SuppFig.6.4) and thus, may have been a relevant factor during adaptation.

Because the predominant class of targets during adaptation relate to the use of aminoacids (SuppFig.6.1), it is likely that glucose is not the preferred carbon and energy source of *E. coli* in mono-colonised mice. This is also supported by the absence of carbohydrates in the identification of metabolites by NMR (SuppFig.6.5 GF) that are usable by *E. coli*, particularly glucose. Raffinose and sucrose are tri- and di-saccharides that could be identified in germ-free mice but would be unusable. Thus, other substrates instead of glucose were tested in the basal medium that included branch-chained aminoacids (BCAA, in fig.3.5a). Additionally, Leucine, the third branch-chained aminoacid and the co-factor of Lrp activity, was tested for its potential regulatory functions, as it should directly modulate the ancestral (*lrp*$^+$) and perhaps *lrp*::IS_{INT} mutants, when functional Lrp is produced.

Next, the analysis of degradative metabolic networks (SuppDiagram.2 degradation) was used to identify particular substrates that could justify an advantage of *lrp* mutants *in vivo*. This analysis indicates that *lrp* mutants may be more efficient in degrading aminoacids, when compared to an *lrp*$^+$. Specifically, Alanine, Serine and Threonine stood out as good candidates (SuppDiagram.2).

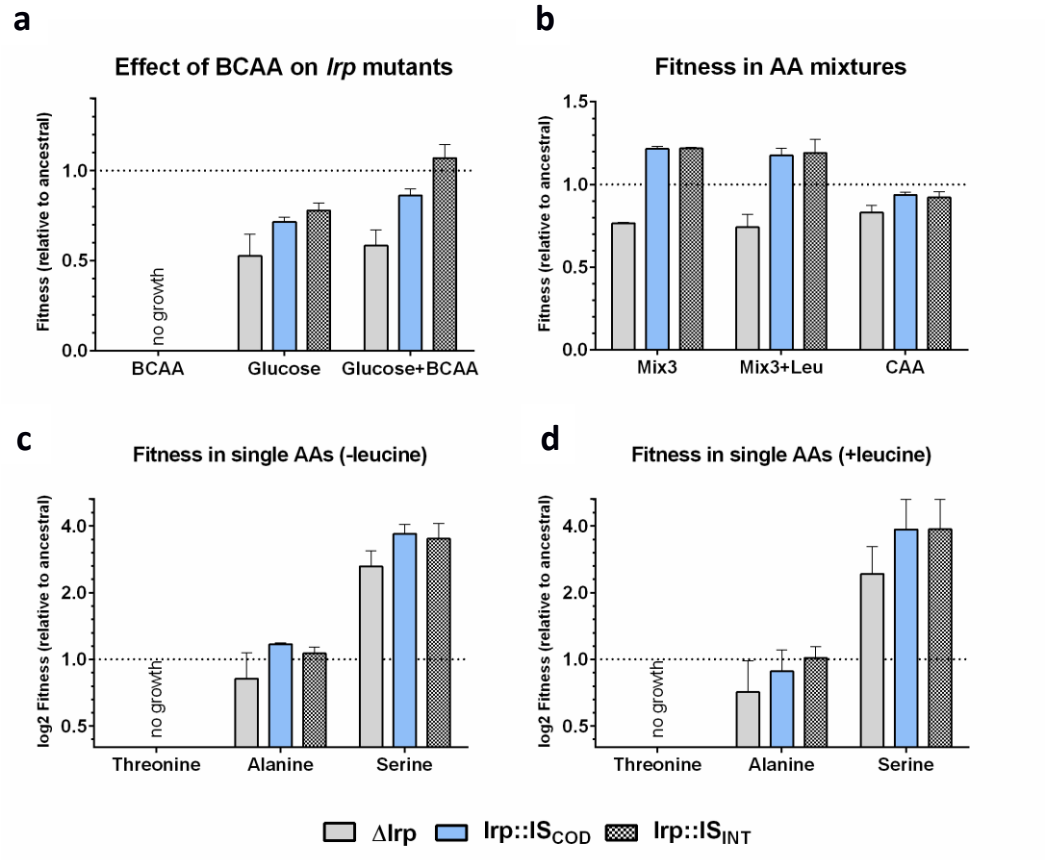


Figure 3.5. Fitness of *lrp* mutants in the presence of aminoacids. Panel a) Effect of BCAA's on relative fitness of *lrp* mutants. Fitness of *lrp* mutants (Δ*lrp*Δ*gatZ* (grey), *lrp*::IS_{COD} (blue), *lrp*::IS_{INT} (dark grey)) relative to the ancestral (DM08 Δ*gatZ*) in monocultures of minimal medium (M9), with either BCAA - Isoleucine(0.25%)+Valine(0.25%), Glucose (0.5%), or the combination of the three (Glucose+BCAA). **Panel b) Fitness of *lrp* mutants in mixtures of aminoacids.** M9 with either a complex mixture of aminoacids (CAA 0.5%) or an equal mixture of L-Alanine, L-Serine, L-Threonine 0.5% plus BCAA 0.5% (Mix3), +/- L-Leucine 0.5%. **Panels c,d) Fitness of *lrp* mutants in single aminoacids.** Alanine, Serine and Threonine were tested individually (0.5%) in M9+BCAA 0.5% +/- L-Leucine 0.5%. Relative fitness is shown as maximum growth rates (*GrowthRates*®) of each strain relative to the ancestral from 3 independent experiments + SD.

In a complex mixture of aminoacids (CAA), the three *lrp* mutants exhibit a lower fitness relative to their ancestral (fig.3.5b CAA), demonstrating that all mutants are capable of growing in the presence of all aminoacids, however not better than the WT. Therefore, if the mutants have advantage in aminoacid degradation this advantage could be specific for a subset of aminoacids as suggested by the degradative metabolic networks (SuppDiagram.2). When present at lower concentrations, as in CAA, their availability may be insufficient to outcompete the ancestral. In agreement, when the same total concentration of the three specific aminoacids alanine, serine and threonine were used as substrate, the *lrp* evolved clones have a 20% higher fitness relative to their ancestral (fig.3.5b Mix3). This is true for the mutants with insertions in either the *lrp* coding or the intergenic region, indicating that these mutants have advantage over the WT in this medium. Furthermore, higher fitness in such conditions was found to be independent of the presence of leucine signalling, as similar results were obtained with and without addition of Leucine (fig.3.5b Mix3+Leu). Unexpectedly, the Δ*lrp* strain

behaved differently from insertion isolates and did not exhibit a fitness advantage relative to the ancestral in the mixture of the 3 aminoacids. To further characterise the preference for each aminoacid, relative fitness was determined by supplementing the BCAA medium with each of the three aminoacids from the mixture tested separately, at the same total aminoacid concentration fig.3.5 c,d). First, relative ability for the catabolism of threonine as a sole carbon source could not be determined, as it did not support growth of neither the WT nor the mutants after 48h of incubation (fig.3.5 c,d). Alanine alone supported growth and, in such conditions, the *lrp*::IS_{COD} mutant outcompeted the ancestral with a 20% higher fitness. The *lrp*::IS_{INT} and Δ*lrp* strain appear neutral or deleterious (fig.3.5c). Leucine presence had no large effect on the relative fitness of the three mutants (fig.3.5d). Finally, serine is clearly the preferred aminoacid of the three, as it leads to an over 2-fold higher relative fitness in all *lrp* strains, independently of leucine treatment.

Together, these results suggest that mutations in *lrp* can increase ability to degrade and thus, compete with the ancestral by utilising a subset of aminoacids, specially serine. This is further supported by the decrease in serine, in caecum contents of germ-free mice after colonisation with *E. coli* (SuppFig.6.4 GF and GF+*E. coli*).

