

Factors affecting the intestinal microbiome homeostasis

Dissertation

In fulfillment of the requirements for the degree “Dr. rer. nat.”
of the Faculty of Mathematics and Natural Sciences
at the Christian-Albrechts-University of Kiel

Submitted by

Richa Bharti

May, 2017

Kiel, Deutschland

First examiner: Prof. Dr. Philip Rosenstiel

Second examiner: Prof. Dr. Thomas Roeder

Date of oral examination: 05, September, 2017

Approved for print on: 05, September, 2017

Approved by: Prof. Dr. Natascha Oppelt, Dean

अनेकसंशयोच्छेदि, परोक्षार्थस्य दर्शकम् ।
सर्वस्य लोचनं शास्त्रं, यस्य नास्त्यन्ध एव सः ॥

It blasts many doubts, foresees what is not obvious ।
Science is the eye of everyone, one who hasn't got it, is like a blind ॥

The Rig Veda

For my Dad and Mom

Table of Contents

Abbreviations	i
List of tables	iv
List of schematics.....	iv
List of figures.....	v
1 The intestinal microbiome.....	1
1.1 Intestinal physiology.....	2
1.2 Intestinal epithelia and microbiota	3
1.3 Inflammatory Bowel Disease (IBD).....	6
1.3.1 Crohn’s Disease (CD).....	8
1.3.2 Ulcerative Colitis (UC).....	8
1.3.3 Role of <i>Atg16l1</i> in the pathophysiology of IBD	9
1.4 Next generation sequencing.....	11
1.4.1 Roche (454) GS FLX sequencing.....	12
1.4.2 Illumina sequencing.....	13
1.5 16S rRNA sequencing.....	15
1.6 Metagenomics	16
1.6.1 Metagenomics and gut microbiome	19
1.7 Objectives of the study.....	21
2 Material and Methods.....	22
2.1 Study groups and sample collection.....	22
2.1.1 Collection of appendix samples.....	22
2.1.2 Collection of mice samples.....	22
2.2 DNA extraction and amplification	23
2.2.1 DNA extraction from samples	23
2.2.2 Amplification of DNA	23
2.3 RNA extraction and cDNA synthesis.....	24
2.3.1 RNA extraction.....	24
2.3.2 cDNA synthesis	25
2.4 Sequence pre-processing	25
2.4.1 Quality control.....	25
2.4.2 Specific pre-processing steps for 454 pyrosequencing	26

Table of Contents

2.4.3	De-multiplexing (454/Roche) and assembly (MiSeq)	26
2.4.4	Downstream pre-processing	26
2.5	Sequence analysis.....	27
2.5.1	OTU based analysis.....	27
2.6	Sequence validation	32
2.6.1	TaqMan quantitative PCR.....	33
2.7	Histological analysis and scoring	33
2.8	Colon mucosa transcriptome	34
2.9	Enrichment analysis of differentially expressed genes	35
2.10	Metagenomics library preparation and Metagenomics analysis	35
3	Microbiome variation in the inflamed and non-inflamed appendix	38
3.1	Background.....	38
3.2	Results	39
3.2.1	Bacterial taxonomy and diversity estimation.....	39
3.2.2	Phylotype analysis	41
3.2.3	Compositional differences in bacterial communities.....	43
3.3	Discussion	45
4	Effect of <i>Atg16l1</i> on microbial resilience after antibiotic stress	47
4.1	Background.....	47
4.2	Results	49
4.2.1	Unaltered baseline intestinal microbiota in <i>Atg16l1</i> ^{ΔIEC} mice	49
4.2.2	Antibiotic treatment induces colitis-like symptoms independent of the genotype	51
4.2.3	Alpha diversity variations in <i>WT</i> and <i>Atg16l1</i> ^{ΔIEC} mice after antibiotic treatment	52
4.2.4	Reciprocal enrichment of <i>Bacteroidetes</i> and <i>Firmicutes</i> after AB treatment	56
4.2.5	Assessment of pathobionts infection in animals.....	58
4.2.6	Analysis of indicator species affecting binary split of community types.....	59
4.2.7	An interaction of bacterial community type and genotype and its association with the inflammatory reaction in the ileal mucosa after antibiotics.....	59
4.2.8	Colon mucosa transcriptome analysis shows distinct subsets of genes regulated by <i>Atg16l1</i> deficiency mice and bacterial community types	62
4.3	Discussion	65
5	Role of <i>Atg16l1</i> in microbial variations during pregnancy and postpartum recovery	68
5.1	Background.....	68
5.2	Results	70
5.2.1	Study design	70

Table of Contents

5.2.2	<i>Atg16l1</i> deficiency produces marked microbial changes in pregnancy.....	70
5.2.3	Compositional differences in microbiota of <i>Atg16l1</i> mutants and control mice.....	73
5.2.4	Microbial compositional dynamics from pregnancy to post-pregnancy.....	76
5.2.5	Physiological features of nulliparous and pregnant mice	78
5.2.6	Analysis of inflammation states of pregnant and non-pregnant mice.....	79
5.2.7	Quality assessment and bacterial profiling of metagenomics data	79
5.2.8	Clustering and functional profiling of metagenomics samples.	82
5.3	Discussion	90
6	Conclusion and outlook	95
6.1	Conclusion	95
6.2	Outlook	97
7	Summary (English).....	99
8	Zusammenfassung (Deutsch)	101
9	References.....	104
10	Annexure	121
11	Curriculum vitae	138
12	Declaration	139
13	Publications	140
14	Acknowledgements	141

List of Abbreviations

Abbreviation	Full form
AMOVA	Analysis of molecular variance
ATP	Adenosine triphosphate
AB	Antibiotic
AMP	Antimicrobial peptide
ATG16L1	Autophagy Related 16-Like 1
BAC	Bacterial artificial chromosome
BT	Bacteroidetes-type
bp	Base pairs
BLAST	Basic local alignment search tool
BC	Bray-Curtis
CCA	Canonical correlation analysis
CDS	Coding DNA sequences
CCD	Coiled-coil domain
cDNA	Complementary DNAPCR
CD	Crohn's disease
D00/d00	Day 00
D21/d21	Day 21
D56/d56	Day 56
°C	Degree Celsius
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate (Deoxynucleotide)
dUTP	Deoxyuridine Triphosphate
DAI	Disease Activity Index
dsDNA	Double-stranded DNA
ER	Endoplasmic reticulum
E.coli	Escherichia coli
FELASA	Federation of European Laboratory Animal Science Association
FT	Firmicutes-type

List of Abbreviations

GIT	Gastrointestinal tract
gDNA	Genomic DNA
GAPDH	Glyceraldehyde 3-Phosphate dehydrogenase
IVC	Individually ventilated cages
IBD	Inflammatory bowel disease
ICAM-1	Intercellular adhesion molecule-1
IFN- γ	Interferon gamma
IL	Interleukin
IEC	Intestinal epithelial cells
KO	Knockout
LPS	Lipopolysaccharides
Mb	Megabase
μ l	Microliter
μ M	Micromolar
ml	Milliliter
mM	Millimolar
min	Minute
APwk1	After pregnancy week 1
M1	Month 1
M2	Month 2
M3	Month 3
MALT	Mucosa-associated lymphoid tissue
MDA	Multiple displacement amplification
MNV	Murine Norovirus
Ng	Nanogram
NGS	Next generation sequencing
NO	Nitric oxide
NOS2	Nitric oxide synthase
NF κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NOD2	Nucleotide-Binding Oligomerization Domain Containing 2
Otu	Operational taxonomic unit
PRR	Pattern recognition receptors

List of Abbreviations

PTP	Picotiter plate
PCR	Polymerase chain reaction
PMN	Polymorphonuclear leukocytes
PCA	Principal Component analysis
PCoA	Principal coordinates analysis
<i>P</i> -value	Probability value
PGHS-2	Prostaglandin-endoperoxide synthase 2
qPCR	Quantitative PCR
ROS	Reactive oxygen species
RDA	Redundancy analysis
RNA	Ribonucleic acid
RDP	Ribosomal Database Project
rRNA	Ribosomal RNA
RT	Room temperature
rpm	Rounds per minute
SFB	Segmented filamentous bacteria
SCFAs	Short-chain fatty acids
SNP	Single nucleotide polymorphism
Spp.	Species
SD	Standard deviation
SEM	Standard error of the mean
S	Svedberg unit
Th	T helper cell
Treg	T regulatory cells
TGFβ1	Transforming growth factor beta 1
T1	Trimester 1
T2	Trimester 2
T3	Trimester 3
TNFα	Tumor necrosis factor
UC	Ulcerative colitis

List of tables

Table 1 Role of the gut microbiota in inflammatory disease in human and model organisms.....	6
Table 2 Overview of the studied projects.	23
Table 3 Overview of the parameters used for sequence alignment and classification.	27
Table 4 Depiction of sample wise diversity estimation at 97% similarity level.	40
Table 5. Metacommunities estimation based on DMM model for pregnancy study.	76

List of schematics

Schematic 1 Flow diagram depicting the major analysis steps for MiSeq/454 data.....	28
Schematic 2 Flow diagram representing the steps followed in downstream analysis.	31
Schematic 3 Typical workflow for the analysis of metagenomics sequencing data.	37

List of figures

Figure 1-1 Distribution of microbiota in GIT and how they affect general physiological processes.	2
Figure 1-2 Barrier function of the intestinal epithelial cells (IECs).	4
Figure 1-3 A compilation of susceptibility/risk genes found associated with inflammatory bowel disease-IBD.	8
Figure 1-4 An overview of the cellular autophagy pathway and role of <i>Atg16l1</i>	10
Figure 1-5 The Roche/454 sequencing approach.	13
Figure 1-6 Sequencing by synthesis approach of Illumina sequencing technology.	14
Figure 1-7 A schematic of the 16S rRNA gene showing locations of major sequencing primers.	15
Figure 1-8 A flow diagram indicating major steps in a metagenomics study.	17
Figure 3-1 Boxplot showing alpha diversity indices.	40
Figure 3-2 Taxonomic classification in the non-inflamed and inflamed appendix samples.	42
Figure 3-3 Venn diagram showing unique and overlapping OTUs in non-inflamed and inflamed samples.	43
Figure 3-4 Sample wise comparative Venn diagrams depicting unique and overlapping OTUs in non-inflamed and inflamed samples.	44
Figure 3-5 A distance based principal component analysis of non-inflamed and inflamed samples.	45
Figure 4-1 Baseline bacterial diversity in <i>WT</i> and <i>Atg16l1^{ΔIEC}</i> mice (n=11/12).	50
Figure 4-2 Taxonomic classification at Phylum level for <i>WT</i> and <i>Atg16l1^{ΔIEC}</i> samples at Day00.	51
Figure 4-3 Longitudinal experimental setup, clinical symptoms and microbial diversity distribution in <i>WT</i> and <i>Atg16l1^{ΔIEC}</i> mice (n=11/12) from Day00 to Day56.	52
Figure 4-4 Longitudinal representation of α -diversity at three time points Day00, Day21 and Day56 for <i>WT</i> (n=12) and <i>Atg16l1^{ΔIEC}</i> (n = 11).	53
Figure 4-5 Venn diagram, showing shared OTUs at different time points - (Day00, Day21, Day56).	54
Figure 4-6 Constrained ordination analysis for <i>WT</i> (n = 12), <i>Atg16l1^{ΔIEC}</i> (n = 11).	55
Figure 4-7 Interaction plots showing two way interactions.	56
Figure 4-8 Temporal development of bacterial phyla.	56
Figure 4-9 Relative bacterial distribution at Day00 and Day56.	57
Figure 4-10 Constrained ordination analysis showing community-type based clustering at Day56.	58
Figure 4-11 Paneth cell morphology and stem cell number in <i>WT</i> and <i>Atg16l1^{ΔIEC}</i> ; n=12/11.	60
Figure 4-12 Ileal inflammation in <i>WT-Bacteroidetes</i> type n=4; <i>WT-Firmicutes</i> type n=8; <i>Atg16l1^{ΔIEC}-Bacteroidetes</i> type n=6; <i>Atg16l1^{ΔIEC}-Firmicutes</i> type n=5.	61
Figure 4-13 Colon mucosa transcriptome.	63
Figure 4-14 Enrichment analysis of differentially expressed genes.	64
Figure 5-1 Study design, bacterial diversity and abundance estimation.	72
Figure 5-2 Distance based β -diversity analysis.	73
Figure 5-3 Correlation based network showing genotype associated (red and green)G unique OTUs and common OTUs (grey) at Trimester 3 (T3).	75
Figure 5-4 Distance based diversity and metastats analysis.	77
Figure 5-5 Histological phenotypes for nulliparous and pregnant mice.	78

List of figures

Figure 5-6 Expression levels of the cytokines in nulliparous and pregnant <i>Atg16l1^{fl/fl}</i> and <i>Atg16l1^{ΔIEC}</i> mice.	79
Figure 5-7 Quality check and rarefaction curves for metagenomics reads.....	80
Figure 5-8 Representative hit based bacterial profile of metagenomics reads.	81
Figure 5-9 Distance based clustering of metagenomics.....	83
Figure 5-10 Level 1 functional categories distribution in metagenomics samples.	83
Figure 5-11 Level 2 functional categories distribution in metagenomics samples.	84
Figure 5-12 Level 4 functional enrichment analysis in metagenomics samples.	84
Figure 5-13 Linear discriminant Effect Size (LEfSe) analysis at level 4 KEGG functional categories.....	86
Figure 5-14 KEGG pathways for Glycolysis and Gluconeogenesis.....	88
Figure 5-15 KEGG pathways showing Fatty acid biosynthesis.	89
Figure 5-16 KEGG pathways showing Glycine, Serine and Threonine metabolism.	90

1 The intestinal microbiome

The importance of the intestines in determining human health has been historically rooted far back as Hippocrates, who stated that “*death sits in the bowels*” and “*bad digestion is the root of all evil*” in 400 B.C (Hawrelak and Myers, 2004). It has been long postulated that intake of excess food or wrong food types result in intestinal toxins. Fermentation of these toxins promotes the growth of bacteria that subsequently leads to disease (Dudgeon, 1926; Windey et al., 2012). Nevertheless, it is now known that the human body is co-inhabited by an immense population of various bacteria, archaea, viruses and other microorganisms. The collection of microbes coexisting with their hosts has been referred as the “*microflora*” or “*microbiota*” (Cani and Delzenne, 2007; Zoetendal et al., 2004). This microbiota flourishes on human skin, gastrointestinal, genitourinary as well as the respiratory tracts (Costello et al., 2009). The microbial components of human microbiota exhibit exceptionally high level of adaptation to their habitat as well as their co-habitants, the phenomenon now termed as “*microevolution*” (Boon et al., 2014; Ley et al., 2006). The gastrointestinal tract (GIT) is the most heavily colonized organ with more than 70% of the entire human microbiota (**Figure 1-1a**). The microbiota is involved in a variety of functions including digestion, promotion of gut motility, nutrient assimilation, xenobiotic metabolism, production of vitamins, short-chain fatty acids (SCFAs) and polyamines as well as modulation of immune response. Thus, owing to its enormous functionalities aiding human health and survival, the microbiome is often considered as an additional organ.

There are several factors that could harm the microbes associated with a healthy gut such as psychological and physical stress, radiation, altered GIT peristalsis, and dietary changes. Besides, gut-specific pathogens, food-borne toxins, antibiotics and certain physiological states such as pregnancy and complex disorders including diabetes and cancer are now been explored for their association with gut microflora (DiGiulio et al., 2015; Gibson et al., 2015; Louis et al., 2014; Maranduba et al., 2015; Moreno-Indias et al., 2014a). In the past few years, advances in sequencing technology greatly improved our understanding of the role of gut microbes in complex diseases such as intestinal bowel diseases (IBD), cancer and diabetes (Kostic et al., 2014; Shreiner et al., 2015). The microbial compositions exhibit a general reduction in bacterial diversity and specific shifts towards certain microbial groups in diseased conditions are reported. Thus, most of the current researches efforts now focus on

identify the functional role of microbiota on the mammalian gut. The microbiome affects the intestine through bacterial signaling molecules and other metabolic products. These molecules can also enter the circulation and get transported to various organs and induce multiple effects in depending on the target organ as shown in **Figure 1-1b**. Thus, identification of factors that result in detrimental changes/dysbiosis in the microbiota composition has become increasingly important in the past decades.

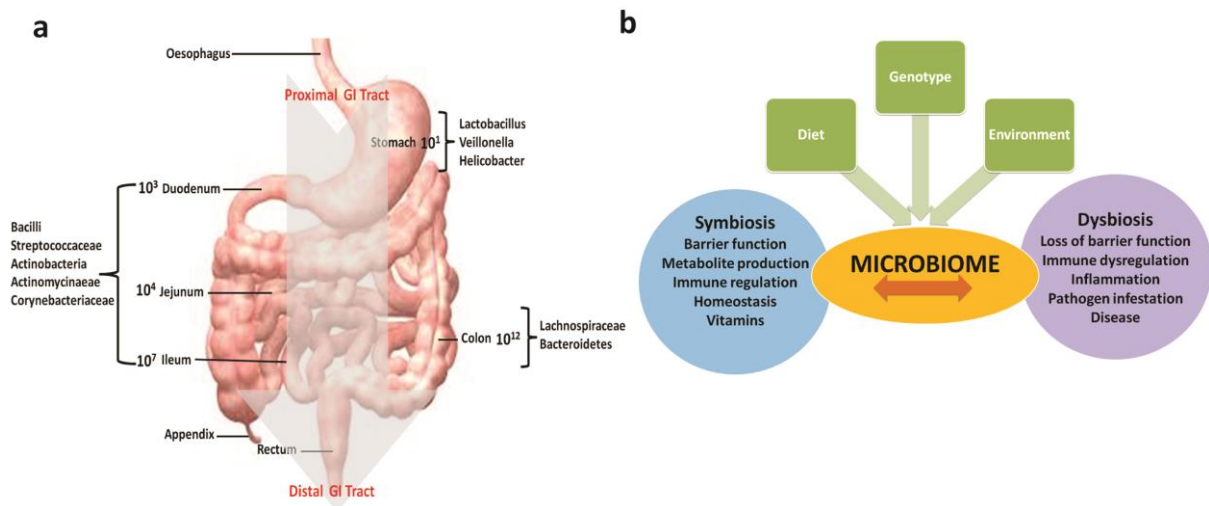


Figure 1-1 Distribution of microbiota in GIT and how they affect general physiological processes.

(a) Variation in microbial composition and numbers across the gastrointestinal tract. (b) The synergistic effect of host genotype, environment and diet in determining regulatory effects of gut microbiota. Figures adapted from (Sartor, 2008) and (Korecka and Arulampalam, 2012).

1.1 Intestinal physiology

The gastrointestinal tract (GIT) is mainly associated with digestion of food and extraction of useful components from it, while continuously eliminating the waste products. The entire system is controlled by a hormonal cascade that is triggered as soon as food enters the GIT and regulates several physiological responses such as acid secretion, increased gut motility and enzyme release (Schneeman, 2002). The released nutrients are then transported to the liver and further broken down, assimilated or distributed across the body.

The majority of GIT is covered by a mucus lining, which is secreted throughout the lumen and acts as a protective layer and mechanical barrier (Salminen et al., 1998). Structurally, the gastrointestinal tract consists of four major anatomical regions *i.e.* the oesophagus, the stomach, the small intestine and the large intestine or colon as shown in **Figure 1-1a**. These four compartments are involved in breaking down of food, its adsorption

and assimilation through mechanical and biochemical processes (Hunt et al., 2015; Komuro, 2006). The intake of food induces sensory response to stimulate the flow of saliva that lubricates the bolus and contains several digestive enzymes. The bolus reaches the stomach by the sphincter actions of the oesophagus. The motility of the stomach continues to process the food with gastric juices containing acid as well as digestive enzymes. The stomach helps in disintegrating food into smaller particles before passage to intestine. After the digested food enters small intestine, peristaltic movement allows further mixing of digestive ingredients of intestine, including the pancreatic enzymes and bile acids. The food macronutrients break down into macromolecular subunits that can be taken up by the absorptive system of the intestinal cells (Kararli, 1995).

Since the GIT executes several crucial physiological functions in human body incorrect metabolic responses in GIT may consequently be associated with an increased risk for diseases. The preciseness of GIT function has been shown to be associated with its barrier function. This barrier function is majorly attributed to the epithelial lining the mucosa in the GIT. These epithelial cells establish and maintain a barrier that acts as central mediator of interactions between body and the external environment (Cesta, 2006). Mucosa-associated lymphoid tissue (MALT) is scattered along mucosal linings in the human body and constitutes the most extensive component of human lymphoid tissue. These surfaces include the tonsils, the Peyer's patches within the small intestine, and the vermiform appendix that help in protecting the body from variety of antigens (Thome, 2008). The mucosal permeability is extremely adaptable and gets regulated by the response to extracellular stimuli such as nutrients, cytokines and pathogens (Arrieta et al., 2006; Gibson, 2004). Apart from the internal immunological responses in the form of cytokines and external stimuli such as food, the microbial component associated with the mucosal lining also plays a crucial role in maintaining epithelial barrier function (Baumgart and Dignass, 2002).

1.2 Intestinal epithelia and microbiota

The intestinal epithelium forms the largest mucosal surface of the entire human body and acts as an interface between the external environment and the host. Moreover, intestinal epithelial cells (IECs) act as a mechanical barrier that separates the host's internal

environment from the external environment (Booth and Potten, 2000; Marchiando et al., 2010) (**Figure 1-2**). This external environment mainly consists of microbes that reside near the mucosal lining and never enter the epithelial cells. The intestinal mucus lining serves as a nutrient source but also presents a habitable milieu accommodating microbial adhesion via lectins and glycosidases. Epithelial cells on the other hand play a crucial role in ion transport, fluid absorption and secretion. Interestingly, the complex carbohydrates generated during these processes are consumed by the bacterial cells growing over epithelial cells. These bacteria are involved in fermentation resulting in product such as SCFAs (short-chain fatty acids) utilized by epithelial cells, liver and muscles (Wong et al., 2006). Besides, many of these microbes are also involved in synthesis of vitamins, facilitating mineral absorption and immune stimulation.