3.2. **Polymicrobial community: Antibiotic-treated Microbiota**

A goal of this work was to test if adaptation could be interpreted from a physiological sense and add insight to mechanisms that benefit this species in the environment of its highest abundance. The closest environment in which *E. coli* adaptation can be easily tested is the streptomycin-treated mouse model. *DcuBfumB* and *focApflB*, together the most frequent and parallel mutations, specific to adaptation in polymicrobial conditions (microbiota-dependent), were chosen as good candidates of high fitness peaks. The mechanistic basis of their expansion was explored here. The structure, ontology, and function of these two operons (*dcuBfumB* and *focApflB)* are seemingly very similar. Both mutations are caused by insertions in the regulatory region of both operons which have been shown to be able to increase downstream transcription[53]; both consist of a transporter followed by a biosynthetic gene[59–61]; both are known to be positively regulated by Fnr[62]. One of the two targets, the *dcuBfumB* operon, codes for a C4-dicarboxylate transporter (*dcuB*) and a fumarate-producing enzyme (*fumB*). The other, in the *focApflB* operon codes for a formate transporter (*focA*) and a formate-producing enzyme (*pflB*). The function of these operons, extensively studied *in vitro*, has been considered to be involved in modulating metabolism to conditions without oxygen, but the strain used in this work bears a known genetic polymorphism in *E. coli* K12 strains: the absence of Fnr, an important oxygen-responsive regulator. Both of these targets were detected in polymicrobial conditions but not in monocolonised mice, supporting the relevance of these metabolic functions in an anaerobic, poymicrobial environment like the GI tract.

3.2.1. *F*itness of *dcuBfumB* and *focApflB* mutants evolved in polymicrobial conditions

In agreement with the similarities in regulation and function, both mutants have also been shown to provide a similar selective advantage *in vitro* (Marta, *2014*[53]), in a medium with carbon sources known to be used by E.coli *in vivo*[76]. A limitation in (Marta, *2014*) was that the fitness of the evolved clones was tested in the presence of oxygen. Since the GI tract is predominantly anaerobic, the evolved *E.coli* strains could have a higher selective advantage in the absence of oxygen. To test this, *in vitro* competitions were carried out with the same carbon sources as before, though in the absence of oxygen. Several reasons support that this composition is a relevant nutritional proxy for the biological environment studied here: first, four of the five carbon sources (gluconate, ribose, glucuronate and mannose) have been previously shown to be preferred by *E. coli*, *in vivo*; and in agreement, as detailed in the introduction, one of these compounds (Gluconate), is very likely available and preferred, in the microbiome of the mice specifically studied here. Because the transcriptional repressor for the transport and utilization of gluconate was disrupted in several mice (SuppFig.6.1 *kdgR*), that selected loss of function likely increased the ability to use gluconate in those mutants. A similar case for the sorbitol repressor (SuppFig.6.1 *srlR*), supports the relevance of this sugar-alcohol in the microbiome of these mice, as the repressor of the operon of sorbitol utilization was another target disrupted during the evolution experiment in polymicrobial conditions[47].

To study the basis of the emergence of mutations in the *dcuBfumB* and *focApflB* operons, competition experiments were performed *in vitro*. Since these mutants were predominantly observed in the gat-negative background and for better comparison, both the mutant and the ancestral strains were deleted for the *gatZ* gene.

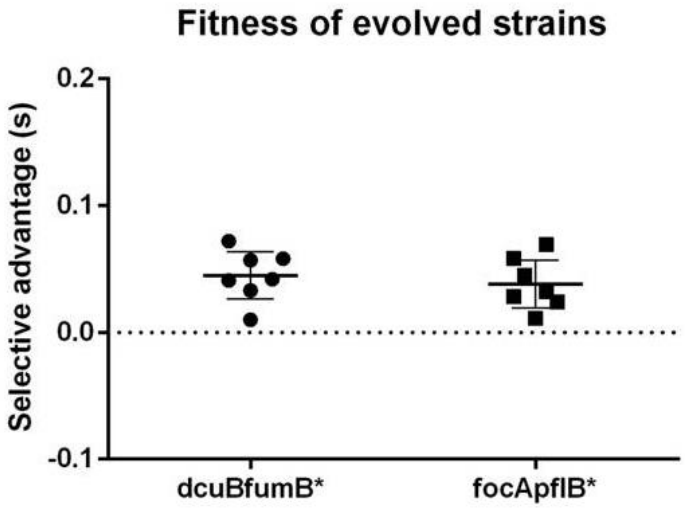


Figure 3.6. Fitness of *dcuBfumB* and *focApflB* mutants evolved in a polymicrobial Microbiota. Fitness of *dcuBfumB** (CP5) and *focApflB** (MFP59) was each tested against their ancestral (Δ*gatZ*) in vitro, under anaerobic conditions with Minimal medium (M9), with an *in vivo*-like mixture of sugars as carbon sources. L-Gluconate, L-Glucoronate, L-Sorbitol, L-Mannose and L-Ribose were each added at 0.01% (w/v). Selective coefficient (s) +/- SD is shown as natural logarithm of the ratio of mutant to ancestral at 24h, over the ln of the initial ratio (1:1), normalised by number of generations of each co-culture. n=7.

As expected, both evolved strains out competed the ancestral in anaerobiosis (fig.3.6), when growing with simple sugars present in the GI tract, shown to be preferred carbon sources *in vivo*[39,76]. Because this selective advantage was tested in the absence of the transporters' substrates, formate (*focA*) and fumarate (*dcuB*), the selective advantage of the evolved strains was likely caused by increased intracellular production by the downstream enzymes, known to produce formate (*pflB*) and fumarate (*fumB*). This likely enabled their utilisation by downstream pathways. If this is true, then increased internalization by virtue of transporter upregulation[53] would be beneficial, if these metabolites are available *in vivo.* Thus, external supplementation should increase the fitness of the evolved strains, even in these *in vitro* conditions. As shown in fig.3.7, the selective advantage increased with addition of both mutants' respective substrate. Fumarate increased selective advantage of *dcuBfumB* mutant (fig.3.7a fumarate) and Formate treatment increased that of the *focApflB* mutant (fig.3.7b formate), against the ancestral. These results further support that both mutations lead to a gain-of-function, increasing transport and use of their respective substrates.

3.2.2. **Physiological basis of fitness advantage of *dcuBfumB* & *focApflB***

The results above successfully demonstrate that an advantage conferred by two major evolved mutants selected in polymicrobial conditions, can be reproduced *in vitro*. However, they do not explain the mechanistic basis of this advantage when growing in the GI tract. Unexpectedly, the absence of oxygen did not increase the selective advantage of either mutant relative to the ancestral

genotype, under these conditions (fig.3.6 compared with Marta, *2014*[53]). This suggests that quantitatively, both the mutants and the ancestral strain may experience the energetic limitations imposed by oxygen restriction with similar effects. Regardless, we assumed that a completely anaerobic environment would be more adequate to study anaerobic respiration and *closer* to *in vivo* conditions, than a well-agitated culture. In such conditions, *E.coli* takes advantage of its respiratory functions to grow in a wide range of environments[77].

If these mutants were selected to favour specific respiratory functions, their fitness should exhibit characteristic features and follow key regulatory signals upon changes in nutritional environment. Specifically, if *focApflB* evolved to divert pyruvate to formate through increased expression of pyruvate-formate lyase (*pflB*), it should be capable of coupling this increased production to nitrate respiration. On the contrary, if *dcuBfumB* evolved to be poised for fumarate respiration, known to be disfavoured when nitrate is available[78], then it should not benefit if nitrate is present. Treatment effect was determined as selective coefficient (s) of each biological replicate with treatment, minus each selective coefficient without treatment, for the same biological replicate ((s) in fig.3.6). As can be seen in fig.3.7 with each evolved mutant co-cultured with its ancestral, nitrate treatment doubled the fitness of the *focApflB* mutant in relation to sugars alone (fig.3.7b +NO_3), whereas the same conditions had no effect on the *dcuBfumB* mutant (fig3.7a +NO_3). Similarly, when nitrate is supplemented with fumarate, no difference is observed in relation to conditions with just fumarate treatment (fig.3.7a +NO_3 and +fumarate&NO_3), whereas addition of nitrate and formate together are somewhat additive for the fitness of the *focApflB* strain (fig.3.7b +NO_3 and +formate&NO_3).

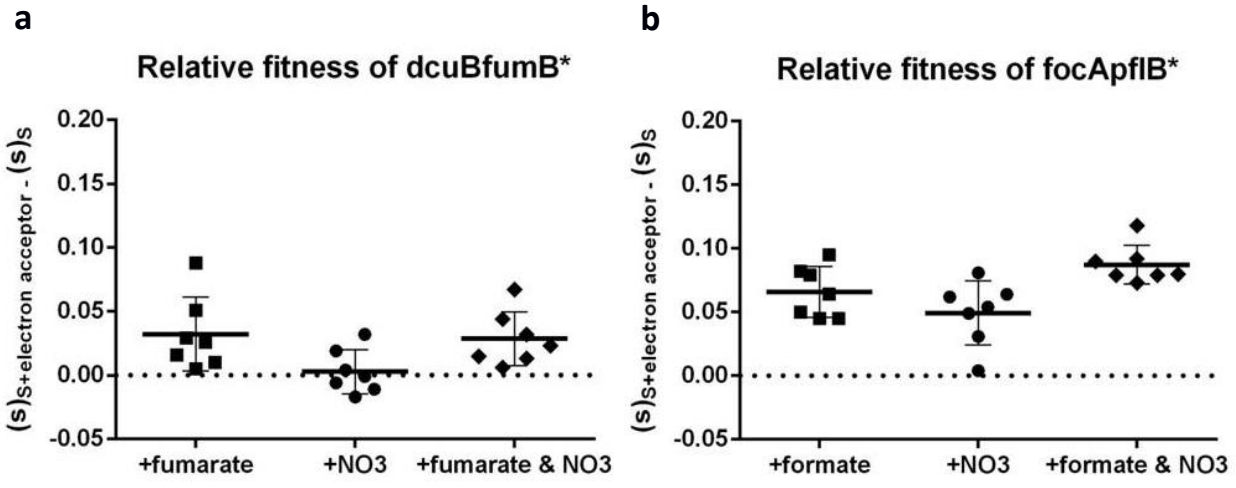


Figure 3.7. Effect of formate or terminal electron acceptors on fitness of evolved clones. Fitness of the *dcuBfumB** (CP5) and *focApflB** (MFP59) mutants evolved in a polymicrobial Microbiota, was tested against the ancestral (Δ*gatZ*) in conditions as in fig.3.6, but with each electron acceptor or formate added at 40mM. *Treatment effect is shown as selective coefficient (s) of each biological replicate with treatment, minus each selective coefficient without treatment ((s) in fig.3.6). A value of 0 indicates the same (s) obtained without treatment. n=7.

4. Discussion

The present work was designed to obtain further insight on how bacteria can establish and persist in the ecological contexts of the complex intestinal environment. An increasingly systems level understanding from physiological studies *in vitro* and metagenomics, immunological and epidemiological studies, has led to renewed interest in the promotion of health and biotechnological processes through intestinal microbial ecosystems. It is thus essential to understand how *in vivo* bacteria extract energy, carbon and nitrogen for successful colonisation.

This work focused on a particular species of enteric bacteria, by studying a particular strain of *E. coli* K-12, one of the best studied model organisms. Use of this strain (MG1655) has been the most prevalent in recent studies, is well characterised *in vitro*[79,80] but has also been studied in the intestinal environment[39,47,64,68]. Additionally, this particular strain bears a known natural polymorphism of a deletion of the *fnr* gene[81], a regulator of respiratory functions, likely constraining its metabolic capacity. Here, we studied the effects of *E. coli* mutants selected during adaptation in germ-free and also in polymicrobial mouse models to address the effect of the microbiota[82].

An immediate realisation was the richness of adaptive targets in both environments. Both in the presence and absence of a microbiota *E. coli* was observed to generate a breadth of beneficial mutations in the span of one month. The majority of these mutations occurred in genes known to affect substrate nutritional use, highlighting insertions in gene regulatory sequences and/or transcriptional regulators. The ability for this bacterium to effectively replicate and persist in the intestinal environment is well established, but these adaptive events suggest that metabolism *in vivo* is not optimal.

Specifically, this founded the hypothesis that the establishment of a metabolic landscape compatible with the state and composition of the intestinal environment was the main adaptive path followed by *E. coli*. This highlights, from an evolutionary scale, the fundamental importance of metabolism for intestinal colonisation. When applied to the mutational profiles, this interpretation suggested that when *E. coli* colonises the GI tract alone, a shift towards the preference for growth with aminoacids was supported by mutations in several genes related with aminoacid use. The most likely mutation to bear a strong effect in metabolism was the master regulator *lrp*. Importantly, mutations in *lrp* were not identified in sequenced populations evolving in polymicrobial conditions, whereas it was observed in mono colonized mice with or without streptomycin treatment (data not shown), supporting that the benefit(s) conferred by such mutations are dependent on absence of microbiota. Because it was also the most parallel event, specific to the mono colonised intestine (10 out of 10 mice), *lrp* was chosen for further characterization.

Likewise, adaptation of *E. coli* in the presence of a polymicrobial microbiota, affected predominantly the ability for nutritional utilisation. Overall, the presence of a complex microbiota brought about an almost completely distinct mutational profile. Besides the galactitol operon (*gat*), all

other targets were exclusive to a polymicrobial condition. Still, targeting specific carbon sources to extract energy from, and thus grow, was a large fraction of the adaptive path followed with a number of targets relating to utilization of sugar-like compounds but not aminoacids. From this functional perspective, however, only the class of respiratory functions was exclusive to the polymicrobial environment. Both targets of this class, *dcuBfumB* and *focApflB*, were chosen for further characterization.

The specific questions raised here were: What are the genetic bases of these beneficial mutations? What are the resulting physiological effects? What is the benefit of acquiring them? How do the answers to these questions relate to the state of the microbiota? To address these questions, experiments at both the level of the population and of the individual were designed.

The genetic basis of these mutations, obtained by sequencing populations after 23 days of adaptation showed that in the three focal targets studied, integration of insertion sequences predominated. In the cases of *dcuBfumB* and *focApflB*, insertions were only observed in regulatory regions, indicating that mutations abolishing gene function were not selected. Indeed, Marta, *2014* showed that the evolved mutants have increased transcriptional expression of downstream genes.
As for *lrp*, this was uncertain, as integration of insertions was not only in regions capable of affecting transcription but also disrupting the protein-coding region. We show that insertions in the coding region of *lrp* indeed lead to the inability to grow in a medium where glycine degradation, normally activated by Lrp, is required for growth. Since a strain deleted for *lrp* also failed to grow but not the ancestral, we conclude that the IS elements in the coding region of *lrp* of evolved clones, lead a loss of function of *lrp*. Insertions in the regulatory region of *lrp* could also affect Lrp levels, for which glycine degradation would serve as a proxy. We show that these lead to small colony sizes and propose that restricted activation of glycine degradation, due to lower levels of Lrp cause this phenotype.

We extended this assay to characterize the distribution of *lrp* mutants evolving in the 10 independent populations and measured their selective effects. Signature dynamics clearly demonstrated a pattern of selection for *lrp* mutants, that is higher for complete loss of Lrp function (S_{lrp}=0.019 (±0.005, 2se)), whereas insertions that cause intermediate effects on Lrp function were selected to very low maximum levels (1-30%). The estimated strength of *lrp*$^-$ mutations was similar to that of *gat*$^-$ mutations (S_{gat}=0.026 (±0.004, 2se)), but occurred later, when the *gat*-negative trait had already stabilised at high frequencies (>90%). We found that *lrp* mutants emerged from a *gat*-negative background, demonstrating that there is no antagonistic pleiotropy between *lrp* and *gat*. Since the latter almost reaches fixation, one could expect a similar expansion of double mutants. However this was not the case, the *gat*$^-$/*lrp*$^-$ population only reached intermediate levels, stabilising at ~50%. Two mutually compatible explanations are that perhaps the mono-colonised GI tract represents a limited niche for *lrp* mutants. This would be compatible with the temporal overlap between the *in vivo* competition and temporal dynamics of the *de* novo mutants, which suggests a deterministic change in the environment and that it is perhaps the initial effects of colonisation that create the environment that

favours the loss of *lrp,* whereby the increased use of aminoacids then limits the continuous expansion of the *lrp* population. The second hypothesis is that adaptation in the mono colonised intestine follows a pattern of clonal interference, whereby other beneficial mutation(s), on an lrp-positive/gat-negative background stably compete(s) with *lrp*/*gat* double mutants preventing fixation of any allele. Other beneficial mutations would require further characterisation.