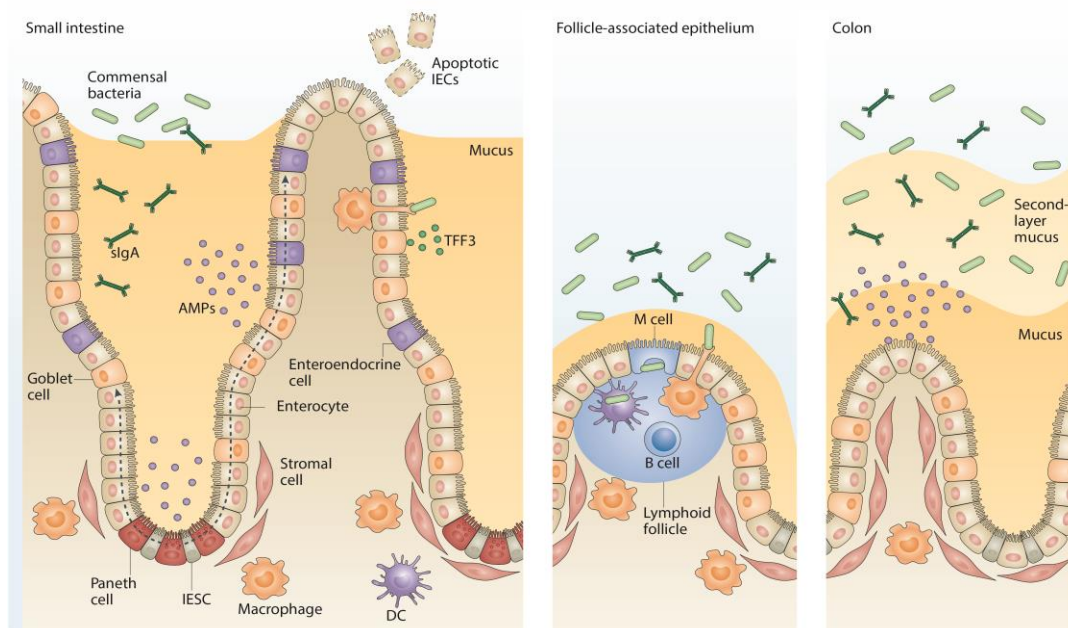


Figure 1-2 Barrier function of the intestinal epithelial cells (IECs).

The intestinal epithelium functions as a barrier between luminal microbial communities and the mucosal immune system. Intestinal epithelial stem cells (IESC) located at the bottom of the crypt proliferate and its progenitor cells repopulate the intestinal epithelium. The dashed arrows show migration of IECs up to the crypt–villus axis. Antimicrobial proteins/peptides (AMPs) and mucus released by secretory goblet cells and Paneth cells helps in exclusion of microbes from the epithelial surface. slgA-Secretory immunoglobulin A, TFF3-Trefoil factor 3, DC-Dendritic cells. Source: (Peterson and Artis, 2014).

It is now widely accepted that the interplay between environmental triggers and host is regulated by two key elements namely, the intestinal permeability and intestinal mucosal defense (Fasano and Shea-Donohue, 2005; Groschwitz and Hogan, 2009). Studies on germ-free and conventionally raised animals have shown that microbes majorly affect mucus

thickness and composition. Compared to the normal mice, germ-free mice have fewer goblet cells and a thinner mucus lining (Bry et al., 1996). A number of *in vitro* and *in vivo* model systems have been extensively used to study the mechanisms used by *Salmonella*, *Yersinia*, *Shigella* and *Listeria* to invade host epithelial cells (Galan and Bliska, 1996). The invasion of pathogenic bacteria into IECs causes an increased expression and secretion of a number of cytokines with chemo-attractant and pro-inflammatory functions. The stimulated epithelia secrete high levels of cytokines such as IL-8, GRO α , GRO β , GRO γ and ENA-78 (Eckmann et al., 1993a; Eckmann et al., 1993b; Jung et al., 1995; McCormick et al., 1993). These cytokines belong to the C-X-C chemokine family that have chemo-attractant abilities and can also activate polymorphonuclear leukocytes (PMN). PMNs are pre-indicators of initial inflammatory response that generally result from infection of the intestinal tract with pathogenic bacteria. Furthermore, the infected human IECs also express and secrete other pro-inflammatory cytokines, including TNF α , GM-CSF, IL-1a, and IL-1 β (Jung et al., 1995; Svanborg et al., 1999). Importantly, this initial immune response is indicative only due to the short lived expression of these cytokines. Nevertheless, this indicates that epithelial cells initiate the innate inflammatory response rather than the antigen-specific response. In addition, the IEC also express a variety of cytokine receptors that include IFN- γ , IL-1, TNF α , TGF β 1, as well as IL-2, IL-4, IL-7, and IL-9 (Reinecker and Podolsky, 1995). This suggests that IECs could very well respond to the immunological signals from the underlying mucosa. Moreover, an important class of membrane adhesion molecules, the Intercellular adhesion molecule-1 (ICAM-1), is also regulated by the IECs. These molecules help in controlling the immunological responses from immune and inflammatory cells (Huang et al., 1996).

Using an *in vivo* model, it has been shown that the infection of invasive bacteria induces expression of ICAM-1 on the apical surfaces of the IECs (Kelly et al., 1992). Also, increase in expression of inducible Nitric oxide (NO) synthase (NOS2) has been reported in IEC during intestinal inflammations. This leads to accumulation of NO which plays an important role in multiple gastrointestinal functions, including blood flow and mucosal inflammation (Eckmann et al., 2000). Overall, these experimental observations suggest that IECs contain a conserved set of functional genes that get activated in response to broad variety of microbial biochemical as well as against number of infections. Unfortunately, the underlying mechanisms and the pathways involved are poorly understood till date. As a

whole, IECs acts as critical communication ports that send signals for initiating major immunological response against pathogenic invasion and loss of native microbiota.

1.3 Inflammatory Bowel Disease (IBD)

Inflammatory bowel diseases (IBDs) are one of the most complex life style diseases that involves chronic inflammation of all or a part of the human digestive tract (Kaser et al., 2010). A dramatic surge in the prevalence of IBD has been observed in Europe and North America in the second half of the twentieth century (Molodecky et al., 2012). The two main disease categories that broadly cover IBD are Crohn's disease (CD) and ulcerative colitis (UC), with overlapping idiopathic chronic inflammatory intestinal conditions. In both cases, the mucosal surface of the gut, especially the intestine is affected which leads to a long-term impairment of the gastro-intestinal structure and function (Sartor, 2006). The pathogenesis of IBD is still not completely understood.

Table 1 Role of the gut microbiota in inflammatory disease in human and model organisms. Adapted from (Kamada et al., 2013).

Commensal bacterium	Host	Genotype	Disease/Inflammation condition	References
Protective microbes against inflammatory responses				
<i>Clostridium</i> spp. clusters IV and XIVa	Mouse	Wild-type	DSS colitis	(Neel et al., 1996)
<i>Bacteroides fragilis</i>	Mouse	Wild-type	TNBS colitis	(Round et al., 2011)
<i>Bacteroides vulgatus</i>	Mouse	<i>I12</i> ^{-/-}	Spontaneous colitis	(Waidmann et al., 2003)
<i>Faecalibacterium prausnitzii</i> , <i>Papillibacter</i>	Human	NA	IBD	(Sokol et al., 2008) (Rehman et al., 2016)
Harmful microbes that promote inflammatory responses				
<i>E. coli</i>	Mouse	<i>I10</i> ^{-/-}	Spontaneous colitis	(Kim et al., 2005)
<i>Enterococcus fecalis</i>	Mouse	<i>I10</i> ^{-/-}	Spontaneous colitis	(Kim et al., 2005)
<i>B. vulgatus</i>	Rat	<i>HLA-B27</i> ⁻ <i>B2m</i> ⁻ transgenic	Spontaneous colitis	(Schultz et al., 1999)
<i>B. thetaiotaomicron</i>	Rat	<i>HLA-B27</i> ⁻ <i>B2m</i> ⁻ transgenic	Spontaneous colitis	(Rath et al., 1999)
<i>Prevotellaceae</i>	Mouse	<i>Nlrp6</i> ^{-/-} , <i>Asc</i> ^{-/-} or <i>Casp1</i> ^{-/-}	DSS colitis	(Elinav et al., 2011)
<i>Helicobacter</i> , <i>Escherichia/Shigella</i> , <i>Clostridium</i> groups	Mouse	<i>I23R</i> ^{ΔIEC}	DSS colitis	(Aden et al., 2016)

Several genetic and environmental factors such as an altered gut microbiota and enhanced intestinal permeability contribute to a collapsed intestinal immunity that concludes into chronic gastrointestinal injury. The current understanding of gut microbiota suggests that altered balances of the gut bacterial species rather than specific pathogens are responsible for the pathophysiological changes associated with several diseases (Eckburg et al., 2005). This alteration or shift in the microbial balance inside the gut is referred to as dysbiosis. The human gut microbiota consists of four major phyla namely; the *Bacteroidetes*, *Firmicutes*, *Proteobacteria* and *Actinobacteria* (Frank et al., 2007; Tong et al., 2013). Out of these, the *Bacteroidetes* and *Firmicutes* form approximately 90% of the entire bacterial phyla dominating the gut. The IBD pathophysiology is associated with a reduced diversity of the gut microbiota. Although the actual cause/s of IBD remains elusive, many recent reports have strongly suggested interplay between the host microbiota and host genetics that determines susceptibility to IBD pathogenesis (**Table 1**). In most IBD patients, a reduction in *Firmicutes* and an increase in *Proteobacteria* sub-populations have been observed (Gophna et al., 2006; Scanlan et al., 2006). Basically, the reduction in bacterial diversity observed in IBD patients is mainly associated with a decline in the diversity of *Firmicutes*.

Long term longitudinal studies examining the gut microbiota in CD patients have shown that the gut microbiota remains unstable even during remission and relapse (Martinez et al., 2008). Similarly, in CD patients, microbial dysbiosis has been reported even in patients with remission (Ott et al., 2008). Apart from this, several medications and antibiotics have been shown to dramatically increase the dysbiosis event in CD and UC patients. One strategy to unravel the reasons of dysbiosis is to compare the gut microbiome compositions in healthy individuals and IBD patients who are related to each other and are likely to share similar genetic and environmental features. Several such comparative studies have now shown that many genes such as *IL23R*, *JAK2*, *NOD2*, *ATG16L1*, *IRGM*, *ATG5*, *ECM1* and *CDH1* act as IBD susceptibility genes associated with bacterial recognition and processing (Frank et al., 2011; Joossens et al., 2011) (**Figure 1-3**). And, the altered bacterial crosstalk with these susceptibility genes affect mucosal barrier function that eventually leads to collapse of bacterial tolerance against the commensal gut bacteria in IBD.

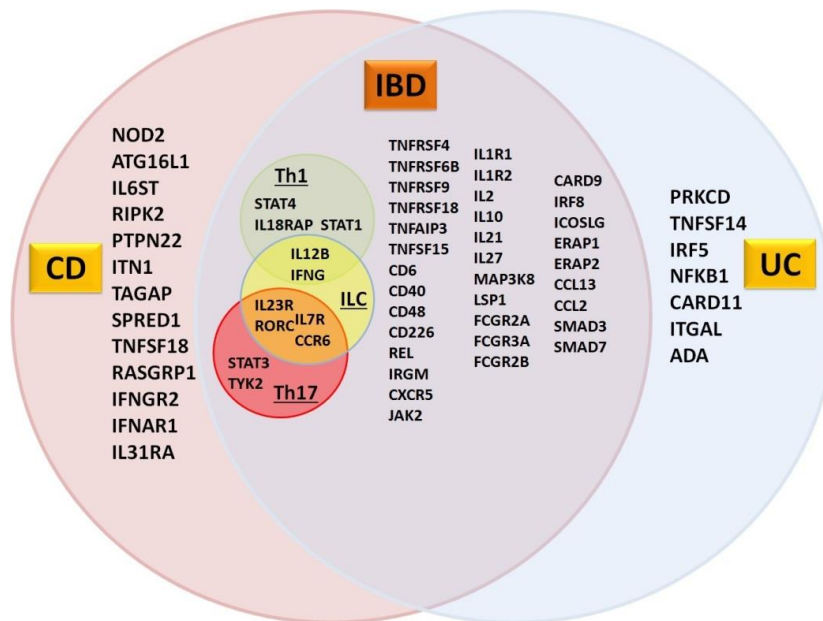


Figure 1-3 A compilation of susceptibility/risk genes found associated with inflammatory bowel disease-IBD. Some of the gene variants are specific for Crohn's disease (CD), few are specific for ulcerative colitis (UC), and the majority, including several T helper Th1 and Th17-related genes, are common to both forms of IBD. The inner circles represent the overlapping genes between both the forms of IBD such as T helper (Th)1- and Th17-related genes which are also important in the phenotype and function of innate lymphoid cells (ILCs), a cell population that is emerging as a possible central player in gut inflammation. Source: (Biancheri et al., 2013).

1.3.1 Crohn's Disease (CD)

Crohn's disease (CD) is a subentity of IBD that is characterized by chronic inflammation of the mucosa with episodic progression. CD may affect any part of the gastrointestinal tract from the mouth to the anus (Bandzar et al., 2013). The typical manifestations of CD include a discontinuous involvement of different segments of the GI tract and the development of complications such as strictures, abscesses, and fistulae. CD is mainly categorized based on the organ-specific localization of inflammation (Baumgart and Sandborn, 2012). The Ileocolic CD accounts for fifty percent of the cases and mainly involves inflammation in both the ileum and the large intestine. Crohn's ileitis and Crohn's colitis mainly affect the ileum and the large intestine respectively.

1.3.2 Ulcerative Colitis (UC)

Ulcerative colitis (UC) is a class of IBD that is characterized by inflammation and ulcers in the colon. The pathophysiology of UC is similar to CD but in contrast to CD where the entire gastrointestinal tract may get affected, it affects only the colon and rectum (Danese and Fiocchi, 2011). Basically, UC is a recurrent inflammatory and ulcerative disease of the colon

and rectum, characterized clinically by the rectal bleeding, diarrhea, cramping abdominal pain, anorexia, and weight loss.

1.3.3 Role of *Atg16l1* in the pathophysiology of IBD

Genetic susceptibility of the host remains one of the most crucial factors contributing to the pathophysiology of IBDs. In the past few decades, several new advances have occurred that have aided in understanding the genetics of human IBD. Numerous studies on IBDs have been undertaken based on single nucleotide polymorphism (SNP), candidate gene approaches and mouse experimental colitis that used transgenic and deletion (knockout) techniques (Biancheri et al., 2013). These studies reported many genes that involved in several physiological functions such as immune regulation, mucosal barrier integrity and microbial clearance and homeostasis (**Figure 1-3**).

A number of genome-wide association studies (GWAS) and metagenomics studies have helped in identifying more than 150 distinct loci that influence IBD susceptibility and are associated with a gamut of metabolic and homeostatic pathways (Van Limbergen et al., 2014). A coding variant (SNP rs2241880; T300→A) in the autophagy gene *Atg16l1* (autophagy related 16-like 1) is associated with an increased risk of CD (Hampe et al., 2007; Rioux et al., 2007). Furthermore, it has been also shown that a CD associated T300A mutation in the coding region of *Atg16l1* gene increases its degradation. This leads to reduced autophagy and contributes essentially to the pathophysiology of CD (Murthy et al., 2014).

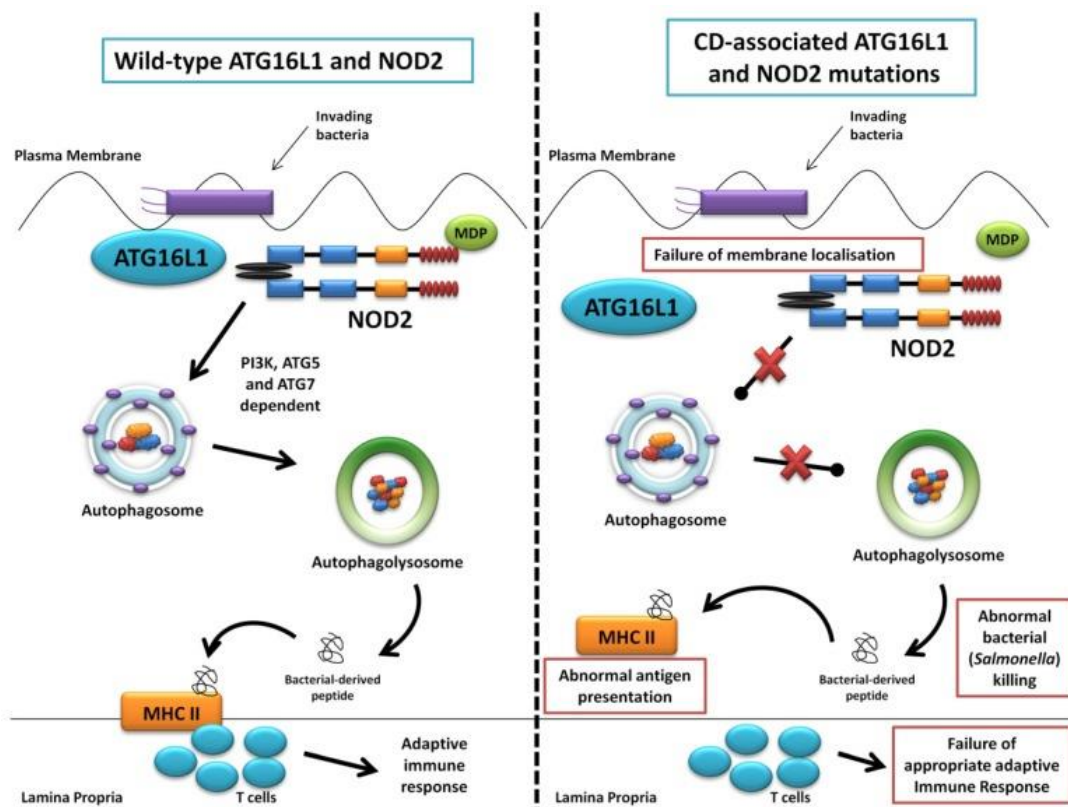


Figure 1-4 An overview of the cellular autophagy pathway and role of *Atg16L1*.

Summary diagram showing the role of ATG16L1 and NOD2 in the autophagy pathway. Figure shows the normal autophagy response with NOD2 recruiting ATG16L1 in the normal autophagosome and autophagolysosome formation, followed by MHCII antigen presentation and an appropriate adaptive immune response. The right panel shows the abnormal membrane localization of ATG16L1 with the LRR and T300A mutations result in abnormal bacterial killing due to defective antigen presentation. Source: (Henderson and Stevens, 2012).

Basically, autophagy is a conserved cellular process targeting unused/harmful cytoplasmic components towards lysosomal degradation. This helps in maintaining cellular homeostasis by recycling large protein aggregates and damaged membranes. The *Atg16l1* gene belongs to the WDR (WD repeat domain containing) gene family and is involved in the formation of mature autophagosome during cellular autophagy. *Atg16l1* along with *Atg5* and *Atg12* helps forming the autophagy complex that facilitates elongation of the initial isolation membrane that engulfs the cargo and forms the autophagosome (Salem et al., 2015) (**Figure 1-4**).

Recently, several studies were performed to identify the mechanisms through which autophagy may regulate intestinal homeostasis and the functional consequences of altered autophagy in gut tissue. Epithelial cells and dendritic cells containing CD-associated *Atg16l1* and Nucleotide-Binding Oligomerization Domain Containing 2 (*NOD2*) gene variants exhibit defects in antibacterial autophagy (Cooney et al., 2010; Travassos et al., 2010). More

specifically, the dendritic cell-associated defects were linked with an impaired ability to present exogenous antigens to CD4⁺ T cells. Together *NOD2*, *Atg16l1* and autophagy link intracellular processing and communication with the adaptive immune system. This also suggests that the genetic polymorphisms associated with IBD may affect these pathways and lead to pathophysiological symptoms. Further, abnormalities associated with CD were observed in mice with defects in autophagy, including hypomorphic *Atg16l1* (*Atg16l1*^{HM}) and IEC-specific *Atg5*-deficient mice (Cadwell et al., 2008). The Paneth cells either from *Atg16l1*^{HM} mice or from CD patients possessing *Atg16l1* (T300A variant) allele showed aberrant granule size, number and location, and reduced AMP secretion (**Figure 1-4**). Alongside, these Paneth cells also showed gain of function, as evidenced by upregulated peroxisome proliferator-activated receptor signaling (Adolph et al., 2013). All these evidences clearly implicate a crucial role of *Atg16l1* in the pathophysiology of Crohn's disease and further investigations on its role during several types of stress conditions still remain to be examined.

1.4 Next generation sequencing

During the last decade, there has been a fundamental shift from the traditionally used Sanger sequencing for gene and genome analysis (Sanger and Coulson, 1975). Earlier, the Sanger method overshadowed the sequencing domain worldwide and was used for a number of massive genome sequencing accomplishments including the human genome project. However, several limitations of Sanger method reflected the need of newer, faster and cost-effective sequencing methods (Schloss, 2008). Advent of newer strategies that mainly use a combination of high throughput experimental and bioinformatics methodologies now dominate the present day sequencing technology. These newer technologies basically rely on template preparation, sequencing and imaging, and data analysis. The unique combinations of these methodologies with their patented variations distinguish one technology from another. All together these are demarcated as the Next generation sequencing technology (NGS). Their major advantage being the ability to process millions of sequencing reads in parallel. Besides, a single instrument acts as a high throughput device and can handle a complete experiment. The conventional vector based cloning methodologies is not required that further cuts down the sequencing time. The

workflow to produce NGS reads are generated by the DNA fragments that may originate from a variety of front end processes. Further, the ability of these different sequencing platforms to create paired end fragments helps in more precise de novo genome assembly.

1.4.1 Roche (454) GS FLX sequencing

Roche 454 was the first sequencing platform that was introduced commercially in early 21st century. It utilizes the principle of pyrosequencing where a pyrophosphate molecule released on nucleotide incorporated by DNA polymerase produces light by a luciferase reaction (Ronaghi et al., 1998) (**Figure 1-5**). Sequencing is based on emulsion PCR that utilizes surfaces of several thousand agarose beads with millions of oligomers attached to them. These oligomers are complimentary to the adaptor sequences that are ligated to the DNA fragment ends during library construction. This involves a vigorously shaken oil-water mixture that helps isolating individual beads having unique DNA fragment hybridized to the oligo-decorated surface. The aqueous micelles formed by these beads incorporate other PCR reagents that initiate the DNA amplification during temperature cycling.