By analysing *in silico* the metabolic network regulated by Lrp we predicted that overexpression of Threonine deaminase (*ilvA*) and alcohol-aldehyde dehydrogenase (*adhE*) would increase threonine degradation. Similarly, alanine, by overexpression of glutamate-pyruvate aminotransferase (*alaC*) and/or D-aminoacid dehydrogenase (*dadA*) and serine by overexpression of serine deaminase (*sdaA*), would also have increased degradation into pyruvate, a central energy metabolite. The three aminoacids would support higher advantage of *lrp* mutants. We show that this was indeed the case with two different isolated *lrp* mutant strains, but observed that the aminoacid profile is a crucial factor, as no fitness advantage was found with a more complex mixture of casaminoacids, at the same concentration. A clear inconsistency should be noted, regarding the reference Δ*lrp* strain. With exception of serine individual medium, this reference strain always appeared to bear a basal cost that several times rendered it deleterious, when the isolated *lrp* mutants did not. The isolated mutants were obtained from early samples exactly to avoid other mutations, but it is possible these were present and may have affected fitness measurements. Alternatively, the disruption of a known promoter of a downstream essential cell-division gene (*ftsK*), located within the coding region of *lrp* (fig.3.1), could explain this apparent cost of the deletion strain, whereas the isolates, only have added insertion sequences. Interestingly, leucine seemed to have little effect on all strains. These competitions were done in monocultures to measure the isolated effects of mutations in *lrp* and thus the fitness's measured assume no interaction. It would be relevant to test this in co-cultures, where we expect the advantage of *lrp* mutants to increase. Not only would loss of *lrp* lead to increased expression of aminoacid and peptide transporters, increasing competitiveness for access to these nutrients but specifically, because *lrp* mutants are known to be scavengers of leucine due to constitutive expression of LIV-type transporters which are specific for branch-chained aminoacids. In such a case, presumably a WT would see less leucine, thus lowering the perceived signal to shift towards aminoacid consumption through Lrp regulation, especially *in vivo* where the scale of competition is more local. Nevertheless, mutations in *lrp,* selected *in vivo,* have indeed the potential to increase fitness in alternative carbon sources like aminoacids.

An interesting comparison with work by Kolter et al., further supports this hypothesis. During prolonged stationary phase in LB medium, *rpoS* and successively *rpoS/lrp* double mutants are selected[83]. Two observations point to important differences between this and environment studied here. First, there is no evidence of growth arrest upon colonising the germ-free GI tract as would occur in stationary phase LB cultures. The authors do not report *lrp*-only mutants and report that first, mutations in *rpoS* serve as a way to evade the group behaviour termed stationary-phase. Upon this

evasion, dividing *rpoS* mutants gain further mutations, of which the *lrp* gene is one of several that allow better growth on the carbon sources available: aminoacids[83,84]. In these conditions, depletion of prefered carbon sources like glucose or similar sugars are believed to be the main cause for LB cultures to progress into stationary phase. This is similar to the results shown here, particularly in relation to the type of nutritional transition.

A second comparison with this work and Kolter et al. is that, even if *rpoS* mutants were to provide a higher threshold for growth arrest in the germ-free GI tract, re-colonisation would likely be hampered: Opposing the rate of intestinal washout is oral re-inoculation, where the first barrier encountered is the stomach's acidic environment. In such conditions an intact *rpos* would be transiently crucial, until a more suitable environment for growth is reached in the intestine. This correlates with *in vitro*, where *rpoS* mutants lose their selective advantage when the pH of LB is lowered[85]. In conclusion, comparable environmental conditions between *in vivo* and *in vitro* environments both selecting for *lrp* allow the extrapolation of commonalities. Together, these suggest that *E. coli* sees a sugar-limited environment in mono colonisation and that absence of a functional Lrp shifts metabolism towards the available nutrients in the absence of the microbiota.

In the case of *dcuBfumB* and *focApflB,* there was no indication of pleiotropic effects as strong as with *lrp* and so, a simpler approach characterised the physiological effects of these adaptive events. Namely, we were interested in comparing both mutations (*dcuBfumB* and *focApflB*), as these were not observed together and appeared to be mutually exclusive. Our hypothesis was that the presence of the microbiota llimits the environment in a way that favours the metabolic functions generated in these mutants. We phenotypically show that, insertions in the regulatory region in these mutants increased ability to uptake either fumarate or formate because their supplementation increased fitness, supporting the increased transcription of *dcuB* and *focA*[53]. Additionally, the likely increased expression of *fumB* and *pflB,* presumably also increased endogenous production since, as we show, both mutants have an advantage even without supplementation, in anaerobic conditions. Increased intracellular availability in the evolved clones could have increased ATP formation indirectly, in *dcuBfumB,* through NADH dehydrogenase-coupled proton export during fumarate respiration[86] and directly, in *focApflB,* through formate dehydrogenase-mediated proton export during formate degradation with or without nitrate reduction. Nitrate (NO_3) served as a distinctive role between mutations. Whereas NO_3 and formate have been shown to be synergistic through shuttling of electrons from formate dehydrogenase to nitrate reductase, fumarate and NO_3, on the other hand, overlap in function and being detrimental together, fumarate respiration is inhibited in the presence of NO_3. We show that indeed nitrate addition increased fitness of *focApflB,* even without formate supplementation, but not that of *dcuBfumB*. Nitrate repression of fumarate respiration has been shown to be independent of oxygen as well as occurring in *fnr* mutants[78]. Addition of nitrate which would repress fumarate reductase (*frd*) and thus oppose fitness effects of mutations observed in *dcuBfumB,* had no large effect

on its fitness. This indicated that fumarate flows through a different enzyme, in these mutants. Indeed it has been shown that succinate dehydrogenase (*sdh*) can replace fumarate reductase (*frd*), by catalysing in reverse of its physiological reaction[87]. We propose this may happen in this strain because it lacks *fnr*, which would replace Sdh with Frd by respectively repressing and activating transcription, when transitioning to an anaerobic environment[87]. Furthermore, we show that oxygen presence did not alter the overall fitness without supplementation of electron acceptors, since a 5% selective coefficient in anaerobic conditions is in agreement with measurements in aerobiosis[53]. We propose that this can be explained by a reversion in the pattern of expression of *dcuBfumB* and *focApflB* in aerobiosis. It is possible that experiments in Marta, *2014* actually measured the ability of the mutants to not express DcuBfumB/FocApflB in aerobiosis, because their functions would be disfavoured by the presence of oxygen, in strains with a functional Fnr[88].

Thus, overall we link the increased function of DcuBfumB and FocApflB with anaerobic respiration, suggesting it as an explanation for the selection of these mutations in a polymicrobial environment where aminoacids are constantly depleted by the rest of the microbiota and the ability for efficient use of simple sugars through anaerobic respiration assumes a crucial role in competitiveness of *E. coli*. These observations are in general agreement with previous work[64] and would support them from an evolutionary scale also. In mono-colonised mice, not only are simple sugars for respiration limited due to the absence of cellulolytic and saccharolytic bacteria necessary to release these nutrients, but the abundance of aminoacids would have undermined the fitness effects of *dcuBfumB* and *focApflB*, compared with *lrp*. On the other hand, it can also be highlighted that the inverse would likely be true. In the presence of a microbiota, balance between aminoacid production and degradation achieved through proper control of *lrp*, should represent a crucial factor in transitioning from the outside to life inside of a host, as has been suggested[89]. Upon colonisation, the efficient respiration of the nutrients made available by other bacteria, through increased proton motive force coupled to ATP production favors proliferation of *E. coli*.

In a stable microbiota where nitrate is depleted[64], degradation of simple sugars through glycolysis produces pyruvate, which supplies the reductive branch of the TCA cycle[77], producing ATP and/or balancing NADH/NAD^+ levels during fumarate respiration. In a disturbed microbiota, where NO_3 is abundant[90], pyruvate can be shuttled to formate, instead of the TCA, and in this way, couple to respiration of NO_3 and ATP production. This would explain why *dcuBfumB* was selected more often than *focApflB* during antibiotic treatment while only *focApflB* but not *dcuBfumB* was selected in mice with compromised adaptive immune system treated with antibiotics[68,90–92].

In conclusion, this work adds insights on the intestinal microbiota and may help future studies to better understand the complex networks that integrate nutrients, microbes and hosts in this complex ecosystem.

5. Bibliography

1 Savage, DC Dwayne C (1977) Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.*, 31(1), pp. 107–133.

2 Sender, Ron et al. (2016) Revised estimates for the number of human and bacteria cells in the body. *bioRxiv*, p. 36103.

3 Aagaard, Kjersti et al. (2014) The placenta harbors a unique microbiome. *Sci. Transl. Med.*, 6(237), p. 237ra65.

4 Dominguez-Bello, Maria G et al. (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. U. S. A.*, 107(26), pp. 11971–5.

5 Fernández, Leónides et al. (2013) The human milk microbiota: Origin and potential roles in health and disease. *Pharmacol. Res.*, 69(1), pp. 1–10.

6 Fallani, Matteo et al. (2011) Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology*, 157(5), pp. 1385–1392.

7 Frese, Steven A et al. (2015) Diet shapes the gut microbiome of pigs during nursing and weaning. *Microbiome*, 3, p. 28.

8 Hasegawa, Mizuho et al. (2010) Transitions in oral and intestinal microflora composition and innate immune receptor-dependent stimulation during mouse development. *Infect. Immun.*, 78(2), pp. 639–650.

9 Huttenhower, C et al. (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), pp. 207–214.

10 Thompson, Jessica Ann et al. (2015) Manipulation of the quorum sensing signal AI-2 affects the antibiotic-treated gut microbiota. *Cell Rep.*, 10(11), pp. 1861–1871.

11 Martínez, Inés et al. (2013) Long-Term Temporal Analysis of the Human Fecal Microbiota Revealed a Stable Core of Dominant Bacterial Species. *PLoS One*, 8(7).

12 Faith, J J et al. (2013) The long-term stability of the human gut microbiota. *Science (80-.).*, 341(6141), p. 1237439.

13 Zoetendal, Erwin G et al. (2012) The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J.*, 6(7), pp. 1415–26.

14 Cornick, Steve et al. (2015) Roles and regulation of the mucus barrier in the gut. *Tissue Barriers*, 3(1–2), p. e982426.

15 Gibson, G. R. et al. (1996) Fermentation of non-digestible oligosaccharides by human colonic bacteria. *Proc. Nutr. Soc.*, 55, pp. 899–912.

16 Gordon, H A et al. (1971) The gnotobiotic animal as a tool in the study of host microbial relationships. *Bact. Rev.*, 35(4), pp. 390–429.

17 Donohoe, Dallas R et al. (2011) The Microbiome and Butyrate Regulate Energy Metabolism and Autophagy in the Mammalian Colon. *Cell Metab.*, 13, pp. 517–526.

18 Li, Hai et al. (2015) The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.*, 6(May), p. 8292.

19 Lawley, Trevor D. et al. (2013) Intestinal colonization resistance. *Immunology*, 138(1), pp. 1–11.

20 Frank, Daniel N et al. (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U. S. A.*, 104(34), pp. 13780–5.

21 Turnbaugh, P J et al. (2009) A core gut microbiome in obese and lean twins. *Nature*, 457(7228), pp. 480–484.

22 Russell, Shannon L et al. (2012) The impact of gut microbes in allergic diseases. *Curr. Opin. Gastroenterol.*, 28(6), pp. 563–569.

23 Yurist-Doutsch, Sophie et al. (2014) Gastrointestinal microbiota-mediated control of enteric pathogens. *Annu. Rev. Genet.*, 48, pp. 361–82.

24 Hibbing, Michael E et al. (2010) Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.*, 8(1), pp. 15–25.

25 Kaper, J B et al. (2004) Pathogenic Escherichia coli. *Nat. Rev. Microbiol.*, 2(2), pp. 123–140.

26 Rohmer, Laurence et al. (2011) Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol.*, 19(7), pp. 341–348.

27 Jones, Shari A. et al. (2007) Respiration of Escherichia coli in the mouse intestine. *Infect. Immun.*, 75(10), pp. 4891–4899.

28 Winter, Sebastian E et al. (2010) Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature*, 467(7314), pp. 426–9.

29 Ijssennagger, Noortje et al. (2015) Gut microbiota facilitates dietary heme-induced epithelial hyperproliferation by opening the mucus barrier in colon. *Proc. Natl. Acad. Sci.*, 112(32), p. 201507645.

30 Hoffman, Lucas R. et al. (2010) Nutrient availability as a mechanism for selection of antibiotic tolerant Pseudomonas aeruginosa within the CF airway. *PLoS Pathog.*, 6(1).

31 Giraud, Antoine et al. (2008) Dissecting the genetic components of adaptation of Escherichia coli to the mouse gut. *PLoS Genet.*, 4(1), pp. 0052–0061.

32 Freter, R. et al. (1983) Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infect. Immun.*, 39(2), pp. 676–685.

33 Freter, Rolf et al. (1982) Survival and Implantation of Escherichia coli in the Intestinal Tract Survival and Implantation of Escherichia coli in the Intestinal Tract. *Infect. Immun.*, 39(2), pp. 686–703.

34 Hoskins, Lansing C et al. (1985) Mucin Degradation in Human Colon Ecosystems Antigens and Oligosaccharides from Mucin Glycoproteins. *Cultures*, 75(March), pp. 944–953.

35 Henrissat, Bernard et al. (1997) Structural and sequece-based classification of glycoside hydrolases. *Curr. Opin.*

Chem. Biol., 7, pp. 637–644.

36 Zoetendal, Erwin G. et al. (2002) Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl. Environ. Microbiol.*, 68(7), pp. 3401–3407.

37 Leatham-Jensen, Mary P. et al. (2012) The streptomycin-treated mouse intestine selects Escherichia coli envZ missense mutants that interact with dense and diverse intestinal microbiota. *Infect. Immun.*, 80(5), pp. 1716–1727.

38 Wadolkowski, E. A. et al. (1988) Colonization of the streptomycin-treated mouse large intestine by a human fecal Escherichia coli strain: Role of adhesion to mucosal receptors. *Infect. Immun.*, 56(5), pp. 1036–1043.

39 Fabich, Andrew J. et al. (2008) Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine. *Infect. Immun.*, 76(3), pp. 1143–1152.

40 Henrissat, Bernard et al. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.*, 7, pp. 637–644.

41 Miranda, Regina L et al. (2004) Glycolytic and Gluconeogenic Growth of coli K-12 (MG1655) in the Mouse Intestine Glycolytic and Gluconeogenic Growth of Escherichia coli O157 : H7 (EDL933) and E . coli K-12 (MG1655) in the Mouse Intestine. *Infect. Immun.*, 7(3), pp. 1666–1676.

42 Autieri, Steven M. et al. (2007) L-fucose stimulates utilization of D-ribose by Escherichia coli MG1655 ??fucAO and E. coli Nissle 1917 ??fucAO mutants in the mouse intestine and in M9 minimal medium. *Infect. Immun.*, 75(11), pp. 5465–5475.

43 Arumugam, M et al. (2011) Enterotypes of the human gut microbiome. *Nature*, 473(7346), pp. 174–180.

44 Subramanian, Sathish et al. (2015) Cultivating healthy growth and nutrition through the gut microbiota. *Cell*, 161(1), pp. 36–48.