The sequencing reaction yields agarose beads with millions of copies of originally conjugated DNA fragment. Following this, thousands of these beads are added to the surface of 454 picotiter plate (PTP), which consists of single wells in the tips of fused fiber optic strands that hold each bead (Berglund et al., 2011). Later, smaller magnetic beads attached with active enzymes are added and the PTP is placed in the sequencer following which reagent solutions are delivered into it in a sequential fashion. The imaging of luciferase activity is utilized for recording the templates that are adding a particular nucleotide. The light produced from the luciferase reaction is directly proportional to the amount of a particular nucleotide being incorporated. Finally, the pre-added adaptor sequences are used to calibrate the level of emitted light and are used to process the downstream base-calling and sequence reading analysis. The currently available GS-FLX 454 sequencer produces an average read length of ~250bp per sample and results in ~100 Mb of sequence data per run (~7 h) (Liu et al., 2012b).

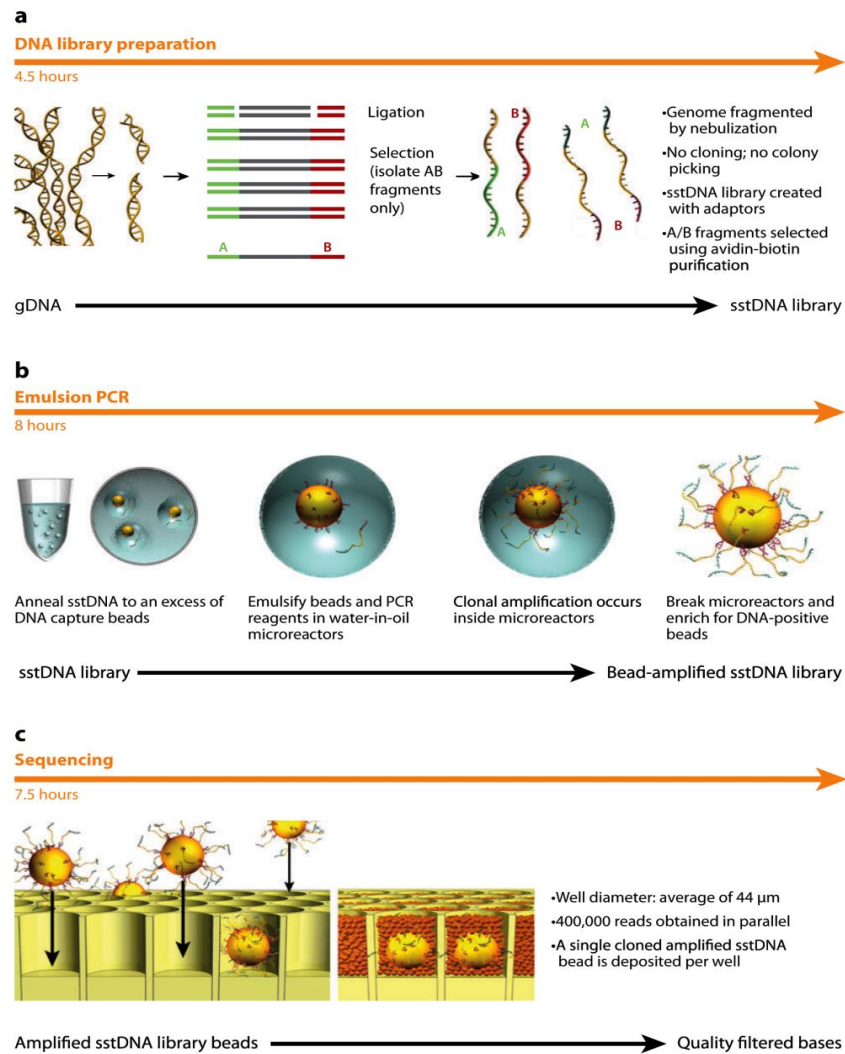


Figure 1-5 The Roche/454 sequencing approach.

Amplification of single-strand DNA copies involves (a) conjugating DNA fragments with agarose beads. The beads contain complementary oligonucleotides to the adapters at the fragment ends that are mixed in an approximately 1:1 ratio. (b) The mixture is encapsulated into aqueous micelles containing PCR reagents and is put into 96-well microtiter plate for PCR amplification. (c) The PCR reaction results into beads decorated with approximately 1 million copies of the original single-stranded fragment. The signal detection during the following pyrosequencing step helps in detecting and recording nucleotide amplification process. Source: (Mardis, 2008).

1.4.2 Illumina sequencing

The Illumina sequencing technology is based on the idea of 'sequencing by synthesis'. The technology is able to produce sequencing reads of ~150-200bp sizes from millions of surface amplified DNA segments (Berglund et al., 2011). A microfluidic cluster station having a number of flow cells is added with a mixture of single-stranded, adaptor ligated DNA fragments. There are 8 separate lanes in individual flow cells that harbor covalently attached oligos complementary to the adapter sequences. Following this, an active heating and

cooling step hybridizes the DNA to the surface attached oligos in the flow cells. The DNA fragments get amplified, by further incubation with an isothermal polymerase along with other reagents. The incorporation of fluorescent labeled nucleotides is monitored during each cycle of amplification that ends with a chemical step that removes the fluorescent group (**Figure 1-6**). The sequencing reads thus generated after each run are processed and quality filtered to remove the low quality reads.

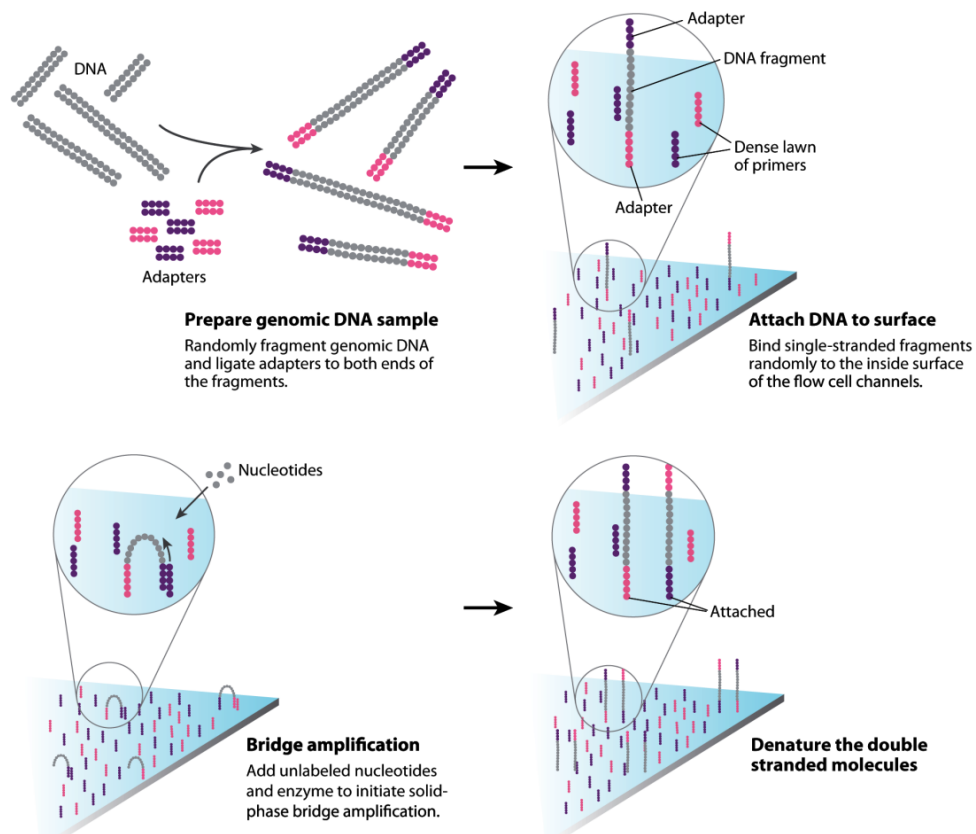


Figure 1-6 Sequencing by synthesis approach of Illumina sequencing technology.

Bridge amplification is utilized for creating cluster strands using fluorescently labelled 3'-OH blocked nucleotides. Each cycle of extension (one nucleotide) is followed by washing away of unused nucleotides and DNA polymerase and imaging of individual lanes of the flow cell. Repeated fluorescent nucleotide incorporation and bridge amplification is done for generating sequencing reads. Source: (Mardis, 2008).

1.5 16S rRNA sequencing

The 16S ribosomal RNA (rRNA) sequencing is a widely used sequencing methodology for identifying and characterizing bacterial populations present in biological and environmental samples. Over the past quarter century, this technique has become a mainstay in studying phylogeny and taxonomy of complex microbiomes that were earlier considered difficult to characterize (Woo et al., 2008). Basically, the 16S rRNA gene is a portion of the prokaryotic DNA found in all bacteria and archaea. The prokaryotic 70S ribosome of bacteria and archaea is composed of two subunits i.e. the large 50S subunit (LSU) and the small 30S subunit (SSU). And the 16S rRNA gene encodes SSU while the LSU is encoded by the 23S rRNA and the 5S rRNA gene. This makes the 16S rRNA gene distinct from the 18S rRNA gene, which is similar to its eukaryotic homologue. Besides, the relatively short size (1542 nucleotides) of the 16S rRNA gene makes it easier to sequence as compared to any other gene. The gene is further subdivided into highly conserved primer binding sites and nine variable regions (V1-V9) (**Figure 1-7**). Studies on microbial ecology have long established that changes in microbial community structure affect the community function. The recent advancements in 16S rRNA based next generation sequencing have greatly helped in testing this phenomenon. Basically, the high levels of conservation in the 16S rRNA sequences imply their important role in determining cellular function and survival in microbes. This forms the basis on which 16S rRNA gene sequencing provides genus and species information for known and unknown microbial taxa. In general, the region at the 5' end of the 16S rRNA gene, the V1, V2, V3 and V4 regions comprising about 500bp sequence is widely utilized for identification of bacterial species from different samples (**red arrows, Figure 1-7**).

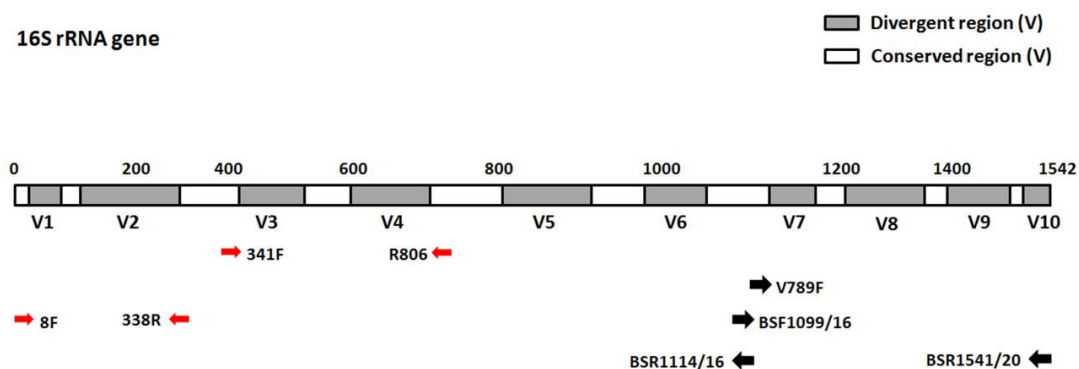


Figure 1-7 A schematic of the 16S rRNA gene showing locations of major sequencing primers
Adapted from Cai H.J Vet Diagn Invest.2003.

Integrating NGS with 16S rRNA has two major advantages namely; massive parallelization and clonal separation of templates without the need to insert gene fragment in host. Thus, this technique has become an essential prerequisite in identification of aetiological agents of infectious diseases as well as identification of previously unknown bacterial strains. For example, identification of thermotolerant *Campylobacter fetus* strains as principal cause of bacteraemia in an immunocompromised patient (Woo et al., 2002). A number of recent 16S rRNA based studies have established an altered microbial composition or dysbiosis in the gastrointestinal tract of humans with IBD (Clarridge, 2004; Janda and Abbott, 2007). Thus, the 16S rRNA sequencing not just allows robust bacterial identification but also generates reproducible and precise information as compared to traditional phenotypic testing.

1.6 Metagenomics

Beyond 16S rRNA sequencing, more precise and deep information on microbial ecologies has been recently drawn using metagenomics. The development of metagenomics has helped in defining direct genetic analysis of genomes from various environmental samples. In contrast to the unimodal phylogenetic analysis based on the diversity of single gene (say 16S rRNA gene), metagenomics catalogues multimodal genetic composition of microbial communities (Riesenfeld et al., 2004). Metagenomics helps in linking function to phylogeny besides creating evolutionary profiles of the microbial community structure. A typical sequence-based metagenome project involves sample processing, sequencing, assembly, binning, annotation, statistical analysis, and finally data storage and sharing (**Figure 1-8**).

Sample processing

The foremost step in initiating a metagenomics project is sample processing. A sufficient amount of a high quality DNA must be obtained, which represents all bacterial cell-types present in a particular sample. This is crucial for creating an unambiguous and precise genomic library. Based on the sample source, selective lysis or fractionation is used to obtain good quality DNA with minimal host impurity. This is important for samples extracted from eukaryotic sources such as humans that have large and complex genomic composition and hence can affect precise annotation of microbial communities. Another important aspect is the yield of DNA that essentially affects library production. Several new separation

techniques now exist that help increasing the yield of genomic DNA such as Multiple displacement amplification (MDA) (Lasken, 2009). MDA utilizes random hexamers and phi29 polymerase to increase DNA yield and hence has been widely utilized in both single cell genomics as well as metagenomics projects.

Sequencing

Metagenomics shotgun sequencing has gradually replaced traditional Sanger sequencing owing to its preciseness and cost-effectiveness. However, low error rate along with the ability to process long read lengths (>650 bp) and insert sizes (> 25-30 Kb for bacterial artificial chromosome or BACs) has still kept Sanger sequencing as a paragon technique (Liu et al., 2012a). Nevertheless, because of huge data sizes in genomics projects, the NGS technology is now implemented extensively. Both 454/Roche and Illumina systems are widely used for analyzing metagenomics data. Both these NGS technologies are discussed in detail in **section 1.4**.

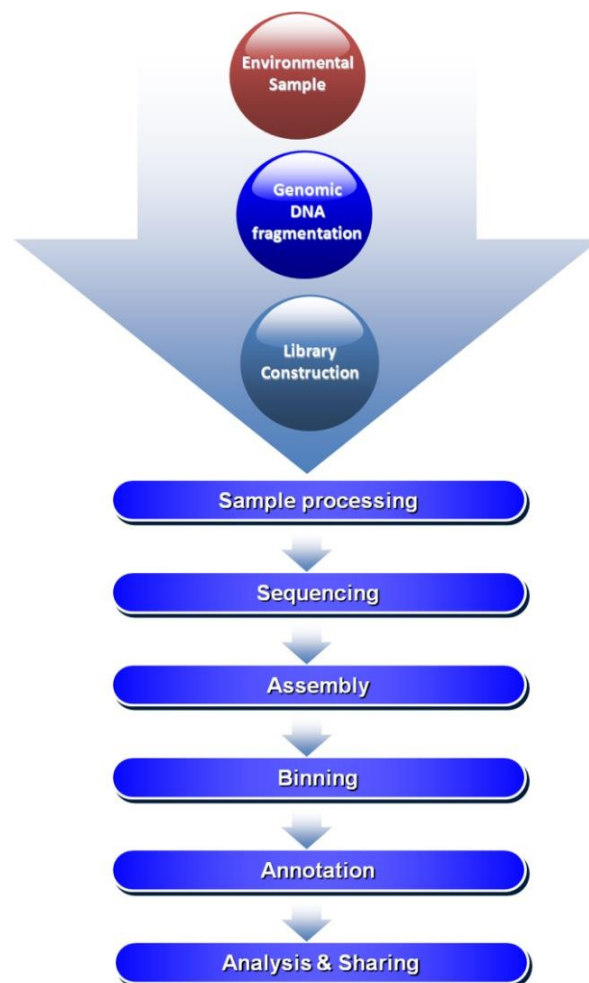


Figure 1-8 A flow diagram indicating major steps in a metagenomics study.

Assembly

The metagenomics analysis is mostly aimed at recovering the genome of microorganisms followed by further characterization of functional genes associated with them. This means a large scale assembly of short read fragments obtained from shot-gun sequencing into full length genomic contigs. Till date, several assembly programs have been reported that are mainly based on two major strategies *i.e.* reference-based assembly and the *de novo* assembly. The reference-based assembly can be implemented using a variety of available software such as MIRA, Newbler (Roche) and AMOS (<http://sourceforge.net/projects/amos/>) (Chevreux et al., 2004). These tools are considerably fast and can be efficiently utilized on a small computing platform. The primary requirement of reference-based assembly is the availability of metagenomics dataset containing sequences closely related to the test data set. Although several curated reference genomes are available, the major differences (insertions, deletions and polymorphisms) present in the sample genome could easily get skipped during assembly. Hence, the preciseness of reference-based assembly still remains low and generally produces fragmented assemblies that do not cover divergent regions. On the other hand, *de novo* assembly programs typically require huge computational resources owing to significantly large amounts of iterations. Basically, the assembly tools search an Eulerian path in a de Bruijn graph. The nodes of the graphs are called k-mers and an edge connecting any two k-mers are positioned consecutively on the same read. Few *de novo* assembly tools/packages have been developed based on de Bruijn graphs such as MetaVelvet and SOAP etc. that can compute genomic assemblies more precisely (Li et al., 2008; Miller et al., 2010; Zerbino and Birney, 2008). Several other tools and pipelines have also been developed recently that use separate algorithms for *de novo* assembly.

Binning

Another important step in metagenomics assembly is binning that means sorting DNA reads into groups that represent single genome or a group of closely related genomes. Most binning algorithms are based on two types of algorithms *i.e.* either conserved nucleotide composition (eg: GC rich regions as k-mers) or gene similarity to known genes in reference database. The composition-based binning algorithm is included in tools such as; Phylopythia, S-GSOM, PCAHIER and TACAO (Chan et al., 2008; Diaz et al., 2009; McHardy et al., 2007; Zheng and Wu, 2010). Besides, the similarity-based algorithm is incorporated in tools such

as; MG-RAST, SOrt-ITEMS, CARMA, MEGAN, IMG/M and MetaPhyler (Haque et al., 2009; Huson et al., 2007; Krause et al., 2008; Liu et al., 2011; Meyer et al., 2008). Besides, few other algorithms such as PhymmBL and MetaCluster use both composition and similarity based binning. All these binning tools group sequences as self-organizing maps (SOMs) or hierarchical clustering that may be operated either in a unsupervised or supervised (with user inputs) manner for cataloguing sequences into defined bins.

Annotation

The metagenome annotation is the most important step that produces important conclusions from a given sample set. There are two ways of metagenomics annotation i.e. either the entire assembled genome is studied as a large contig and is functionally annotated or else the annotation is performed on an entire community using the unassembled, co-clustered contigs. There are several pipelines that may be used for the first approach such as RAST and IMG (Aziz et al., 2008; Markowitz et al., 2009). However, this approach requires long contig lengths (> 30000 bp) for obtaining useful genomic annotation. Overall both these paths utilize two general steps namely; feature prediction that identifies features of interest or genes and functional annotation where putative gene functions and taxonomic neighbors are assigned. The feature prediction also requires precise demarcation of genes and genomic elements. Several tools have been designed that identify coding DNA sequences (CDS) that include FragGeneScan, MetaGeneAnnotator and Orphelia (Hoff et al., 2009; Noguchi et al., 2008; Rho et al., 2010). Besides, several non-coding gene identifier tools have been also reported that largely utilize SILVA, Greengenes and RDP databases (Cole et al., 2003; DeSantis et al., 2006; Pruesse et al., 2007). Finally, the availability of several reference databases such as KEGG, eggNOG, COG/KOG, PFAM and TIGRFAM make the interpretation and functional annotation of metagenomics sequences more precise (Finn et al., 2010; Kanehisa et al., 2004; Muller et al., 2010; Selengut et al., 2007; Tatusov et al., 2000).

1.6.1 Metagenomics and gut microbiome

Metagenomics has been lately identified as a better and comprehensive tool for deciphering the role of intestinal microbiome under normal and diseased states. The crucial information on functional genes, metabolic pathways, microbial dysbiosis and antibiotic resistance obtained through metagenomics has aided in identifying new host-microbiota interactions. The availability of Human Microbiome project along with the European MetaHIT project, has

led to an enormous addition to the reference gene catalog (Human Microbiome Project, 2012; Qin et al., 2010). Both these projects have immensely aided in metagenomics analysis in several microbiome based studies. The MetaHIT reports 3.3 million non-redundant genes in human gut microbiome which is more than 150 times larger than the total human gene complement. These genes are mostly (>99%) attributed to bacterial communities that sum up more than 1000 species (Qin et al., 2010). Moreover, metagenomic studies together with other studies have identified several factors that could influence the diversity and abundance of gut microbiome. This mainly includes geographical locations, environmental factors, diet, age and administration of drugs and antibiotics (Biagi et al., 2013; David et al., 2014; Guo et al., 2014; Perez-Cobas et al., 2013; Yatsunencko et al., 2012).

1.7 Objectives of the study

This study focused on exploring perturbing factors that affect the gut microbiota diversity and abundance. The work encompasses three major factors: (i) inflammation as pathological state, (ii) antibiotic administration under genetic susceptibility, and (iii) pregnancy, and its subsequent effects on intestinal microbiome. Through this work, the corollary between microbiome composition and physiological components (stable or disturbed) of the host (human/mice) has been explored. The basic setup utilizes NGS (454 Roche/Illumina) which is followed by detailed bioinformatics and statistical workflow. Based on the study design and analysis in each case, the following questions have been addressed:

1. How does the pathological state acute appendicitis with its severe inflammatory response affect microbial diversity? Is the intestinal microbiome of healthy individuals different from appendicitis patients? **(Chapter 3)**
2. Does *Atg16l1* play a role during antibiotics-based perturbation of microbial homeostasis in the intestine? Is there any defect in microbial recovery after antibiotic treatment in the absence of *Atg16l1*? **(Chapter 4)**
3. Are there any compositional variations in the intestinal microbiome during various phases of pregnancy and post pregnancy? Are these variations correlated with inflammatory responses or *Atg16l1* functionality during different pregnancy phases? **(Chapter 5)**

Chapter 2 summarizes the methods used in the thesis. For increased readability I have separated the individual studies in the following chapters. The chapters 3, 4 and 5 correspond to the individual questions 1-3 and contain background, study design, results and individual discussion for each part.