45 Steidler, Lothar et al. (2011) Treatment of Murine Colitis by Lactococcus Secreting Interleukin-10. *Adv. Sci.*, 289(5483), pp. 1352–1355.

46 Meibom, Karin L. et al. (2010) The unraveling panoply of Francisella tularensis virulence attributes. *Curr. Opin. Microbiol.*, 13(1), pp. 11–17.

47 Barroso-Batista, João et al. (2014) The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. *PLoS Genet.*, 10(3).

48 De Paepe, Marianne et al. (2011) Trade-Off between Bile Resistance and Nutritional Competence Drives Escherichia coli Diversification in the Mouse Gut. *PLoS Genet.*, 7(6), p. e1002107.

49 Borody, Thomas J et al. (2012) Fecal microbiota transplantation and emerging applications. *Nat. Rev. Gastroenterol. Hepatol.*, 9(2), pp. 88–96.

50 Leatham, Mary P. et al. (2009) Precolonized human commensal Escherichia coli strains serve as a barrier to E. coli O157:H7 growth in the streptomycin-treated mouse intestine. *Infect. Immun.*77(7), pp. 2876-2886.

51 Gordo, Isabel et al. (2011) Fitness Effects of Mutations in Bacteria. *J. Mol. Microbiol. Biotechnol.*, 21(1–2), pp. 20–35.

52 Martin, William F. et al. (2016) Early microbial evolution: The age of anaerobes. *Cold Spring Harb. Perspect. Biol.*, 8(2).

53 Lourenço, Marta Mansos (2014) Evolution of Escherichia coli in the mouse gut. (Master's Thesis). Retrieved: http://hdl.handle.net/10451/15760.

54 Barrick, Jeffrey E et al. (2013) Genome dynamics during experimental evolution. *Nat. Rev. Genet.*, 14(12), pp. 827–39.

55 Paul, Ligi et al. (2007) Integration of regulatory signals through involvement of multiple global regulators: control of the Escherichia coli gltBDF operon by Lrp, IHF, Crp, and ArgR. *BMC Microbiol.*, 7(72), p. 2.

56 Chen, C. F. et al. (1997) Metabolic regulation of Irp gene expression in Escherichia coli K-12. *Microbiology*, 143(6), pp. 2079–2084.

57 E B Newman, and et al. (2003) Leucine-Responsive Regulatory Protein: A Global Regulator of Gene Expression in E. Coli. *Annurev.Mi.*, 49, pp. 747–75.

58 Newman, E B et al. (1995) Leucine-responsive regulatory protein: a global regulator of gene expression in E. coli. *Annu. Rev. Microbiol.*, 49, pp. 747–75.

59 Golby, Paul et al. (1998) Transcriptional Regulation and Organization of the dcuA and dcuB Genes , Encoding Homologous Anaerobic C 4 -Dicarboxylate Transporters in Escherichia coli Transcriptional Regulation and Organization of the dcuA and dcuB Genes , Encoding Homologous Anaerob. *J. Bact.*, 180(24), p. 6586.

60 Woods, Sally A et al. (1987) Differential roles of the Escherichia coli fumarases and fnr-dependent expression of fumarase B and aspartase. *FEMS Microbiol. Lett.*, 48, pp. 219–224.

61 Beyer, Lydia et al. (2013) Coordination of FocA and pyruvate formate-lyase synthesis in escherichia coli demonstrates preferential translocation of formate over other mixed-acid fermentation products. *J. Bacteriol.*, 195(7), pp. 1428–1435.

62 Constantinidou, Chrystala et al. (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as Escherichia coli K12 adapts from aerobic to anaerobic growth. *J. Biol. Chem.*, 281(8), pp. 4802–4815.

63 Leonhartsberger, Susanne et al. (2002) The molecular biology of formate metabolism in enterobacteria. *J. Mol. Microbiol. Biotechnol.*, 4(3), pp. 269–76.

64 Jones, Shari A. et al. (2011) Anaerobic Respiration of Escherichia coli in the Mouse Intestine. *Infect. Immun.*, 79(10), pp. 4218–4226.

65 Nobelmann, Barbara et al. (1996) Molecular analysis of the gat genes from Escherichia coli and of their roles in galactitol transport and metabolism. *J. Bacteriol.*, 178(23), pp. 6790–6795.

66 Faber, Franziska et al. (2016) Host-mediated sugar oxidation promotes post-antibiotic pathogen expansion. *Nature*,

534(7609), pp. 697–699.
67 MA. Mandrand-Berthelot, G. Condemine, and N. Hugouvieux-Cotte-Pattat (2004) Catabolism of hexuronides, hexuronates, aldonates and aldarates. *EcoSal Plus*, (Module 342).
68 Barroso-Batista, João et al. (2015) Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut bacteria. *Nat. Commun.*, 6, p. 8945.
69 Lester, R L et al. (1971) Effects of molybdate and selenite on formate and nitrite metabolism in E. coli. *J Bacteriol*, 105(3), pp. 1006–1014.
70 Marques Dissertação orientada por, Daniel et al. (2009) Influência do sistema biossensor de quórum via AI-2 para a colonização do ambiente intestinal de mamíferos.
71 Datsenko, K A et al. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.*, 97(12), pp. 6640–6645.
72 Chevin, Luis-Miguel (2011) On measuring selection in experimental evolution. *Biol. Lett.*, 7(2), pp. 210–3.
73 Ambartsoumian, G et al. (1994) Altered amino acid metabolism in lrp mutants of escherichia coli k12 and their derivatives. *Microbiol. Uk*, 140(Part 7), pp. 1737–1744.
74 Stauffer, Lorraine T. et al. (1999) Role for the leucine-responsive regulatory protein (Lrp) as a structural protein in regulating the Escherichia coli gcvTHP operon. *Microbiology*, 145(3), pp. 569–576.
75 Keseler, Ingrid M. et al. (2013) EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Res.*, 41(D1), pp. 605–612.
76 Chang, Dong-Eun et al. (2004) Carbon nutrition of Escherichia coli in the mouse intestine. *Proc. Natl. Acad. Sci. U. S. A.*, 101(19), pp. 7427–32.
77 Guest, J R (1995) The Leeuwenhoek Lecture, 1995. Adaptation to life without oxygen. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 350(1332), pp. 189–202.
78 Jones, H. M. et al. (1987) Regulation of Escherichia coli fumarate reductase (frdABCD) operon expression by respiratory electron acceptors and the fnr gene product. *J. Bacteriol.*, 169(7), pp. 3340–3349.
79 Blount, Zachary D (2015) The unexhausted potential of *E. coli*. *Elife*, 4, pp. 1–12.
80 Orth, Jeffrey D et al. (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol. Syst. Biol.*, 7(535), p. 535.
81 Soupene, Eric et al. (2003) Physiological Studies of Escherichia coli Strain MG1655 : Growth Defects and Apparent Cross-Regulation of Gene Expression Physiological Studies of Escherichia coli Strain MG1655 : Growth Defects and Apparent Cross-Regulation of Gene Expression. *J. Bacteriol.*, 185(18), pp. 5611–5626.
82 Ahern, Philip P. et al. (2014) Mining the human gut microbiota for effector strains that shape the immune system. *Immunity*, 40(6), pp. 819–823.
83 Zinser, Erik R et al. (2000) Prolonged Stationary-Phase Incubation Selects for lrp Mutations in Escherichia coli Prolonged Stationary-Phase Incubation Selects for lrp Mutations in Escherichia coli K-12. *J. Bacteriol.*, 182(15), pp. 4361–4365.
84 Zinser, Erik R et al. (1999) Mutations Enhancing Amino Acid Catabolism Confer a Growth Advantage in Stationary Phase Mutations Enhancing Amino Acid Catabolism Confer a Growth Advantage in Stationary Phase. *J. Bacteriol.*, 181(18), pp. 5800–5807.
85 Farrell, Michael J. et al. (2003) The Growth Advantage in Stationary-Phase Phenotype Conferred by rpoS Mutations Is Dependent on the pH and Nutrient Environment. *J. Bacteriol.*, 185(24), pp. 7044–7052.
86 Hon Tran, Quang et al. (1997) Requirement for the proton-pumping NADH dehydrogenase I of Escherichia coli in respiration of NADH to fumarate and its bioenergetic implications. *Eur. J. Biochem*, 244, pp. 155–160.
87 Maklashina, Elena et al. (1998) Anaerobic Expression of Escherichia coli Succinate Dehydrogenase : Functional Replacement of Fumarate Reductase in the Respiratory Chain during Anaerobic Growth Anaerobic Expression of Escherichia coli Succinate Dehydrogenase : Functional Replacement of F. *J. Bacteriol.*, 180(22), pp. 5989–5996.
88 Tomasiak, Thomas M et al. (2007) Succinate as Donor; Fumarate as Acceptor. *EcoSal Plus*, 2(2), pp. 1–24.
89 Newman, E B et al. (1992) The leucine-Lrp regulon in E. coli: a global response in search of a raison d'être. *Cell*, 68(4), pp. 617–619.
90 Winter, Sebastian E et al. (2013) Host-derived nitrate boosts growth of E. coli in the inflamed gut. *Science*, 339(6120), pp. 708–11.
91 Erdman, S E et al. (2009) Nitric oxide and TNF-alpha trigger colonic inflammation and carcinogenesis in Helicobacter hepaticus-infected, Rag2-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.*, 106(4), pp. 1027–1032.
92 Tiso, Mauro et al. (2015) Nitrate Reduction to Nitrite, Nitric Oxide and Ammonia by Gut Bacteria under Physiological Conditions. *PLoS One*, 10(3), p. e0119712.
93 Long, T et al. (2009) Quantifying the integration of quorum-sensing signals with single-cell resolution. *PLoS Biol*, 7(3), p. e68.

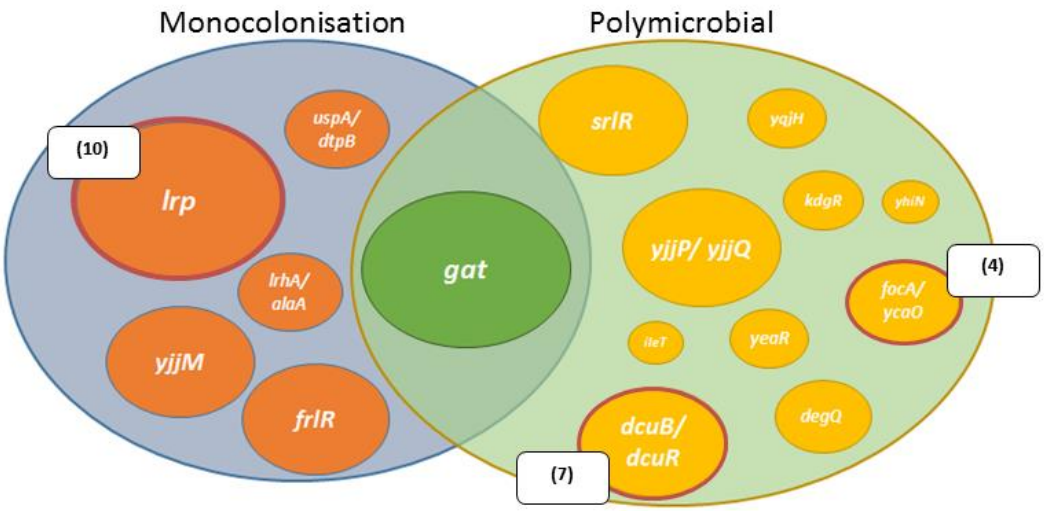
6. Annexes

6.1. Strains, plasmids and primers used

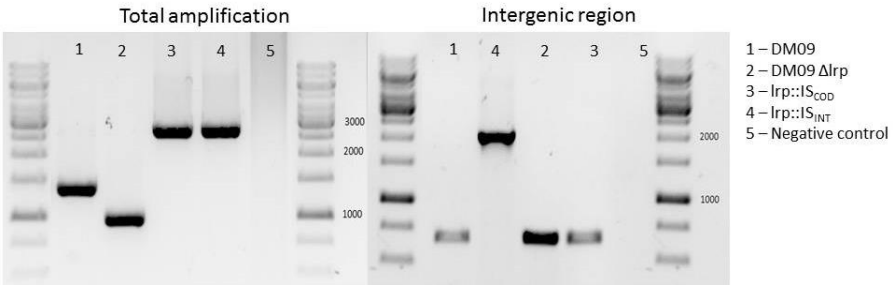
SuppTable 6.1. Strains, plasmid and primer sequences used in this work.

Element ID	Characteristics	Parent Strain	Origin
MG1655	Wild type reference	K-12	-
DM08	Δ*lacIZYA* YFP:: amp^R str^R	NO21	*Daniel Marques 2009*[70]
DM09	Δ*lacIZYA* CFP:: amp^R str^R	NO23	*Daniel Marques 2009*[70]
pKD3	pKD3 cm^R amp^R	-	This work[71]
MFP1	pKD46 $temp^S$ amp^R	MG1655	Plasmid pKD46 electroporated
MFP02	*lrp*::cm^R	MFP1	This work[71]
MFP04	Δ*lacIZYA* YFP:: amp^R str^R *lrp*::cm^R	DM08	P1 transduction JW2082
MFP06	Δ*lacIZYA* CFP:: amp^R str^R *lrp*::cm^R	DM09	P1 transduction JW2082
pCP20	cm^R amp^R	BT340	This work[71]
pTL17	kan^R	-	*Tao Long et al, 2009*[93]
MFP08	Δ*lacIZYA* YFP::amp^R str^R Δ*lrp*	MFP04	Cm^R removal with pTL17
MFP11	Δ*lacIZYA* CFP::amp^R str^R Δ*lrp*	MFP06	Cm^R removal with pTL17
MFP14	Δ*lacIZYA* YFP::amp^R str^R Δ*lrp gatZ*::kan^R	MFP08	P1 transduction JW2082
MFP16	Δ*lacIZYA* CFP::amp^R str^R Δ*lrp gatZ*::kan^R	MFP11	P1 transduction JW2082
MFP18	Δ*lacIZYA* YFP::amp^R str^R Δ*lrp* Δ*gatZ*	MFP14	Kan^R cassette with pCP20
MFP21	Δ*lacIZYA* CFP::amp^R str^R Δ*lrp* Δ*gatZ*	MFP16	Kan^R cassette with pCP20
MFP88	Δ*lacIZYA* CFP::amp^R str^R *lrp*::IS_{COD} (coding region) *gat*-negative	DM09	Clone #1 from Mouse 9 day 14 of EvoExp#3 (GF-evolved clone)
MFP89	Δ*lacIZYA* YFP::amp^R str^R *lrp*::IS_{INT} (intergenic region) *gat*-negative	DM08	Clone #1 from Mouse 1 day 16 of EvoExp#1 (GF-evolved clone)
MFP48	Δ*lacIZYA* YFP::amp^R str^R *gatZ*::kan^R	DM08	P1 transduction JW2082
MFP50	Δ*lacIZYA* CFP::amp^R str^R *gatZ*::kan^R	DM09	P1 transduction JW2082
MFP52	Δ*lacIZYA* YFP::amp^R str^R Δ*gatZ*	MFP48	kan^R removal with pCP20
MFP54	Δ*lacIZYA* CFP::amp^R str^R Δ*gatZ*	MFP50	kan^R removal with pCP20
CP1	Δ*lacIZYA* YFP::amp^R str^R *dcuBfumB**::IS_{INT} *gat*-negative	DM08	Isolate 1.2 of EvoExp (SPF-evolved clone)[47]
CP3	Δ*lacIZYA* YFP::amp^R str^R *dcuBfumB**::IS_{INT} *gatZ*::kan^R	CP1	P1 transduction JW2082
CP5	Δ*lacIZYA* YFP::amp^R str^R *dcuBfumB**::IS_{INT} Δ*gatZ*	CP3	Cassette removal with pCP20
MFP56	Δ*lacIZYA* CFP::amp^R str^R *dcuBfumB**::IS_{INT} *gat*-negative	DM09	Clone 1.10 of EvoExp (SPF-evolved clone) *11cfp*
MFP57	Δ*lacIZYA* CFP::amp^R str^R *focApflB**::IS_{INT} *gatZ*::kan^R	MFP56	P1 transduction JW2082
MFP59	Δ*lacIZYA* CFP::amp^R str^R *focApflB**::IS_{INT} Δ*gatZ*	MFP57	Cassette removal with pCP20
Primer (a)	5' ATCAGCACAGGTTGCAGGTT 3'	-	fw primer upstream of *lrp*
Primer (b)	3' TGTCTCTCTGTATTCCTTCCCT 5'	-	rev primer adjacent to *lrp* ATG
Primer (c)	3' GCGGCCGCTACTTAACTTTG 5'	-	rev primer downstream of *lrp*