2 Material and Methods

2.1 Study groups and sample collection

2.1.1 Collection of appendix samples

Tissue samples were obtained by surgical removal of appendix from human appendicitis patients and healthy individuals. The extracted samples were immediately washed in ice-cold NaCl and snap frozen in liquid nitrogen before storing at -80°C . A total of twenty-nine appendicitis patients with a mean age of 38 years, having an equal male: female ratio, were analyzed. Age and sex matched control group consisting of 18 healthy individuals were also analyzed; of which 8 samples were used for 16S rRNA analysis explained in **(Table 2)**. Histological analysis was done to confirm phlegmonous appendicitis (inflamed) or healthy appendices (non-inflamed). A written consent prior to a clinical investigation was taken and the ethics committees of the University of Kiel, Germany, and Kuopio University Hospital, Finland, approved the study (AZ158/01).

2.1.2 Collection of mice samples

The genetics of mice studied for each objective is explained in **(Annexure-I)**. Mice were housed in individually ventilated cages (IVC) under specific-pathogen free conditions (SPF) and were supplied with autoclaved food, bedding and water. All animals were re-transferred to autoclaved cages (type 2 long, polycarbonate Makrolon) every week to rule out any possibility of fecal exchange among cages. A routine testing of animals for possible pathogen infection was also carried out according to protocols described by the Federation of European Laboratory Animal Science Association (FELASA). The numbers of mutant and control mice used in each study are given separately in **Table 2**, and the study design and objectives are explained in detail in each chapter separately. During the period of experimentation, mice were monitored for clinical parameters on a daily and weekly basis. The clinical parameters consisted of body weight, stool consistency and rectal bleeding evaluated by hämoccult testing (Beckman Coulter). The clinical parameters were represented as a combined score in the form of Disease Activity Index (DAI) and plotted as $\text{mean} \pm \text{SEM}$. In all cases, the fecal samples were collected at reported time points and were stored at -80°C until DNA extraction. For extracting tissue specific RNA, intestinal organs

were cut longitudinally after sacrificing the mice and were used for extracting RNA using described protocol.

Table 2 Overview of the studied projects.

Project	Name	Number of samples	Gender/Number of samples	Region	Platform	Location
1.	Appendicitis study	8 Samples (4 inflamed X 4 non inflamed)	n=8 4 male X 4 female	V1-V2	454	Ileum/ colon
2.	<i>Atg16l1</i> ^{ΔIEC} antibiotics study	Day 00 - 23 Samples (11 <i>Atg16l1</i> ^{ΔIEC} X 12 WT), Day21- 23 Samples (11 <i>Atg16l1</i> ^{ΔIEC} X 12 WT), Day56 - 23 Samples (11 <i>Atg16l1</i> ^{ΔIEC} X 12 WT), Control group (6 - Day 00, 6 - Day56)	<i>Atg16l1</i> ^{ΔIEC} = 5 male X 6 female WT = 6male X 6 female	V1-V2	454	Feces
3.	<i>Atg16l1</i> ^{ΔIEC} pregnancy study	a. Pregnancy period (Day00, T1, T2, T3)	All females(n=55) 28 <i>Atg16l1</i> ^{ΔIEC} X 27 WT (16SrRNA sequencing)	V3-V4	Miseq	Feces
		b. Until 3 months (Day00, T1, T2, T3, APWk1, Month1, Month2, Month3)	All females(n=48) 24 <i>Atg16l1</i> ^{ΔIEC} X 24 WT (16SrRNA sequencing)	V3-V4	Miseq	Feces
		c. Pregnancy period (Day00 and T3) Metagenomic sequencing	All females (n=28) (7-Day00WT, 7-Day00 <i>Atg16l1</i> ^{ΔIEC} 7-T3WT, 7-T3 <i>Atg16l1</i> ^{ΔIEC})	WGS	Hiseq	Feces

2.2 DNA extraction and amplification

2.2.1 DNA extraction from samples

The DNA extractions from the fecal or tissue samples were performed using the 'Fast DNA spin kit' or the MOBIO 'PowerSoil DNA Isolation kit' or 'AllPrep DNA/RNA Mini kit'. Initially, the samples were homogenized in 200 μl of lysis buffer (200 mM HEPES, pH 7.5, 1 M KCl, 100 mM MgCl₂, 1 mM EDTA, 0.2% NaN₃). Immediately after homogenization, 25 μl proteinase K was added and the samples were incubated at 55°C for the next 2 hours (Pierce, Erlangen, Germany) with additional mechanical force to maximize bacterial cell wall lysis. Finally, the resulting lysate was used to isolate DNA according to the manufacturer's protocol.

2.2.2 Amplification of DNA

After the lysis step, the bacterial 16S rRNA gene was amplified using broad-range forward of primers. Primers and barcodes used for each study are shown in details in **Annexure II**. For

preparing samples for 454 pyrosequencing, a 10-base multiplex identifier (XXXXXXXXXX) was added to the reverse primer to tag each PCR product. Similarly, samples for MiSeq platform were prepared by adding Illumina sequencing adapters and dual-index barcodes to each amplicon by running a limited cycle PCR using a 1 ng PCR product. For each sample, 100 ng DNA was amplified in a total reaction volume of 50 μ l using a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, California, USA). The amplification reaction contained 1X PCR buffer, 15 mM MgCl₂, 0.2 mM of each dNTPs, 1 U DNA Taq polymerase (all Qiagen, Germany) along with 0.2 mM of each primer (Carl Roth, Germany). Each reaction constituted following cycling conditions: 95°C for 5 minutes, 30 cycles of 95°C for 30 seconds, 57°C for 30 seconds, and 72°C for 1 minute, following a final extension of 10 minutes at 72°C. All PCR products were analyzed for correct sizes by running on 1.5% agarose gels. The specific bands of the amplified PCR products were purified using a QIA quick gel extraction Kit (Qiagen, Germany). The concentrations of these purified DNA fractions were measured using Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies GmbH, Germany). The resulting PCR products were mixed in single tubes in equal amounts and subsequently sequenced using the 454 pyrosequencing or Illumina MiSeq platforms.

2.3 RNA extraction and cDNA synthesis

2.3.1 RNA extraction

RNA isolation was carried out using the RNeasy Mini kit (Qiagen, Germany). Briefly, cell pellets or tissue samples were mixed with 350-600 μ l RLT (lysis) buffer and were disrupted by rapid agitation in a TissueLyser system. Following this, the lysate was centrifuged at 13000 rpm for 2 min and the resulting flow through was mixed with equal volumes of 70% ethanol. After adding RNeasy mix to the resulting solution, it was further centrifuged for 30 seconds at 10000 rpm. The flow through was then discarded and 350 μ l RW1 buffer was added to the column followed by centrifugation for 30 seconds at 10000 rpm. After discarding the flow through, the column was dried by centrifugation. Next, a master mix containing 80 μ l DNase I and 70 μ l RDD buffer was added onto the column and left to incubate for 30 minutes. Following incubation, the flow/solution was discarded and column was washed twice using 500 μ l RPE buffer (30 sec at 10000 rpm). After this, 30-50 μ l Rnase free water was added

onto to the column and kept for 5 minutes in incubation. Finally, the purified RNA was eluted by centrifugation for 1 minute at 12000 rpm and was stored at -80°C.

2.3.2 cDNA synthesis

The purified RNA samples were reverse-transcribed into single-stranded cDNA using the QuantiTect Reverse Transcription Kit (Qiagen, Germany). The eluted RNA fractions were initially incubated in gDNA Wipeout Buffer at 42°C for 2 minutes for removing any contaminated genomic DNA. Following this, the purified RNA fractions were added to a reaction mixture containing Quantiscript Reverse Transcriptase, Quantiscript RT Buffer and RT Primer Mix, and mixed. This reaction mixture was incubated for 15-20 minutes at 42°C. An additional incubation at 95°C for 3 minutes was done to inactivate Quantiscript Reverse Transcriptase. Finally, the quality of RNA and cDNA was checked by analysis of expression of housekeeping genes (GAPDH).

2.4 Sequence pre-processing

2.4.1 Quality control

After the initial pre-processing, the sequencing reads generated from MiSeq or 454 platforms were checked for quality control using the 'fastqc' file of raw data. Further, the raw sequences were trimmed based on quality score analysis using Trimmomatic (Illumina specific) tool.

2.4.1.1 Trimmomatic

The Illumina MiSeq raw sequences were trimmed using the command line of Trimmomatic platform. Briefly, the low quality bases were removed using the sliding window trimming procedure that cuts every time the quality within the window falls below the threshold. In the trimming procedure, a sufficiently accurate match in each read sequence is identified and subsequently clipped. Finally, the quality scores are converted to Phred-33 or Phred-64 and analyzed (Bolger et al., 2014).

2.4.2 Specific pre-processing steps for 454 pyrosequencing

- Firstly, after converting the binary .sff file into fasta, qual, and flow files, each flowgram was separated according to the barcode and primer combination. At this stage, length based screening of raw sequences was done and the number of flowgrams were capped according to specified length. For example: Minflow=360 and maxflow=710.
- Secondly, the shhh.flows command in the Mothur platform (implementation of the PyroNoise component of the Amplicon Noise suite of programs) was employed. The script utilizes an expectation-maximization algorithm to correct flowgrams and helps in identifying their idealized form. Each flowgram was then translated into a DNA sequence (Schloss et al., 2009).

2.4.3 De-multiplexing (454/Roche) and assembly (MiSeq)

- In this step, the multiplex identifiers (MID) and primer sequences was removed using the Meta datafile/oligos file containing primer and barcodes. The reads were further refined using a set criteria of mean quality score ≥ 25 , maximum homopolymer count of 6 and zero ambiguities using an in-house shell script based on Mothur (http://www.mothur.org/wiki/Download_mothur, version 1.32.1). Forward and reverse reads (fastq) were assembled to contigs sequences, and discarded if, were more than alignment length shown in **Table 3**.

2.4.4 Downstream pre-processing

The sequences passing the above mentioned quality control step were subjected to alignment against 16S rRNA gene using the Silva reference database. Initially, the sequences not aligning to the target region of 16S rRNA gene were discarded from analysis. Next, chimeric sequences were detected using the Chimera.Uchime algorithms and were removed from the analysis. At this stage, sub-sampling was done using the command line implementation of the Mothur platform (**Schematic 1**). In addition, sequence classification was done using the green genes taxonomic database (release August 2013, containing 202,421 bacterial and archeal sequences) or RDP taxonomic database (release May 2015 10,244 bacterial and 435 archaeal 16S rRNA gene). Classification was performed in Mothur using the k-Nearest neighbor algorithm (k-mer size=8, iteration=1000 and

cutoff=60). Reads showing matches with eukaryotic, chloroplast or mitochondrial sequences were excluded from subsequent analysis. The detailed parameters for sequence alignment and classification information for data sets studied in the thesis are described in **Table 3**.

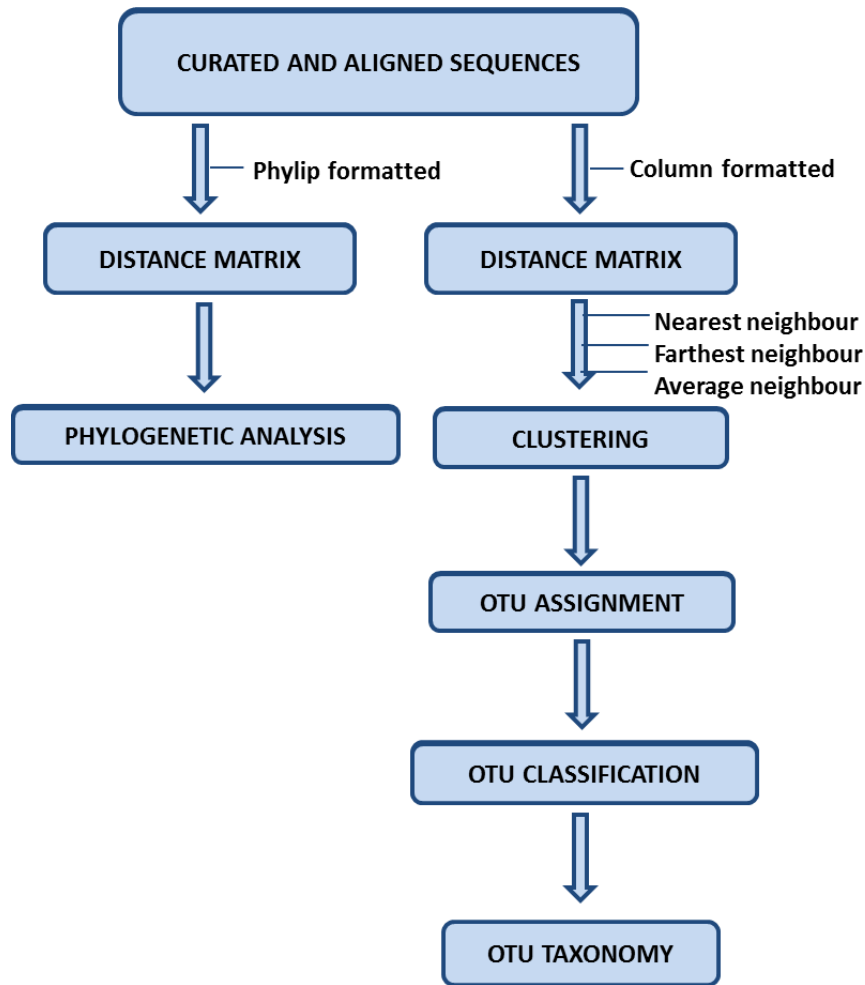
Table 3 Overview of the parameters used for sequence alignment and classification.

SN	Region	Alignment length	Alignment algorithm	Alignment Reference	Taxonomy Reference	Classification algorithm
1.	V1-V2	Start=1044, end=6333 Length=~320 bp	Search = kmer based Ksize = 8mers, Method= Needleman, Match = +1, Mismatch = -1, Gap opening = -2, Gap extending = -1	SILVA	Greengenes RDP	Method=k-nearest neighbor, K-mer size = 8, Cutoff = 60, Iteration = 1000
2.	V3-V4	Start=6428, End 23444, Length=~450 bp	Search = kmer based Ksize = 8mers, Method= Needleman, Match = +1, Mismatch = -1, Gap opening = -2, Gap extending = -1	SILVA	Greengenes RDP	Method=k-nearest neighbor, K-mer size = 8, Cutoff = 60, Iteration = 1000
3.	V4	Start=13862, End=23444, Length=~253 bp	Search = kmer based Ksize = 8mers, Method= Needleman, Match = +1, Mismatch = -1, Gap opening = -2, Gap extending = -1	SILVA	Greengenes RDP	Method=k-nearest neighbor, K-mer size = 8, Cutoff = 60, Iteration = 1000

2.5 Sequence analysis

2.5.1 OTU based analysis

A standard cut-off of minimum 1000 sequences/sample was kept to eliminate sampling bias and uneven depth of coverage. Following this, the aligned and sub-sampled dataset was used to compute a distance matrix for assigning sequences into operational taxonomic units (OTU) by a 97% similarity cut-off. An OTU represents a terminal node in phylogenetic analysis that defines a distinct 'species' in sequence-based microbiological analysis. Typically, an OTU is assigned using rDNA sequences, where a percent similarity threshold is set for classifying microbes within the same, or different, OTUs (**Schematic 1**). Following this, a sample/OTU table stating the number of sequences belonging to a certain OTU for each sample was created. This sample/OTU table was further utilized for comparative analysis of microbial diversity in the samples (**Schematic 2**).



Schematic 1 Flow diagram depicting the major analysis steps for MiSeq/454 data.

2.5.1.1 Phylotype analysis

The bacterial community composition was categorized into taxonomic groups for analyzing their distribution in samples. For brevity, mostly 'Phylum' and 'Genus' level distributions were visualized and plotted using R scripts. Based on relative abundance, the community composition was broadly divided into high abundant taxa (>1%) and low abundant taxa (<1%) respectively. Both distributions were plotted and analyzed to study sample wise variability in bacterial taxa in each case.

2.5.1.2 Alpha diversity

An estimate of diversity in taxa within the same sampling groups was done to estimate alpha diversity. Initially, Species richness (using Sobs (number of observed OTUs)) and Species evenness (using Pielou's evenness Index (J')) for analysis were utilized. Besides Chao's, Shannon's, simpson even diversity indices were also used. However, to incorporate the richness and evenness correctly, number equivalents (Legendre P, 1998) were employed as

a parameter for calculating alpha diversity. Shannon number equivalent has been used unambiguously throughout the present work. After plotting the diversity of the samples, analysis of differences between different sampling groups based on treatment (genotype/time point) was done. Paired and unpaired tests were done based on longitudinal and groups differences. In all cases, repeated-measurements of ANOVA were done to ascertain the significant differences. All analysis was performed using the Mothur platform and customized Perl and Shell scripts. Statistical analysis was done using R environment (version 3.2.1; <http://www.R-project.org>).

2.5.1.3 Indicator species analysis

Indicator species analysis was done to identify the OTUs responsible for differentiating samples/treatments. The Mothur platform was utilized to cluster the samples by average-neighbor clustering. After extracting the abundance information for all the OTUs present in the samples, indicator species representing specific sample clusters were obtained. Briefly, an indicator value (range, 0 to 1) was assigned for each OTU-sample cluster combination using both frequency of occurrence and relative abundance information. Following this, a 1,000-bootstrap test was used to generate a *P*-value for each indicator value. The resulting indicator value for an OTU is considered high when it is present in all the samples and is highly abundant for a specific sample (Dufrene M, 1997; Legendre P, 1998).

2.5.1.4 Metastats analysis

The Metastats analysis is a series of statistical methods compiled together for detecting differentially abundant features in clinical metagenomics samples (White et al., 2009). The abundance differences are tested between the groups and individual features in each datasets that distinguish between the two populations are identified. The datasets are compiled into a *Feature Abundance Matrix*, where the rows represent specific features and the columns correspond to individual metagenomics samples. The inputs for compiling the *Feature Abundance Matrix* can be either the 16S rRNA or the random shotgun data using available software packages. The 16S rRNA data processed using RDP Bayesian classifier (Wang et al., 2007) and Greengenes SimRank (DeSantis et al., 2006) or clustered using DOTUR (Schloss and Handelsman, 2005) into OTUs can be directly utilized. Also, the shotgun data classified using MEGAN (Huson et al., 2007), CARMA (Krause et al., 2008), or MG-RAST (Meyer et al., 2008) can be directly used for creating the *Feature Abundance Matrix*. To

ascertain different levels of sampling across multiple sampling sets, the raw abundance measures are converted into fractions representing the relative contribution of each feature with respect to each individual in the sampling sets. Following this, the two-sample *t* statistics is applied and features displaying *t* statistics exceeding a specific threshold are considered differentially abundant. Similarly for complex systems comprising multiple features, taxa or subsystems, the *t* statistics is not applied owing to large number of false positives. Here, *q* values are chosen assuming *P*-values of truly null tests are uniformly distributed.

2.5.1.5 Enterotype/community analysis

Presence of multiple factors such as environmental or genotypic factors may modulate the bacterial composition of gut microbiome. This could initiate the development of different bacterial community types or enterotypes indicative of different factor specific effects. To further analyze the enterotype/community type development in the sampling groups, a Dirichlet multinomial mixtures (DMM) for the probabilistic modelling of microbial metagenomics data into meta communities, or enterotypes is done (Holmes et al., 2012). This approach helps in identifying the number of communities in a particular sampling group by increasing the number of partitions until a local minimum Laplace value is reached.

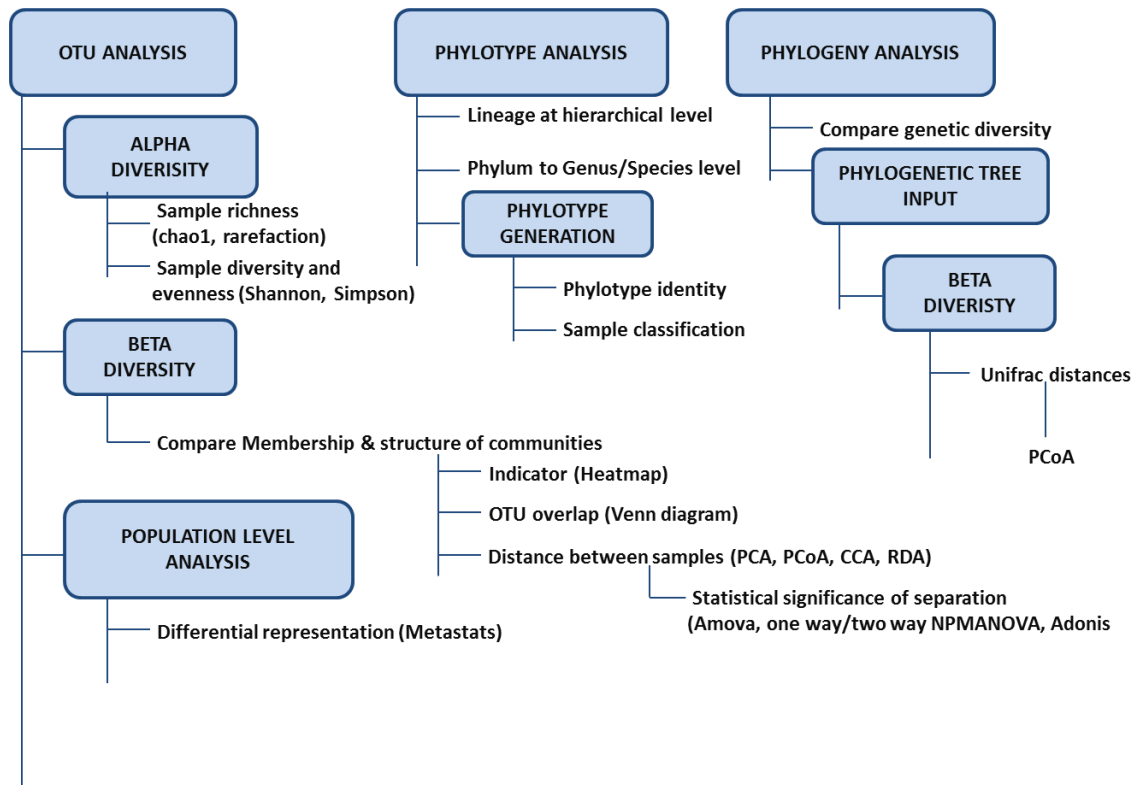
2.5.1.6 Network analysis

To deduce correlations among phylotypes, the 'sparcc' algorithm was used to generate 1000 permutations in OTUs and a spearman correlation was measured keeping minimum abundance > 10 across all samples (Friedman and Alm, 2012). Finally, OTUs with *P*-value ≤ 0.01 were chosen and were further divided into two categories of (i) OTUs with positive correlation $\leq +0.6$ and (ii) OTUs with negative correlation ≤ -0.6 . Only the OTUs showing strong correlations (± 0.6) were considered for network analysis. All individual networks were visualized and analyzed using 'DyNet' application of Cytoscape (Goenawan et al., 2016).