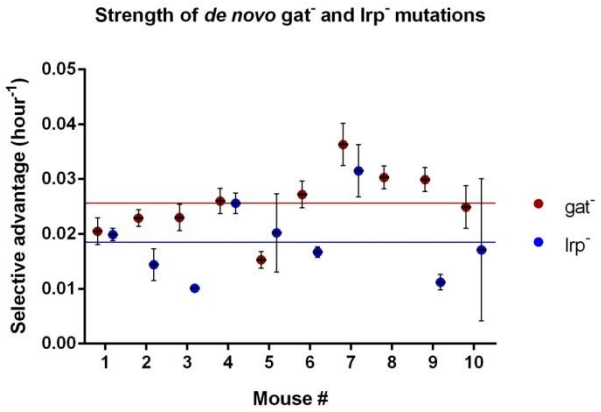
6.2. **Supplementary Figures**



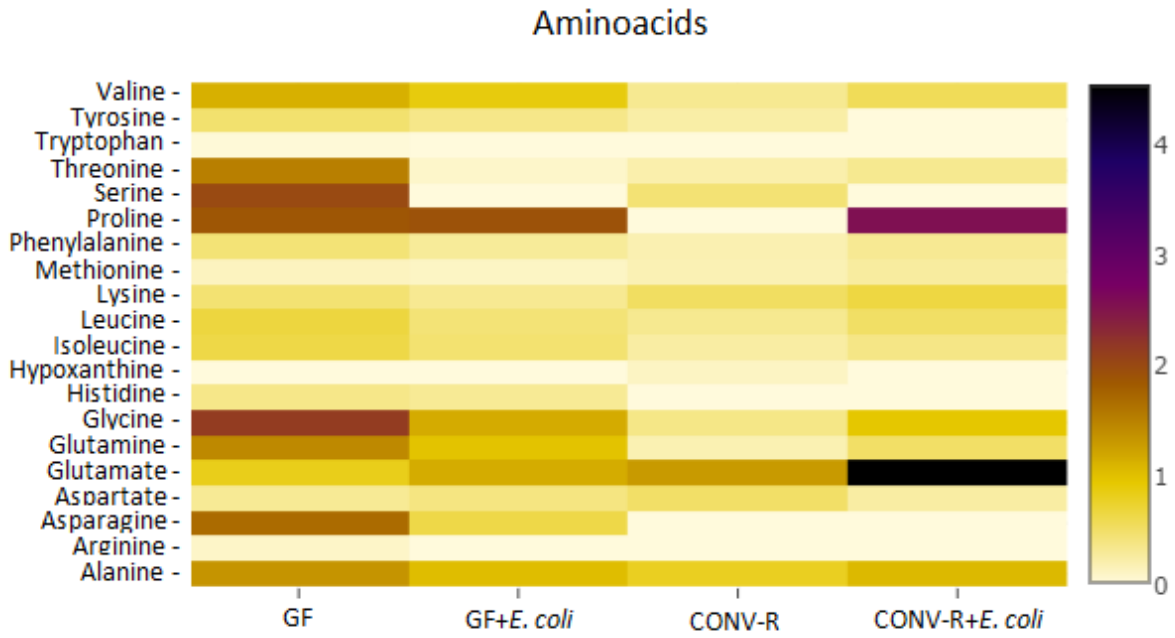
SuppFig. 6.1. Adaptive targets during adaptation to the mouse GI tract. Venn diagram of all parallel mutations in *E. coli* identified by population sequencing at day 23 of colonisation in germ-free mice (Dias, J unpublished) (n=8) and streptomycin-treated mice[47] (n=15). Bars indicate intergenic regions between indicated genes. Red circles identify the targets chosen in this work, with parenthesis indicating the number of independent mice where the same gene(s) were targeted.



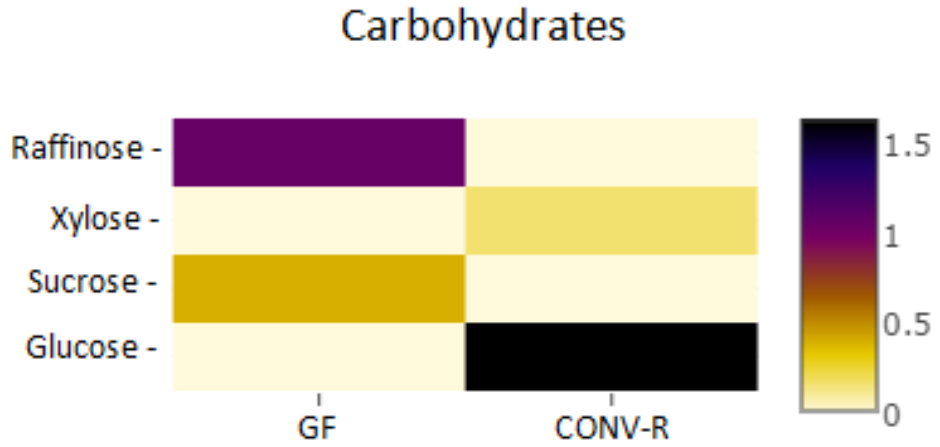
SuppFig. 6.2. Genotype of *lrp*::IS isolated mutants. The 2 types of insertions in the *lrp* gene in isolated clones represented with *lrp*::IS$_{COD}$ strain MFP88 (lane 3) and *lrp*::IS$_{INT}$ MFP89 (lane 4). A WT fragment size with primers ***a*** and ***c*** (primers as indicated in fig.3.1), indicates no IS (lane 1); a larger band indicates the presence of IS (lanes 3 and 4); a larger band, but a WT size for the intergenic region with primer set ***a*** and ***b***, indicates an IS within the open reading frame (lane 3); and larger fragments with both pairs of primers indicates an IS in the upstream intergenic region (lane 4). Δ*lrp* included for comparison (lane 2).



SuppFig. 6.3. Selective strength of *lrp* and *gat* mutations. The strength of selection during emergence of *lrp* and *gat*, loss of function mutants in mono colonised mice was estimated by taking the linear slopes of ln(x/1-x) +/- 2se, where x are frequencies from red and blue temporal dynamics of figure.3.3. Lines represent average of all mice.



SuppFig. 6.4. Metabolomic identification of aminoacid profiles *in vivo*. Caecum contents of germ-free (GF) and conventionally-raised mice (CONV-R, polymicrobial) or, the same, with *E. coli* (GF+*E. coli* and CONV-R+*E. coli*), after 28 days of colonisation were analysed by 1H NMR, (Pinto, C unpublished). Units (dimensionless) obtained from ratio of the concentration measured for each metabolite over the concentration of an internal control (DSS) present in each sample. n=2-9.



SuppFig. 6.5. Metabolomic identification of carbohydrate profiles *in vivo*. Caecum contents of germ-free (GF) and conventionally-raised mice (CONV-R, polymicrobial) or, the same, with *E. coli* (GF+*E. coli* and CONV-R+*E. coli*), after 28 days of colonisation, were analysed by 1H NMR, (Pinto, C unpublished). Units (dimensionless) obtained from ratio of the concentration measured for each metabolite over the concentration of an internal control (DSS) present in each sample. n=2-9.