2.5.1.7 Beta diversity

Beta diversity analysis was done for identifying differences between sampling groups based on treatment (genotype/environment) and to check unique as well as common bacterial taxa among the groups. Initially, the distributions of observed OTU's in each sampling groups were compared using Venn diagrams. UniFrac distance matrices were calculated and were represented as Tree with branch lengths denoting similar or dissimilar communities

(Lozupone et al., 2007). While, shared branch lengths represent similarities, presence of unique branch lengths signifies differences. However, for discriminating sampling groups based on treatment, distance based constrained and non-constrained statistical methods were employed. The multivariate analyses of datasets were done using various ordination methods. This helped in identifying potential factors that influence clustering in similar groups.



Schematic 2 Flow diagram representing the steps followed in downstream analysis.

2.5.1.7.1 Principal component analysis (PCA)

PCA is an unconstrained analysis mainly applied to the raw or quantitative data and generally preserves the Euclidean distance among sampling groups. Initially, a series of data transformations were done to convert the OTU/abundance data of samples into single data points. These were used to create an association matrix containing the variances and covariances of the variables with their correlations in the form of Eigen values. Following this the original system of axes (defined by variables) was rotated to produce new axes called the principal components. Basically, the cumulative variance of sampling groups was described by each principal component and the first principal component defined the highest variance. On matching the experimental outcomes (effects) in data with each point, the variances

were used to explain a specific treatment or parameter (e.g. mutant/KO versus wildtype/WT). This was finally visualized as clusters of specific variables in the datasets. PCA was done using matrix equations using the “*Vegan*” package of R environment (R Development Core Team R; 2013).

2.5.1.7.2 Principal coordinates analysis (PCoA)

PCoA is generally employed when the similarity or distances between samples are utilized to calculate their positions in the multidimensional framework in an unconstrained analysis. Distance calculations can be done through a number of tools such as Bray-Curtis, Jaccard or weighted and unweighted unfrac metrics. It provides a Euclidean representation of a set of data points whose relationships are measured by the chosen similarity or distance measure chosen. Here, the abundance data in sampling groups were either normalized by log transformation or raw OTU counts were used. Following this, the differences between groups were tested with anosim/One-way NPMANOVA and plotted using the *Vegan* package in R and PAST v.3.0. (Øyvind Hammer, 2001).

2.5.1.7.3 Canonical correlation analysis (CCA)

A constrained ordination method is required when a relationship between response variables (e.g. OTUs) and explanatory variables (treatment e.g. mutation/time) is required. Canonical Correlation Analysis (CCA) was utilized to identify a maximum linear relationship between the sampling groups. Basically, CCA is a constrained multivariate analysis method that provides measures of the strength (canonical correlations) of multivariate association between variable sets. This helps in interpreting the role of variables in undermining the multivariate relationship between the sampling groups. CCA is based on an estimate of the covariance matrix between the two variable sets and hence can be used for precise ecological interpretation of species assemblages in the sampling groups. CCA function was applied directly to the raw, untransformed abundances using the *Vegan* package in R environment. The CCA triplots, thus created contained response variables (OTUs) represented by points to display the organization of bacterial community with respect to the environmental constraints or treatments (e.g. mutation/time).

2.6 Sequence validation

The NGS analysis in most cases is also supported by certain other experimental procedures for validating the sequencing data. This also includes *ex vivo* and *in vivo* validation of the

obtained results to correlate with the sequencing information. In the presented work, experimental sections were performed by or with assistance from few experimental biologist working at the University Hospital Schleswig-Holstein and IKMB namely Alexander Arlt and Susanne Billmann-Born.

2.6.1 TaqMan quantitative PCR

To validate the sequencing data, significantly differing bacterial taxa identified by sequencing analysis (454/MiSeq) were also confirmed using TaqMan quantitative PCR. TaqMan was also used to quantify tissue level expression of certain cytokines (pro- and anti-inflammatory). Briefly, 50 ng DNA/sample was assayed for specific bacterial taxa on the 16S rRNA region or gene specific sequences for quantifying cytokines. PCR with bacteria/gene specific TaqMan probes was performed in a final volume of 50 μ l. The reaction mixture contained 1 \times PCR Reaction Buffer (20 mM Tris-HCl (pH 8.4), 50 mM KCl); 125 μ M (each) of dATP, dCTP and dGTP; and 250 μ M dUTP; 1 μ M forward and reverse primers, 75 nM probe, 5 mM MgCl₂, and 2 U Platinum[®] Taq DNA polymerase (Life Technologies). The PCR reaction constituted initial denaturation at 95°C/10 minutes followed by 50 cycles of 95°C/15 seconds denaturation and 60°C/1 minute for annealing and extension respectively. All results were expressed as the number of 16S rRNA gene copies/ml or CTMs. Data was normalized to total bacterial load. For tissue expression analysis, all results were normalized to GAPDH. The sample wise differences were further verified statistically using the paired and unpaired tests. In Paired, t-test with Welch correction or Wilcoxon test was used and for unpaired, non-parametric Mann-Whitney rank sum test was used.

2.7 Histological analysis and scoring

Quantitative analysis of Paneth cell disorder was performed for corroborating results described in chapter 4 (Adolph et al., 2013; Bevins and Salzman, 2011; Cadwell et al., 2008; Clevers and Bevins, 2013; Salzman et al., 2010). In brief, Paneth cells in 10 spatially distributed crypts per mouse were counted and classified for lysozyme allocation patterns: normal (D0), vesicles disordered (D1), vesicle depleted (D2) and diffuse (D3). Interstitial stem cells were identified by morphology and counted in 10 spatially distributed crypts per mouse. Quantitative analysis of inflamed ileum-length was performed with the polyline tool

of the Leica LAS AF software on merged high-resolution images. In total 326 mm of mouse small intestine were counted, with a mean of 253 crypts per mouse. Criteria for classification as inflamed were: majority of Paneth cells disordered and whole crypts and also villi lysozyme positive or loss of crypt/villus architecture.

2.8 Colon mucosa transcriptome

Total RNA from Colon tissue pieces was extracted and submitted for mRNA sequencing. For all samples, mRNA focused libraries were generated using the TruSeq RNA Sample Preparation Kit v2 (Illumina) following the manufacturer's protocol. The prepared paired-end libraries (100 + 100 bp) were sequenced on the Illumina HiSeq 2000. Prior to mapping, reads that did not pass the Illumina chastity filter were excluded. Further, adapter trimming was done using cutadapt (v1.3; minimum overlap set to 5 bp) and quality trimming was performed using prinseq(v0.20.3). Subsequently, processed sequences were mapped to the murine genome (mm10, including random chromosomes) using TopHat2 (v2.0.9; Bowtie2 2.1.0). For expression estimation, count based expression values were estimated using HTSeq (v0.5.4p3).

Differential expression was calculated using the R package DESeq2. A brief description follows:

1. HTSeq read counts were imported using the 'DESeqDataSetFromHTSeqCount' function to form the count data set object.
2. Size factors were estimated using the estimateSizeFactors' function.
3. Dispersion values were estimated using the 'estimateDispersions'. The 'maxit' parameter was increased to 200.
4. Differentially expression was identified via the 'nbinomWaldTest' function. Again, maxit was increased to 200.
5. Variance stabilized and rlog transformed data was generated using the 'varianceStabilizingTransformation' and 'rlog' functions. This resulted in the first data set that did not use outlier replacement. However, this data was not used for subsequent analysis.
6. Outliers were replaced using the 'replaceOutliersWithTrimmedMean' function (minReplicates=3) and size factors, dispersion values and differentially expressed genes were

calculated as described above and transformed using variance stabilization and rlog transformation.

7. All subsequent analysis is based on the outlier-replaced data set.

The following models were investigated using DESeq:

- 1) ~cage+community type+sex+genotype
- 2) ~sex+condition, where condition is a combined term for genotype and community type.
- 3) ~sex+community type
- 4) ~sex+genotype

The Gene Ontology analysis was conducted as previously described (Tavazoie et al., 1999), while the biological processes were retrieved from the Gene Ontology Consortium (www.geneontology.org).

2.9 Enrichment analysis of differentially expressed genes

In order to interpret the biological significance of the differentially expressed genes (up- and down- regulated) for BT vs FT pair-wise comparison and genotype comparison, Gene Ontology (GO) and KEGG pathway enrichment analysis performed to investigate their functional distribution using the InnateDB database (www.innatedb.com) (Breuer et al., 2013). Transcription factor binding sites (TFBS) that were overrepresented in the promoter region of differentially expressed genes (up- and down- regulated) were identified by innateDB (Breuer et al., 2013), integrating predicted transcription factor binding site data from the CisRED database (www.cisred.org) (Robertson et al., 2006). Top ten significant GO-terms, KEGG pathways and TFBD ids obtained from up- and down- regulated DEGs (P -value < 0.1) are described in **Figure 4-15** in **Chapter 4**.

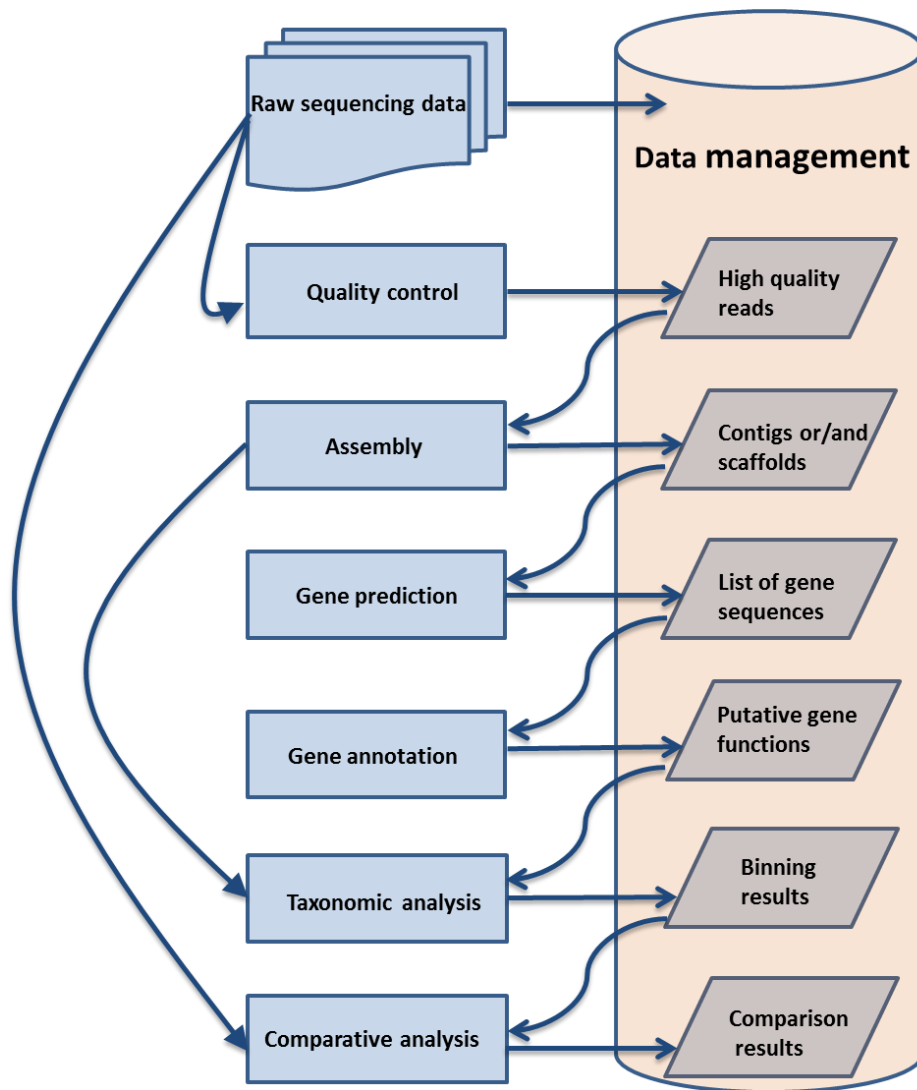
2.10 Metagenomics library preparation and Metagenomics analysis

The metagenomics DNA libraries from fecal samples were constructed using the Nextera XT kit (Illumina Inc. USA) according to the manufacture's protocol. The protocol utilizes an engineered transposome to simultaneously fragment and tag ("tagment") input DNA, adding

unique adapter sequences in the process. A limited-cycle PCR reaction uses these adapter sequences to amplify the insert DNA. Finally, the PCR reaction also adds index sequences on both ends of the DNA, thus enabling dual-indexed sequencing of pooled libraries on any Illumina Sequencing System - HiSeq 2000 platform. The paired-end reads have an average length of 150 bps per sample. Raw Fastq sequences were submitted to MG-RAST (<http://metagenomics.anl.gov>) where the initial quality controls: preprocessing, dereplication, Duplicate Read Inferred Sequencing Error Estimation and Screening were performed before the downstream analysis.

Representative hit based taxonomic profiles was performed with Min. Alignment Length Cutoff of 15, Min. % Identity Cutoff of 80% (similar to 16S rRNA profiling), Max. E-Value Cutoff of $1e-5$, Annotation source MNR5. A KEGG based subsystem functional annotation of genes predicted by FragGeneScan (trained for prokaryotes) using the BLAT program referencing the M5NR database (**Schematic 3**) (Kanehisa and Goto, 2000).

Normalization on raw functional abundances data was performed using Bioconductor package DeSeq2 implemented in R (Love et al., 2014). Functional enrichment analysis was performed using bioconductor package 'mmnet' (Cao et al., 2015). Linear discriminant analysis (LDA) was performed on functional level 4 and significant differences between groups were tested using default parameters of LDA score ≥ 2 and Kruskal–Wallis rank sum and Wilcoxon tests ($\alpha = 0.05$) as implemented in LEfSe (Segata et al., 2011). Statistical significance such as Kruskal wallis test was performed using R (version 3.2.1; <http://www.R-project.org>).



Schematic 3 Typical workflow for the analysis of metagenomics sequencing data.
Source: (Meyer et al., 2008).

3 Microbiome variation in the inflamed and non-inflamed appendix

3.1 Background

In the present chapter, perturbation of intestinal microbiota during acute appendicitis and its correlation with the inflammatory response was studied. Acute appendicitis is caused by partial or complete obstruction of the appendix. These obstructions may occur due to several factors such as worms, trauma, enlarged lymphoid follicles and tumors (Miettinen et al., 1996; Niederreiter and Kaser, 2011). The pathophysiology of appendicitis is often found associated with bacterial/enteric pathogen infection inside the organ. This results into the formation of pus and the increase in the pressure that compresses local blood vessels. At later stages, lymphatic tissue in the appendix swells and blocks the appendix incurring acute inflammation and pain (Lamps, 2004; Yamamoto et al., 2005). The common treatment strategy involves administration of anti-inflammatory drugs and/or removal of the inflamed organ in complicated cases. However, several studies have shown that at early stages of the inflammation, antibiotic treatment is equally effective (David R. Flum, 2015).

The analyses of inflamed appendices have shown the presence of mixed infiltrate of lymphocytes and eosinophil's suggestive of immune response similar to enteric pathogens. Although, higher levels of complications are reportedly associated with higher load of lymphocytic and monocytic cells, a clear time-line of this progression is not yet known. Recent reports have shown that bacterial lipopolysaccharides (LPS) may be a significant factor involved in the pathophysiology of appendicitis making it a cause-action scenario involving body's general anti-inflammatory response (Peterfi et al., 2006). The dissection of microbial composition of inflamed appendices has shown presence of number of bacterial species, such as *Clostridium sp.*, *Enterococcus sp.*, *Enterobacter cloacae*, *Lactobacillus sp.*, *Eubacterium sp.*, *Bacteroides vulgatus* and *Bacteroides distasonis*. These bacterial groups are observed to be significantly lower in abundance in inflamed appendices when compared to healthy colon (Guinane et al., 2013; Sugawara et al., 2005). A recent report also highlighted critical role of *Fusobacterium nucleatum/necrophorum* in the pathophysiology of appendicitis (Swidsinski et al., 2011). However, a detailed picture of exact bacterial composition in the inflamed organ remains elusive.

In general, the epithelial cells form gastro-intestinal barrier for preventing infection against enteric pathogens under both healthy and compromised states (Peterson and Artis, 2014). Apart from this, the epithelial lining is also responsible for initiating the innate and the acquired immune responses. Alterations in this barrier function results in the collapse of microbial growth patterns in gut and results unbalanced growth of atypical bacteria and fungi (Kelly et al., 2015; Marchiando et al., 2010). The resulting disruption of this gut microbiota induces heavy inflammatory responses and activation of pro-inflammatory cytokines as well as pattern recognition receptors (*PRR*) and antimicrobial peptides (*AMP*) (Otte et al., 2003; Rosenstiel et al., 2007). However, post-infection response against the bacterial incursion in appendicitis remains largely unknown. **Thus, this study aimed to elucidate the compositional differences in the microbial communities of inflamed and non-inflamed appendices. The main objective of this work was to find a correlation between microbial diversity or its components with the inflammatory response in acute appendicitis.** The study design and details (see **Material and Methods section 2.1.1 and Table 2**) involved human inflamed and non-inflamed appendicitis samples. Further bar-coded primers for 16S rRNA sequencing of the samples were utilized and 454 pyrosequencing was performed as described earlier (see **Annexure-II**).

3.2 Results

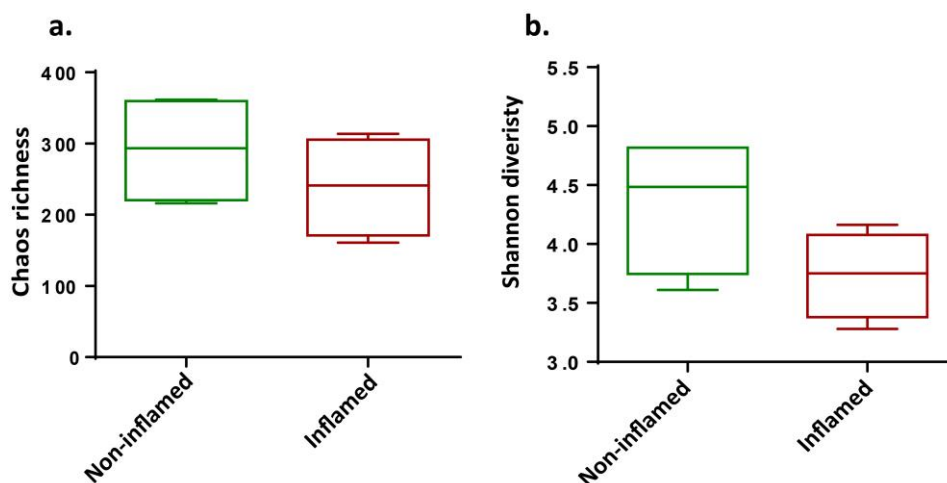
3.2.1 Bacterial taxonomy and diversity estimation

To decipher the role of bacterial communities in the pathophysiology of acute appendicitis, a comparative analysis of bacterial diversity was carried out. A 16S rDNA sequencing analysis of four inflamed and four non-inflamed appendices was performed. This resulted in 20,556 and 26,664 raw sequencing reads for inflamed and non-inflamed samples, respectively.

Table 4 Depiction of sample wise diversity estimation at 97% similarity level.

Sample	Number of OTUs	Chao I richness	Shannon diversity index	Good's estimated coverage
Inflamed_1	196	313.5	3.82	0.91
Inflamed_2	114	201	3.67	0.90
Inflamed_3	134	281.4	4.16	0.89
Inflamed_4	73	160.8	3.27	0.91
Non-inflamed_1	207	353.2	4.15	0.87
Non-inflamed_2	192	233.8	4.81	0.87
Non-inflamed_3	141	216.1	3.61	0.93
Non-inflamed_4	278	361.6	4.81	0.92

The resulting sequences were trimmed and processed to produce 6836 unique sequences, which were further clustered at 97% similarity level using Silva reference database to obtain 789 OTUs. These OTUs were used to compare the bacterial diversity between the inflamed and non-inflamed appendices. For both sampling groups, the Good's estimated coverage averaged >90% depicting satisfactory sampling (**Table 4**). Further, a comparative of Shannon index and Chao1 richness estimate showed lower bacterial diversity in inflamed appendices (**Figure 3-1**). These results indicated lowering of bacterial phylotypes under inflammatory conditions in the appendices.

**Figure 3-1** Boxplot showing alpha diversity indices.

(a) Chaos's richness and (b) Shannon diversity in non-inflamed and inflamed appendix samples. A decrease in bacterial richness and diversity in inflamed samples is clearly seen.

3.2.2 Phylotype analysis

After identifying significant lowering in abundance of bacterial phylotypes in the inflamed appendices, I next investigated the phylum and class level differences in both sampling groups. Each comparison was separated at two levels *i.e.* high abundance taxa (mean relative abundance >1%) and low abundance taxa (mean relative abundance \leq 1%). The comparative phylum level distribution of both sampling groups showed the presence of five major phyla namely; *Bacteroidetes*, *Firmicutes*, *Fusobacteria*, *Proteobacteria*, and *Tenericutes* (**Figure 3-2a**). On the other hand, there were only two major low abundance phyla namely; *Actinobacteria* and *Lentisphaerae*. In all eight samples, *Firmicutes* and *Bacteroidetes* formed the most abundant phyla with an average distribution of more than ~50% and ~34% respectively. Interestingly, among the other three major phyla, *Fusobacteria* showed higher abundance in the inflamed samples. Besides, a drastic reduction in *Actinobacteria* and total depletion of *Lentisphaerae* in the inflamed samples was also noted.

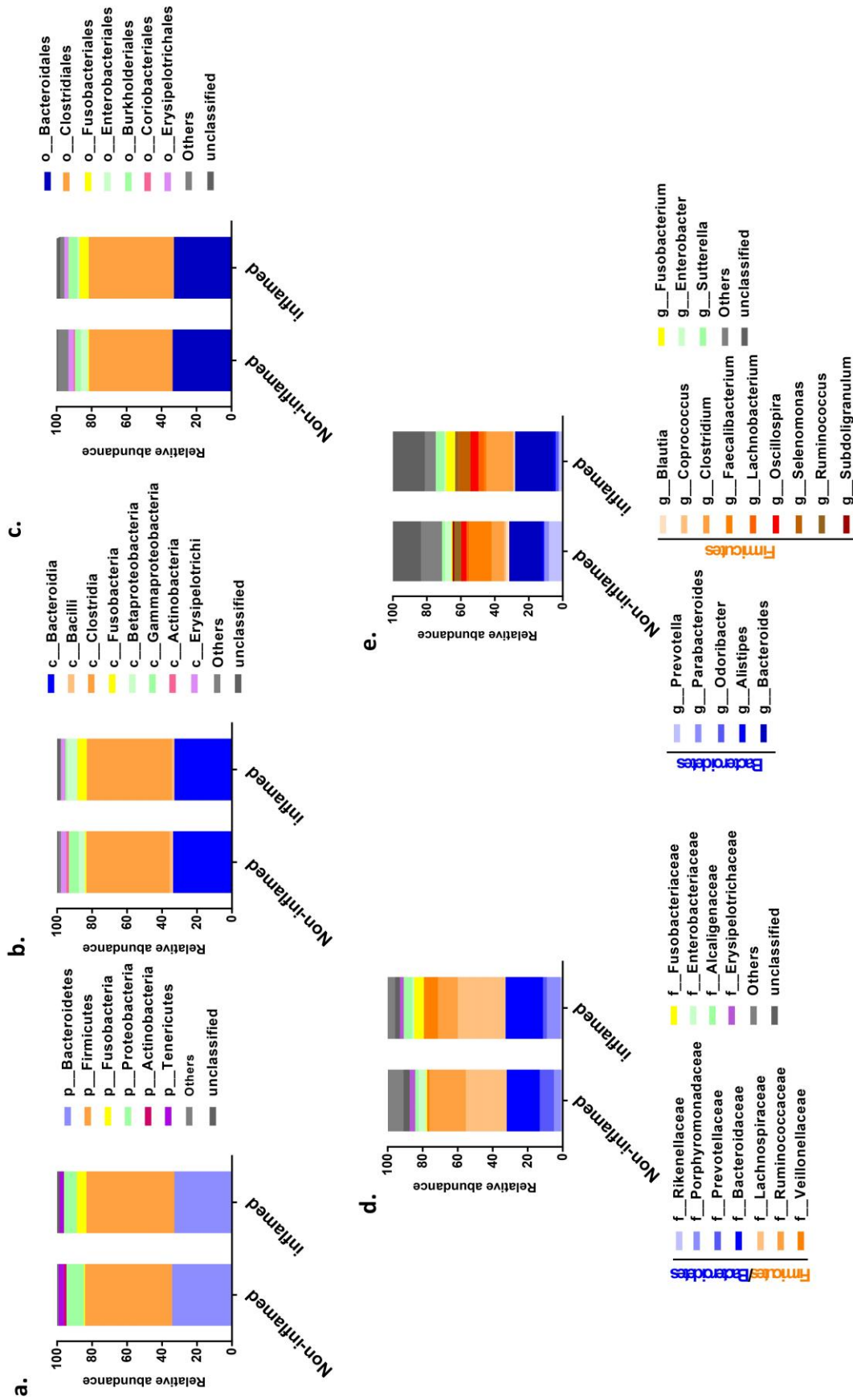


Figure 3-2 Taxonomic classification in the non-inflamed and inflamed appendix samples. (a) Phylum level, (b) Class level, (c) Family level, (d) Order level and (e) Genus level.

Further, a class level distribution analysis of all sampling groups showed the presence of seven major bacterial taxa viz. *Bacilli*, *Bacteroidia*, *Betaproteobacteria*, *Clostridia*, *Erysipelotrichi*, *Fusobacteria* and *Gammaproteobacteria* (**Figure 3-2b**). Among these, *Clostridia* and *Bacteroidia* formed the major phlotypes, with an average distribution of ~48% and ~34% respectively. The higher abundance of *Fusobacteria* in the inflamed samples was along with a significant reduction in *Gammaproteobacteria* and *Actinobacteria*. However, the major variations at class level were noted in the low abundance taxa. A sharp decline in *Alphaproteobacteria*, *Flavobacteriia*, *Sphingobacteria* and *Lentisphaerae* populations in the inflamed samples was found. Besides, a significant increase in *Deltaproteobacteria* and *Mollicutes* in the inflamed samples was also noted. This trend was also confirmed by the analysis at genus level where an increase in abundance of *Fusobacterium* was found (**Figure 3-2a-e**). Apart from this, increase in *Clostridium*, *Selenomonas*, *Sutterella* and *Lachnobacterium* were also evident. Interestingly, the non-inflamed samples were found rich in *Faecalibacterium*, *Prevotella* and *Ruminococcus* as compared to inflamed samples. A detailed phlotypes distribution at different taxa levels in both sampling groups is shown in (**Annexure-III**).

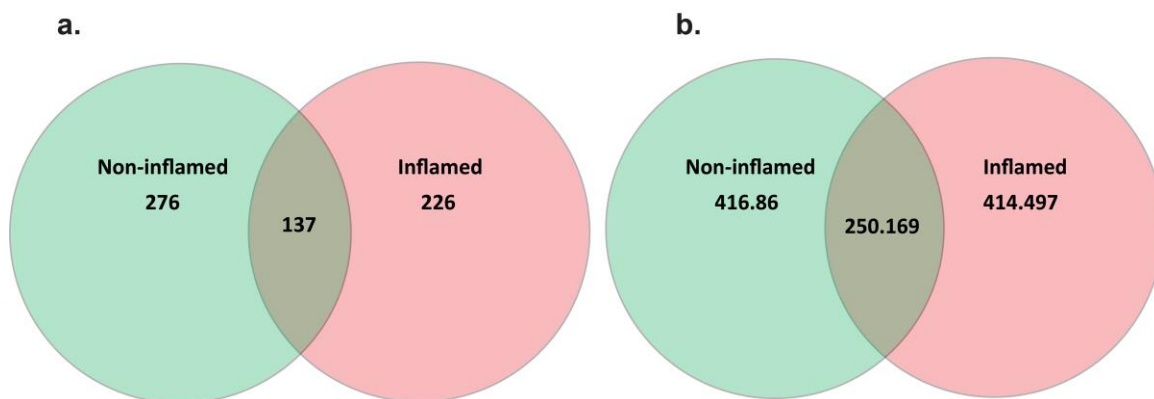


Figure 3-3 Venn diagram showing unique and overlapping OTUs in non-inflamed and inflamed samples. (a) shared OTUs and (b) Chaos richness.

3.2.3 Compositional differences in bacterial communities

After assessing the differences in phlotype abundance, I next analyzed the overlap of OTU clusters between both the sampling groups to identify any variations in the community composition. Approximately 21% OTU were shared between the inflamed and non-inflamed samples that amount to 137 OTUs (**Figure 3-3**). Besides, there were 275 and 278 unique

OTUs in the inflamed and non-inflamed samples, respectively. The inflamed samples showed higher sample variability shown forming separate clusters and with different OTU numbers. These differences were further confirmed by the analysis of sample-wise comparative Venn diagrams (**Figure 3-4A** significantly lower number of OTU sharing was observed in inflamed samples 1 and 2 (17) as well as 3 and 4 (22), indicating inter individual differences in the community composition in inflamed samples. However, non-inflamed samples showed significantly higher number of shared OTUs, for instance samples 1 and 2 (59) and samples 3 and 4 (84) depicted higher level of similarity.

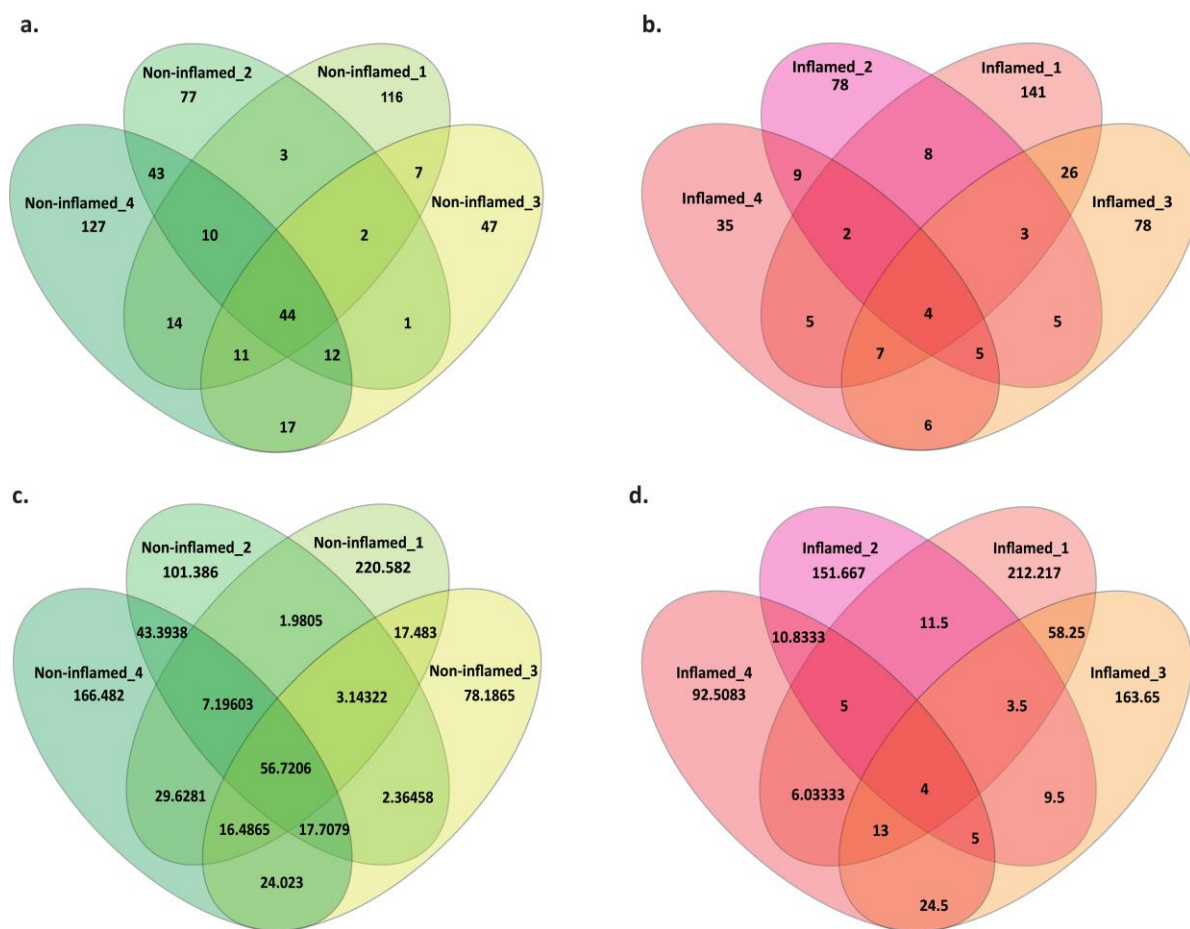


Figure 3-4 Sample wise comparative Venn diagrams depicting unique and overlapping OTUs in non-inflamed and inflamed samples. (a-b) shared OTUs, and (c-d) Chaos richness.

The community composition was further dissected using distance based principal component analysis (dbPCoA). Distance based analysis using Bray and Jaccard distances constrained with inflammation as factor was performed (**Figure 3-5**). The differences between inflamed and non-inflamed samples were tested using Adonis

(Factor=Inflammation; Bray Curtis; $R^2=0.22852$; P -value=0.03197*) (Factor=Inflammation; Jaccard; $R^2=0.19784$; P -value=0.02997*) where R^2 values were obtained using 1000 random permutations. The ordination plot showed that the OTU composition of non-inflamed samples showed less variability and resembled each other.

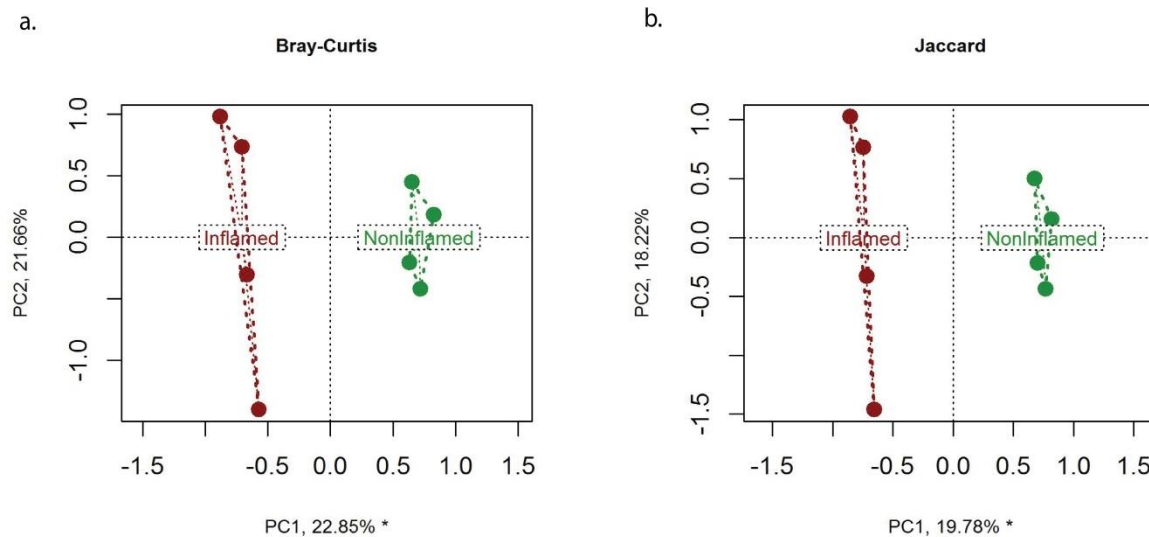


Figure 3-5 A distance based principal component analysis of non-inflamed and inflamed samples.

(a) Bray-Curtis and (b) Jaccard distances; significantly higher variation in community composition in the inflamed samples (red) is seen when compared to non-inflamed samples (green).

3.3 Discussion

Acute appendicitis is a pathological state that is characterized by abdominal pain and gastric obstruction (Smith et al., 2009). Historically, appendix is believed to act as a reservoir for microbial communities that help in recapitulating intestinal microbiome (Bollinger et al., 2007; Laurin et al., 2011). However, there is still a dearth of information on actual microbial composition of appendix. Also, how this microbial composition varies during appendicitis and its correlation with appendix associated inflammation is not yet clear. Several reports have highlighted the central role of bacterial infection in appendix inflammation (Merchant et al., 2012; Roberts, 1988). Thus, the present study was aimed at investigating the involvement of bacteria in the pathogenesis of appendicitis. The analysis of microbial communities associated with inflamed and non-inflamed appendices showed an overall decrease in bacterial abundance and diversity in the inflamed appendices. This observation suggested that decreased microbial flora in fact may contribute to initial inflammatory responses in the appendix, which, in turn, may lead to the swelling and blockade of the appendix opening. Interestingly, *Fusobacteria spp.* is increased in the inflamed appendicitis

which correlates well with previous findings (Swidsinski et al., 2011). Apart from this, abrupt decrease in *Actinobacteria* and *Lentisphaerae* in the inflamed samples indicates their role in pathophysiology. Moreover, there are compositional differences in bacterial flora of inflamed appendices as compared to non-inflamed ones. This suggests that compositional variations in bacterial community affect inflammatory responses in diseased states. These observations also strongly suggest the presence of bacteria specific responses in both normal (promoting homeostasis) and diseased (promoting inflammatory responses) states.

Another pertinent question that may arise is the actual cause of the observed compositional differences between both states. The answer lies in the previously reported observations of bacterial infection in appendicitis patients. This infection consequently leads to bacterial overgrowth due to blockage of the appendix lumen. And this blockage acts as the major pathophysiological impetus that further aids more bacterial infection that leading to severe inflammatory response in acute appendicitis (Bhangu et al., 2015). Later, the inflammatory response mediated by the infiltration of neutrophils, monocytes and lymphocytes itself may act as a principle cause, which changes the appendix-associated microflora (e.g. by ROS and/or antimicrobial peptides). Previously, it has been shown that acute appendicitis involves the expression of PRRs such as NOD2 and also inflicts high levels of AMPs (Bass et al., 2012; Shimizu et al., 2016). In the present work the α -defensins human neutrophil peptides 1–3, HD5 and HD6, as well as the two β -defensins, human β -defensins (hBD)-2 and hBD-3, were up-regulated, whereas hBD-1 was down-regulated in acute appendicitis (Arlt et al., 2015). Previously, β -defensins such as hBD-1, hBD-2 and hBD-3 are also known to affect anti-microbial/bactericidal defense mechanisms during enteric infections (Steubesand et al., 2009; Wehkamp et al., 2005). In the same line, acute appendicitis was found to down-regulate a number of β -defensins indicating a possible link with bacterial infiltrations (Rivera-Chavez et al., 2004; Rivera-Chavez et al., 2003). As a whole, the present work showcases differential regulation of AMPs in inflamed appendices and its correlation with changes in bacterial diversity during acute appendicitis.

4 Effect of *Atg16l1* on microbial resilience after antibiotic stress

4.1 Background

In this chapter, a comprehensive outlook on intestinal microbiome changes during and after antibiotic administration is presented. The study focuses on undermining comparative differences in microbial recovery (during and after antibiotic treatment) in the presence and absence of genetic susceptibility (*Atg16l1* intestinal knockout) in mice models. In general, the gut microbiome remains in a dynamic coalescence while its components constantly get remodeled under different environmental and developmental variations (Caporaso et al., 2011). The microbiome keeps accommodating new components and transforms into a consortium of complex community as the age progresses. Several extrinsic factors such as changes in dietary constituents, alcohol, drugs, antibiotics and several pathogen infestations affect this complex community structure. Nevertheless, average human intestinal microbiota remains resilient towards perturbations over lifetime and different nonspecific and specific fluctuations in its composition are indicative of acute or long term pathological states (Clemente et al., 2012; Holmes et al., 2011; Neish, 2009). Microbial resilience describes the amount of stress that a system can tolerate before its homeostatic state shifts towards a new equilibrium (Lozupone et al., 2012). Thus, resilience of the microbiota essentially defines if a particular insult will disrupt its stable state or it will regenerate back to its initial state after a short disturbance.

It has been recently shown that the gut microbiome comprises specific bacterial enterotypes that vary under stress and environmental changes. Bacterial enterotypes can be defined as different bacterial sub-populations that form the compositional basis of bacterial ecosystem of an organism gut. The functional repertoire of these gut residing bacterial sub-populations has been characterized into three robust enterotypes (Arumugam et al., 2011; Wang et al., 2014). The type 1 enterotype is characterized by high abundance of *Bacteroides*, type 2 with *Prevotella* and the type 3 enterotypes are rich in *Ruminococcus*. Interplay between these enterotypes and their combinations determine functioning of microbiota in different regulatory arrangements such as metabolite production, gut barrier function and various immune and inflammatory responses. Thus, the complex intestinal bacterial communities are not only characterized by their different regulatory roles but also by their self-regenerative capacity or resilience phenomenon. Catastrophic events such as acute

infectious diarrhea, dietary life events (e.g. phases of malnutrition) and/or antibiotic (AB) treatment may lead to a transient intestinal state, which has been termed “dysbiosis”, representing an ill-defined destruction of the sensible intestinal host-microbial balance typically related to loss of diversity. Numerous studies have linked long-term microbiome dysbiosis to human diseases, e.g. metabolic syndrome/type II diabetes, colorectal cancer or inflammatory bowel disease (Moreno-Indias et al., 2014b).

Antibiotics are generally prescribed against numerous common ailments as well as during terminal pathological states all over the world. However, the high antibiotic prescription rate doesn't account for social, economic, food, health as well as age factors (Fierro et al., 2014). In most cases, long term antibiotics usage leads to several profound short- and long-term effects leading to increased risk of allergy and asthma (Risnes et al., 2011), inflammatory bowel disease (Kronman et al., 2012; Shaw et al., 2010) or overweight, adiposity and diabetes (Azad et al., 2014; Boursi et al., 2015) later in life. These aftereffects may be attributed to the level of microbial dysbiosis depending on the age and amount of antibiotics abuse (Biedermann and Rogler, 2015). Besides, development of multiple antibiotic resistances in pathogenic strains also creates a void in treatment of several infections (Aleksun and Levy, 2007; Walsh, 2000).

It has been reported that a commonly used antibiotic amoxicillin not just reduces the level of *Lactobacillus* spp. in the small intestine but it also reduces the expression of MHC class I and II genes (Schumann et al., 2005). Similarly, clindamycin administration has been found associated to loss of *Bacteroides* population that fails to return to its original composition even after 2 years of treatment (Jernberg et al., 2007). Several other studies have shown that combination of antibiotics leads to severe alterations in the intestinal microbiota that persist long after antibiotics treatment (Buffie et al., 2012; Heinsen et al., 2015; Ubeda et al., 2010). Besides microbial dysbiosis, long term antibiotic administration has been also implicated in severe secondary effects in the host such as antibiotic associated diarrhea (AAD) involving pertinent loss of butyrate-producing bacteria (Young and Schmidt, 2004). Thus, it is important to understand the link between disease susceptibility and microbiome variations during and after antibiotics administration to unravel the factors that initiate disease like features. This could also help in identifying important parameters that define healthy or diseased microbiota of an individual.

In the present chapter, the role of functional Paneth cells that form a major source of antimicrobial peptides during the recovery after a catastrophic event and its link to *Atg16l1* gene is explored under antibiotic stress. Thus, mice harboring conditional knockout of the *Atg16l1* gene in their intestinal epithelium were studied during and after antibiotic administration. A dysbiosis of the microbiota in *Atg16l1* mutated individuals might therefore represent a possible link to understand the association of *Atg16l1* mutations to the etiology of Crohn's disease. The mechanistic basis of the *Atg16l1* dependent Paneth cell phenotype remains unsolved, but a recent study revealed the role of autophagy and particularly *Atg16l1* in resolving ER-stress, which might be crucial to cells with high ER-export and vesicle turnover (Adolph et al., 2013). Indeed, *Atg16l1^{ΔIEC}* mice develop discontinuous, fissuring ileitis with increasing age, which is dependent on hyperactivation of IRE1 α , an ER stress sensor (Tschurtschenthaler et al., 2017). Moreover, autophagy as a general cellular clearance mechanism may also assist in eliminating intracellular bacterial pathogens (xenophagy) and hence may help preventing long-term infection. In this work, I have studied whether disrupted secretion of antimicrobial peptides in mice genetically deficient in *Atg16l1* in the intestinal epithelium leads to altered microbial community structures in adulthood and/or results in impaired dynamics of diversity recovery after a course of antibiotics.

4.2 Results

4.2.1 Unaltered baseline intestinal microbiota in *Atg16l1^{ΔIEC}* mice

The intestinal microbial composition of adult *Atg16l1^{ΔIEC}* mice and the *WT* littermates (12 weeks of age) were compared before the antibiotic (AB) treatment to detect any significant alterations due to the genotype at steady state level. The bacterial population in both groups closely resembled each other as evident from similar α - and β -diversities (**Figure 4a-b**) (**Table 1**). Interestingly, *Bacteroidetes* dominated both the sampling groups at phylum and genus level (**Figure 4-1c** and **Figure 4-2**). The bar-coded primers for 16S rRNA sequencing of the samples were utilized and 454 pyrosequencing was performed as described earlier (**see Annexure-II**).

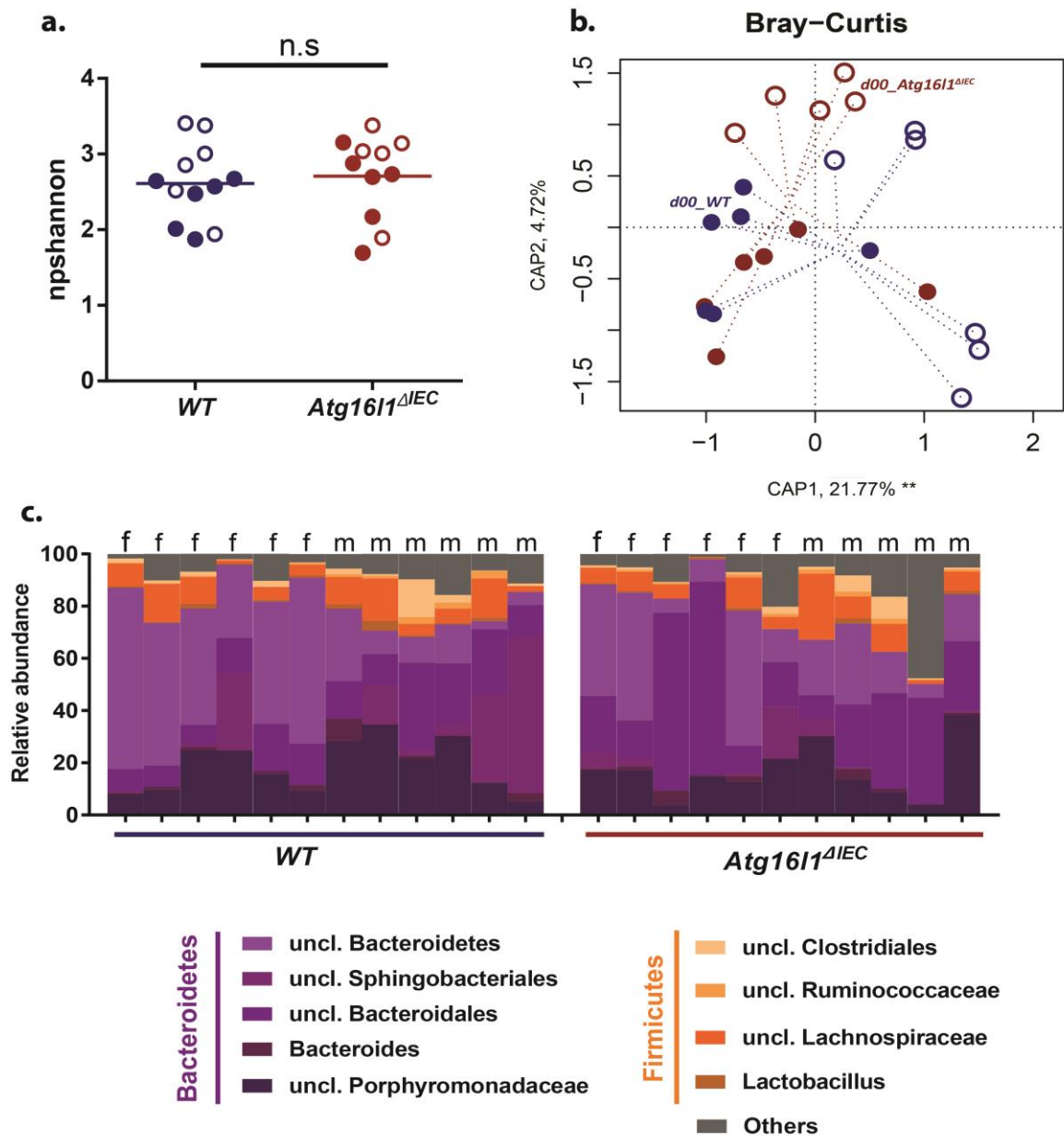


Figure 4-1 Baseline bacterial diversity in WT and *Atg16l1*^{ΔIEC} mice (n=11/12).

No significant difference at basal diversity in WT and *Atg16l1*^{ΔIEC} mice. (a) Day00 diversity before antibiotic administration is represented by non-parametric shannon (npshannon). (b) Unconstrained ordination analysis by using Bray Curtis distances shows no baseline differences between WT and *Atg16l1*^{ΔIEC}. (c) Relative abundance of microbial taxa at genus level representing fecal microbiota of WT and *Atg16l1*^{ΔIEC}. Filled and hollow circles represents female and male in each genotype respectively.

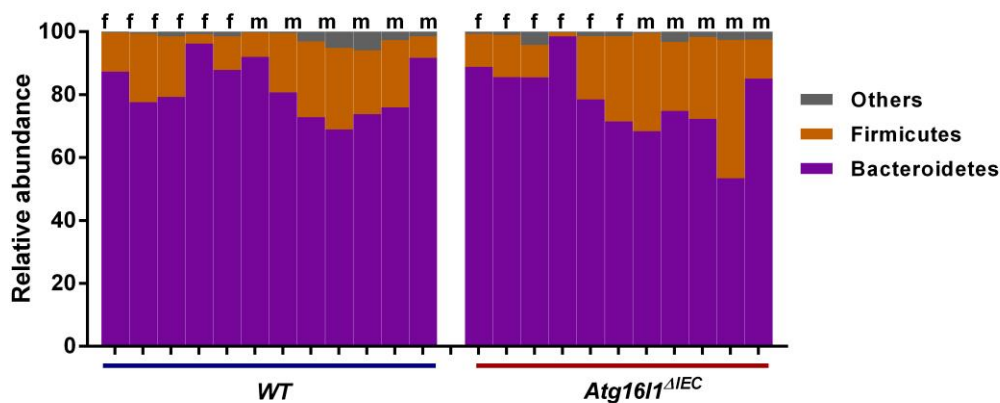


Figure 4-2 Taxonomic classification at Phylum level for *WT* and *Atg16l1*^{ΔIEC} samples at Day00.

4.2.2 Antibiotic treatment induces colitis-like symptoms independent of the genotype

To test the influence of intestinal epithelial deletion of *Atg16l1* on intestinal microbiome dynamics, a longitudinal study that involved AB administration for two weeks followed by a recovery phase of six weeks was next performed (**Figure 4-3a**). The first week (Day00-Day07) of AB treatment showed considerable weight loss with a positive disease score for both *WT* and *Atg16l1*^{ΔIEC} mice (**Figure 4-3b**). Further, in the second week (Day07-Day14), along with weight gain, both mice groups developed inflammatory-disease like symptoms with diarrhoea and rectal bleeding. The weight gain might be attributed to the development of megacaecum observed in mice sacrificed individuals at Day14. Further, ileal histologies of Day14 mice revealed moderate inflammation and a positive histological inflammation-score regardless of the genotype. Clinical symptoms persisted during the subsequent 2 weeks (i.e. post-AB treatment), and complete clinical recovery was observed only at 4 weeks post-AB administration in both groups. Further, AB treatment efficiency was also estimated by assessing fecal DNA content that showed dramatic reduction in week 1 and week 2 samples (**Figure 4c-d**).

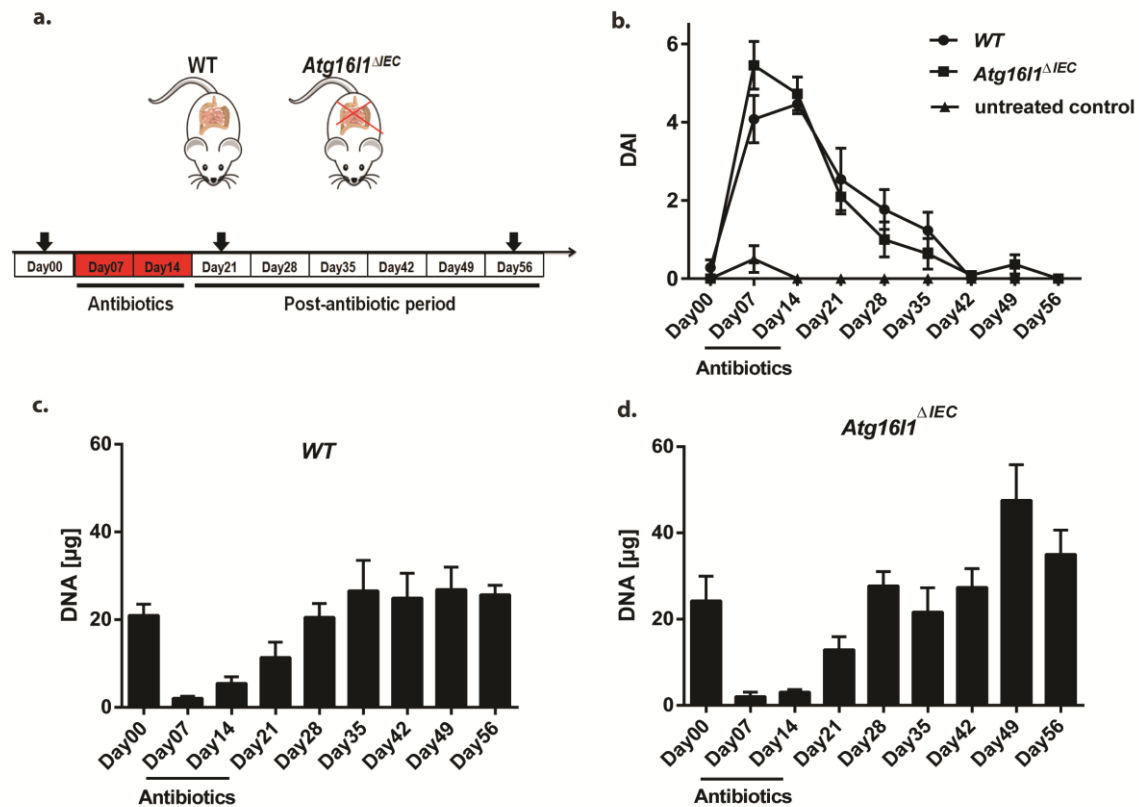


Figure 4-3 Longitudinal experimental setup, clinical symptoms and microbial diversity distribution in *WT* and *Atg16l1*^{ΔIEC} mice (n=11/12) from Day00 to Day56.

(a) A schematic of experimental design for the study. (b) High clinical disease symptoms during antibiotic administration that slowly decline to normal four weeks post-treatment. Disease Activity Index (DAI) is plotted as mean \pm SEM; *WT* (n=12); *Atg16l1*^{ΔIEC} (n=11). (c-d) Reduced fecal DNA content during antibiotic administration, DNA content of 1 pellet per sample is plotted as mean \pm SD (c) for *WT* (n=5) and (d) *Atg16l1*^{ΔIEC} (n = 5).

4.2.3 Alpha diversity variations in *WT* and *Atg16l1*^{ΔIEC} mice after antibiotic treatment

For comparing the pre-AB treatment (Day00), AB treatment (Day21) and post-AB treatment (Day56) states of microbial diversity, fecal DNA was used to quantify bacterial 16S rRNA signatures. It is important here to mention that Day07 and Day14 fecal samples could not be analyzed owing to a very low DNA content. A total of 632,328 raw sequence reads were obtained for Day00, Day21 and Day56 sampling groups.

A quantitative overview of the raw and processed 16S rRNA data of three sampling groups *i.e.* Day00, Day21 and Day56 are also represented (**Annexure-IV Table 1**). The Good's estimated coverage averaged more than 90% for all three sampling groups depicting satisfactory sampling (**Annexure-IV Table 2**). Firstly, longitudinal α -diversity measures were compared between *WT* and *Atg16l1*^{ΔIEC} mice. The Day00 and Day56 samples showed higher

diversity indices suggestive of relatively higher phylotype abundance as measured by Shannon index and Chao1 richness estimate. A comparative diversity analysis showed decline in microbial diversity at Day21 in both *WT* and *Atg16l1^{ΔIEC}* samples. The pair-wise comparisons among the *WT* sampling groups showed significant difference between the Day00 and Day21 as well as between Day21 and Day56 samples (**Figure 4-4a and 4-5a**). However, the Day56 diversity measures *Atg16l1^{ΔIEC}* failed to revert back to Day00 levels (**Figure 4-4b and 4-5b**) indicating incomplete recovery at the α -diversity level. A relatively low diversity observed in Day21 samples suggests starting phase microbial diversity after depletion by antibiotics during the AB regime.

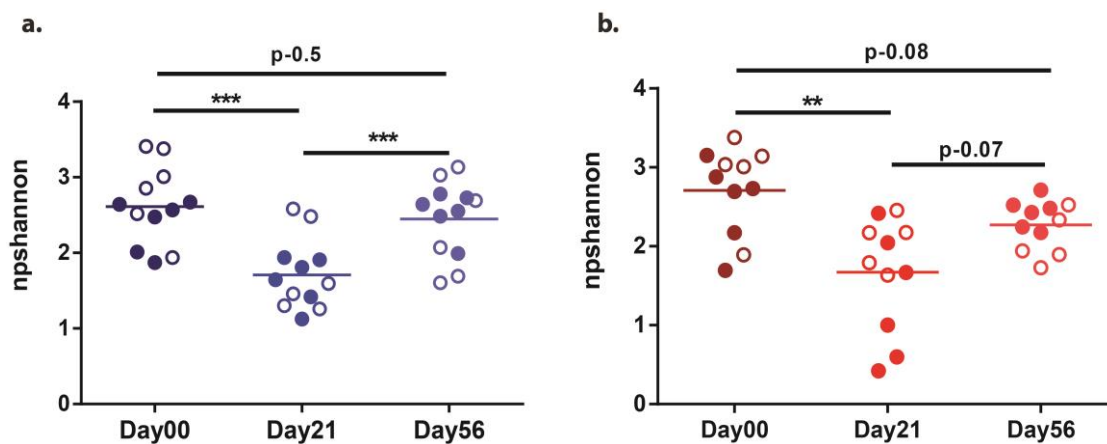


Figure 4-4 Longitudinal representation of α -diversity at three time points Day00, Day21 and Day56 for *WT* (n=12) and *Atg16l1^{ΔIEC}* (n = 11).

(a) Decline in bacterial diversity at Day21 followed by recovery at Day56 in *WT* animals. (b) Failure in recovery of diversity at Day56 in *Atg16l1^{ΔIEC}* mice following decline at Day21. Bacterial diversity is represented by Shannon numbers equivalents in (a) and (b), Pair-wise Wilcoxon rank-sum test (between same sampling groups) and Mann-Whitney test (between different sampling groups); ** $P \leq 0.001$, * $P \leq 0.05$. Filled and hollow circles represents female and male in each genotype respectively.

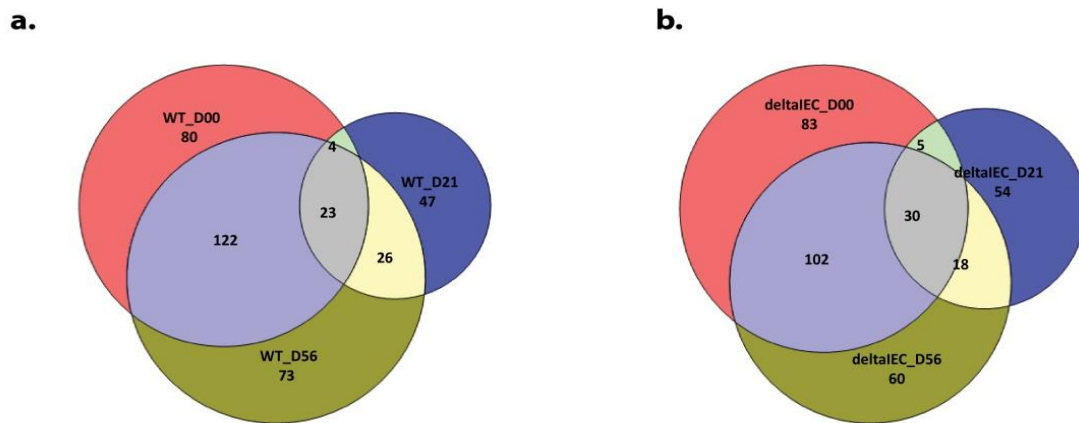


Figure 4-5 Venn diagram, showing shared OTUs at different time points - (Day00, Day21, Day56).
(a) *WT* and (b) *Atg16l1*^{ΔIEC}

4.2.4. β -diversity variations in *WT* and *Atg16l1*^{ΔIEC} mice after antibiotic treatment

β -diversity analyses for comparing *WT* and *Atg16l1*^{ΔIEC} mice groups at three time points (Day00, Day21 and Day56) showed an overall similar microbial composition (adonis; Bray-Curtis; Day, $R^2=0.22795$, P -value (as P)=0.000999***; Genotype, $R^2=0.01054$, $P=0.447552$; Gender, $R^2=0.02085$, $P=0.016983^*$; Jaccard; Day, $R^2=0.16288$, $P=0.000999$ ***; Genotype, $R^2=0.01162$, $P=0.497502$; Gender, $R^2=0.02205$, $P=0.012987^*$) (**Figures 4-6a-b**) was observed. This observation was re-affirmed by Unifrac-weighted and unweighted analysis too (Unifrac-unweighted; Day, $F=4.7033$, $P=0.001$ ***; Genotype, $F=0.9252$, $P=0.668$; Gender, $F=1.2494$, $P=0.058$; Unifrac-weighted; Day, $F=15.9093$, $P=0.001$ ***; Genotype, $F=0.9852$, $P=0.376$; Gender, $F=1.5392$, $P=0.116$) (**Figures 4-6c-d**). Although, major differences were observed due to Day effect which is the AB period, gender effect was also visible at Day56 (**Figures S1a-S1f, Annexure-IV Table 3**).

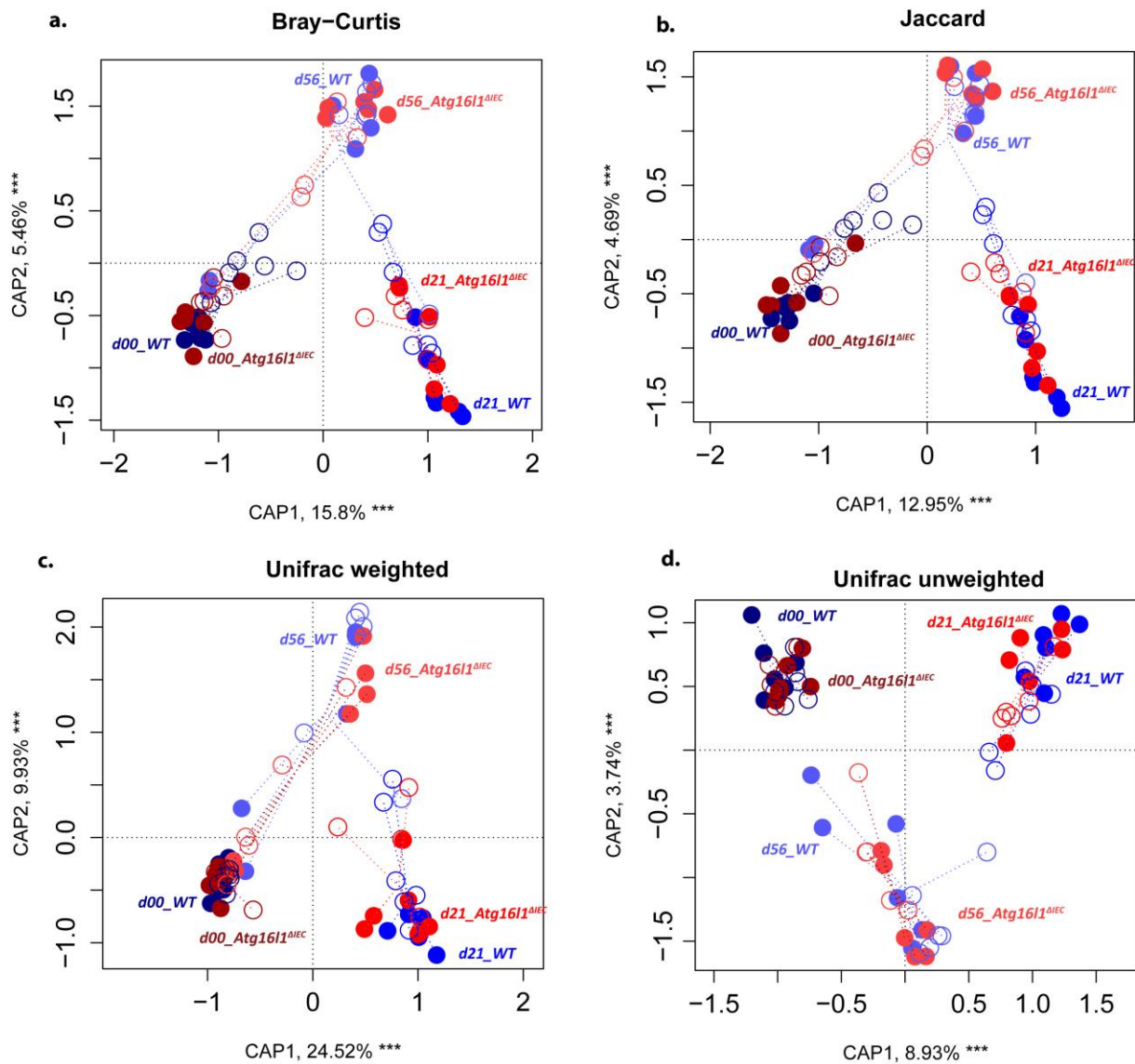


Figure 4-6 Constrained ordination analysis for WT (n = 12), *Atg16l1*^{ΔIEC} (n = 11).

(a) Jaccard index, (b) Bray-Curtis dissimilarity, (c) unweighted Unifrac distances, and (d) weighted Unifrac distances for Day/AB, genotype and gender as factors. All the distance types are showing stronger clustering based on Day/AB effect. Filled and hollow circles represents female and male in each genotype respectively.

Further, a linear mixed-effect models (LME) analysis for ascertaining interactions between Day/AB, genotype and gender effects highlighted the critical role of antibiotic treatment in determining diversity in the sampling groups (**Figures 4-7a-d**). This analysis also revealed a subtle, yet significant effect of genetic deletion and gender, with *Atg16l1*^{ΔIEC} male mice being most impaired in their recovery. To further confirm the results from the longitudinal analysis, samples from weekly time points were analyzed using TaqMan qPCR for selected bacterial taxa (*Bacteroidetes* and *Firmicutes*). The results showed high correlation to the sequencing results and were additionally consistent in their temporal development (**Figures 4-8a-b**).

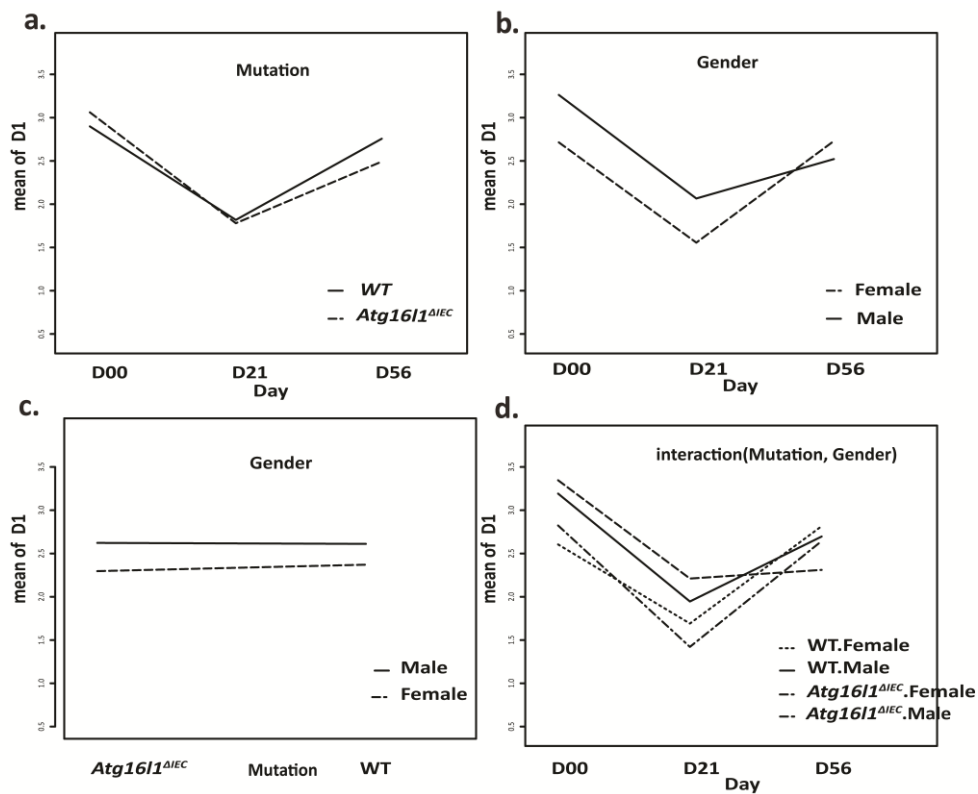


Figure 4-7 Interaction plots showing two way interactions.

(a) Day/Genotype, (b) Day/Gender, (c) Genotype/Gender and all together (d) Day, Genotype and Gender.

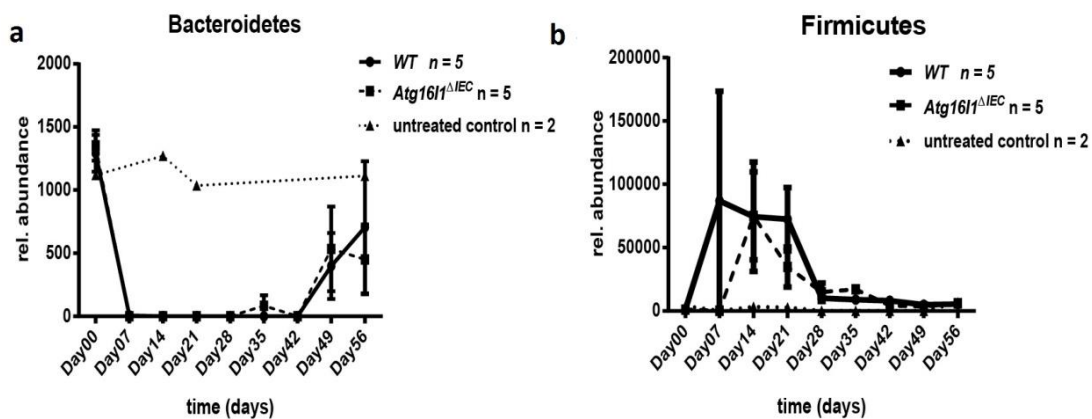


Figure 4-8 Temporal development of bacterial phyla.

(a) *Bacteroidetes* and (b) *Firmicutes* during and after antibiotic treatment. TaqMan data of bacterial taxa relative abundances are depicted as mean ± SEM.

4.2.4 Reciprocal enrichment of *Bacteroidetes* and *Firmicutes* after AB treatment

Noting the high variability of distance-based diversity measures at day 56, I next analyzed the individual data sets longitudinally (**Figures 4-9a-b**). Strikingly, a near binary split of post-antibiotic fecal bacterial communities due to mutually exclusive occurrence of OTUs from

the *Bacteroidetes* or *Firmicutes* phylum in individual animals was noted. A more detailed analysis at genus level showed the existence of two bacterial community types, later termed as *Bacteroidetes* type (BT) and the *Firmicutes* type (FT) (Figures 4-9c).

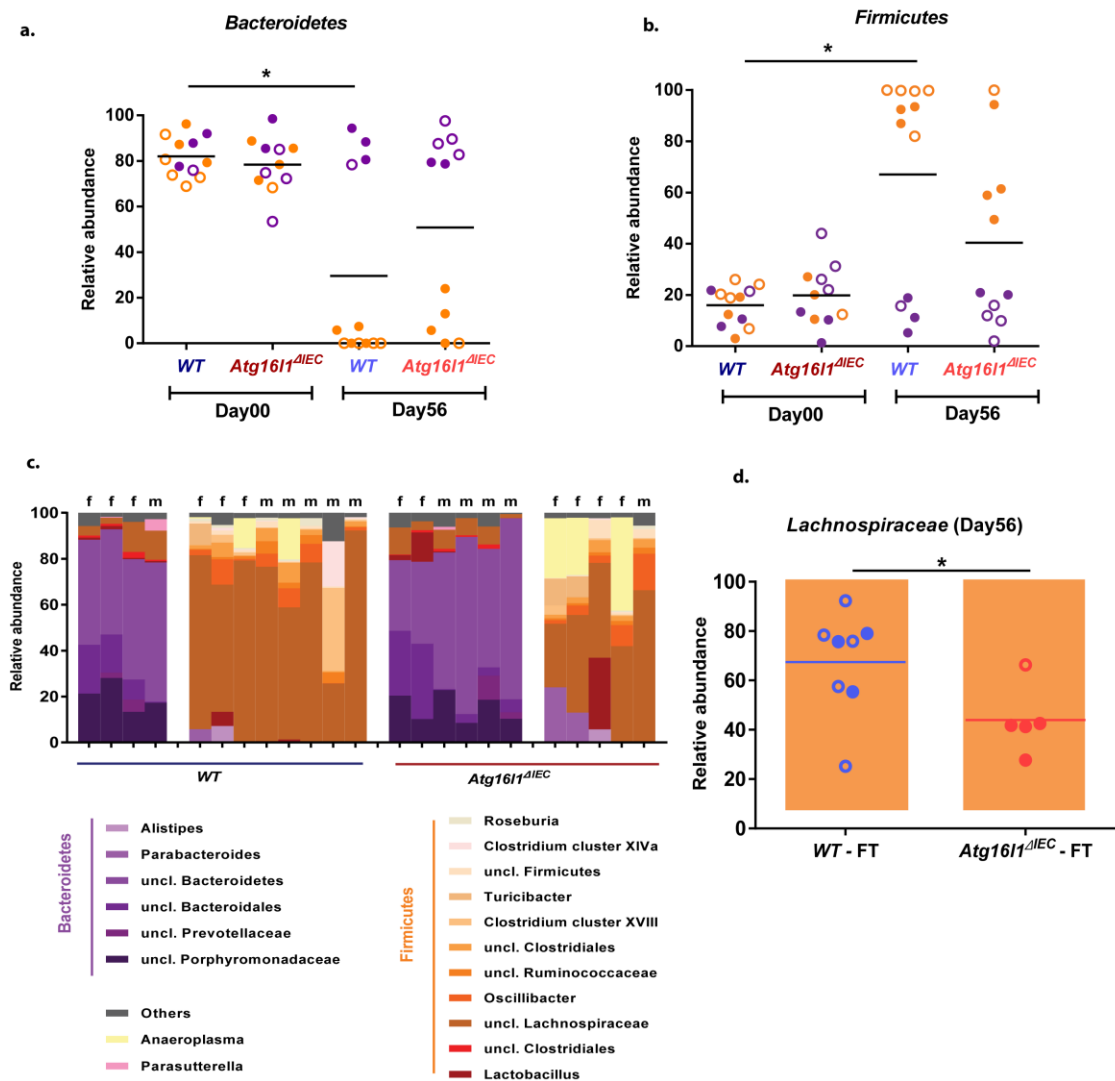


Figure 4-9 Relative bacterial distribution at Day00 and Day56.

(a-b) Distribution of *Bacteroidetes* and *Firmicutes* per sample/mice at Day00 and Day56. Reciprocal distribution of *Bacteroidetes* (blue) and *Firmicutes* (orange) in sequencing results is shown in color at Day00 and Day56. (c) Relative abundances in WT and *Atg16l1*^{ΔIEC} samples were plotted. Taxonomic assignments at genus level for samples of both genotypes (WT and *Atg16l1*^{ΔIEC}; n=12/11) showing binary splitting into two community types i.e. *Bacteroidetes* type (BT, blue) and *Firmicutes* type (FT, orange). (d) *Lachnospiraceae* of FT type showing significant differences between WT and *Atg16l1*^{ΔIEC} (n=8/5) mean ± S.E.M.; two way t-test; ** P<0.001, * P<0.05. Filled and hollow circles represents female and male in each genotype respectively.

The BT community was mainly composed of *Porphyromonadaceae*, *Bacteroidetes* and *Bacteroidales* and was reminiscent to the original flora from Day00 animals. On the other hand, the “novel” FT community was majorly dominated by *Lachnospiraceae* spp.

Interestingly, this group also showed significantly higher number of reads associated with OTUs belonging to *Lachnospiraceae* in floxed *WT* littermates compared to *Atg16l1*^{ΔIEC} mice (**Figures 4-9d, Annexure-IV Table 4 Electronic copy**).

Further, Capscale constrained analysis (Adonis) using Bray-Curtis and Jaccard distance matrices was again performed using the factors genotype, gender and community type. The analysis revealed an unusual contrast in the *Bacteroidetes* and *Firmicutes* proportions of the recovered and unrecovered samples (adonis; Bray-Curtis; community type, $R^2=0.19569$, $P=0.000999^{***}$; Genotype, $R^2=0.03154$, $P=0.582418$; Gender, $R^2=0.05244$, $P=0.145854^*$; Jaccard; community type, $R^2=0.14327$, $P=0.000999^{***}$; Genotype, $R^2=0.03627$, $P=0.539461$; Gender, $R^2=0.05513$, $P=0.070929$). Also, FT distribution in *WT* was more homogeneous compared to the *Atg16l1*^{ΔIEC} samples (**Figures 4-10a-b**). Accordingly, based on community type distribution, the analysis was further split into four sub-groups: *Bacteroidetes* -type *WT* (n = 4), *Firmicutes*-type *WT* (n = 8), *Bacteroidetes*-type *Atg16l1*^{ΔIEC} (n = 6) and *Firmicutes*-type *Atg16l1*^{ΔIEC} (n = 5).

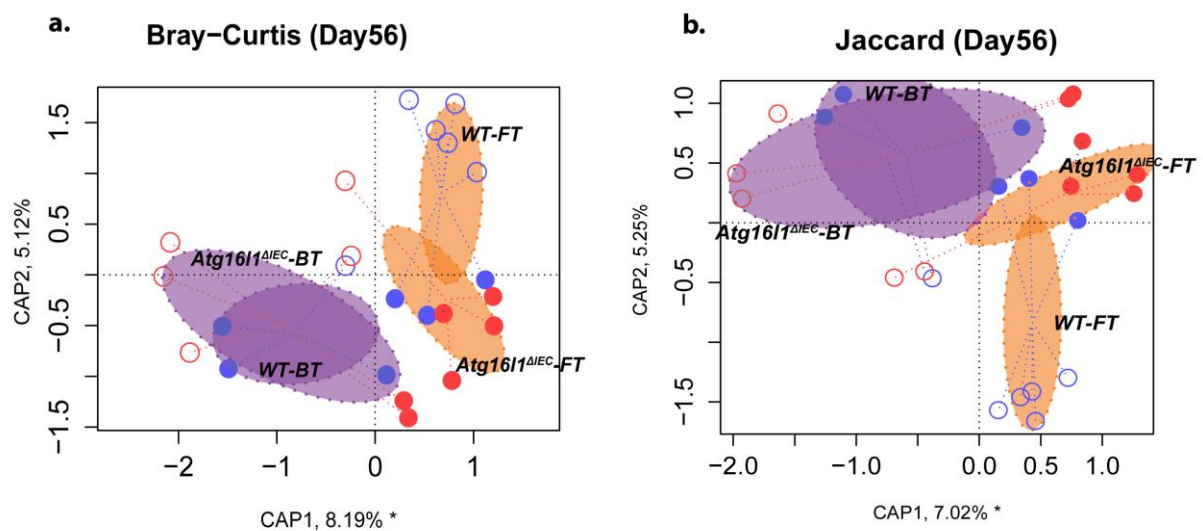


Figure 4-10 Constrained ordination analysis showing community-type based clustering at Day56.

(a) Bray-Curtis and (b) Jaccard distances. BT types were more similar and clustered together whereas FT type showed differences in *WT* and *Atg16l1*^{ΔIEC}; n=4/8/6/5. Filled and hollow circles represents female and male in each genotype respectively.

4.2.5 Assessment of pathobionts infection in animals

Although the detailed phylogenomic analysis did not reveal any overt signs of infection (e.g. expansion of *Helicobacter*, *Salmonella* or *Shigella*), further investigation was performed to

access whether the presence of particular pathobionts could explain the development of the two community types. Initially segmented filamentous bacteria (SFB) infection in the host was ascertained to rule out any pathogenic infestation that may lead to aberrant results. SFB have been shown to be strong modifiers of immune responses in the gut (Ivanov et al., 2009; Ivanov and Littman, 2010). SFB in mice are mostly found in ileum growing close to the intestinal epithelial cells. However, any signs of SFB infection were neither found in sequencing nor by specific TaqMan qPCR in any of the samples (including ileal tissue from Day56 mice). Besides, Murine Norovirus (MNV) status was also tested in all mice in two independent samples *i.e.* random hexamer transcribed cDNA of feces and of mucosa scrapings, which were subjected to Norovirus specific TaqMan qPCR. However, in the presence of appropriate positive controls, all experimental samples were negative.

4.2.6 Analysis of indicator species affecting binary split of community types

To further identify any bacterial groups responsible for the dichotomous community types; the indicator bacterial sub-populations in Day21 samples, which might predict the development of the respective community types in individual mice, were compared. Indicator species were filtered to choose significant indicator taxa in each group (**Annexure-IV Table 5 Electronic copy**). *Ruminococcaceae* (OTU 0072) belonging to the *Clostridium cluster IV*, and *Lachnospiraceae* (Otu0060), a member of *Clostridium cluster XIVa*, were found as indicator taxa for the BT group whereas *Clostridium cluster XI* (Otu0142) and more in FT group. Overall, this analysis indicated that the *Firmicutes* sub-population in both BT and FT groups at day 21 may dictate the diversion to either community type.

4.2.7 An interaction of bacterial community type and genotype and its association with the inflammatory reaction in the ileal mucosa after antibiotics

At this stage, it was important to ascertain whether *Atg16/1* genotype (*i.e.* the presence/absence of the intestinal epithelial deletion) and/or type of bacterial community may influence the histopathology of the ileal and colonic mucosa, as the clinical DAI showed inflammation in both genotypes (**Figure 4-3b**). Histological examination at day 56, confirmed the formerly described alternations of *Atg16/1*^{ΔIEC} crypts in Paneth cell vesicle and lysozyme distribution (**Figure 4-11a-c**). A significant reduction in stem cells in the crypts of *Atg16/1*^{ΔIEC} mice was also found (**Figure 4-11d**). Both these observations were not correlated with the presence of BT and FT community types. This indicated that alterations in Paneth cell and

stem cell numbers are only genotype dependent and are not influenced by microbiome composition (BT or FT community type).

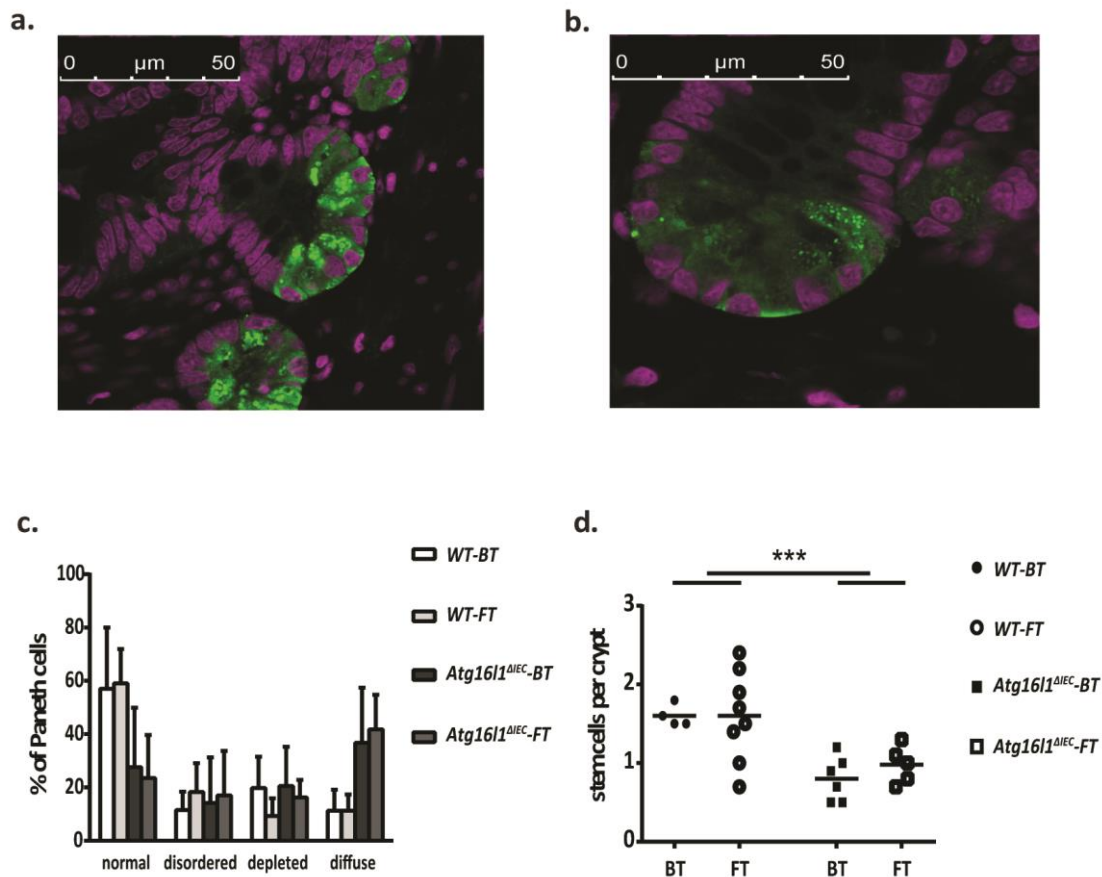


Figure 4-11 Paneth cell morphology and stem cell number in WT and *Atg16l1*^{ΔIEC}; n=12/11.

FFPE sections of day56 are immunostained for Lysozyme (green) and nuclei are counterstained with DRAQ5 (magenta). Scale 50 μ m. (a) WT and (b) *Atg16l1*^{ΔIEC} (c) Paneth cells in 10 spatially distributed crypts per mouse were counted and classified for lysozyme allocation patterns: normal, vesicles disordered, vesicle depleted and diffuse (WT-*Bacteroidetes* type n=4; WT-*Firmicutes* type n = 8; *Atg16l1*^{ΔIEC}-*Bacteroidetes* type n=6; *Atg16l1*^{ΔIEC}-*Firmicutes* type n=5; unpaired Student's t-test; mean \pm SD *** P \leq 0.001). (d) Interstitial stem cells per crypt (unpaired Student's t-test; mean \pm SD; *** P \leq 0.001).

However, *Atg16l1*^{ΔIEC} mice displayed more and longer stretches of the ileal mucosa which were characterized by an expansion of lysozyme production in epithelial villus and crypt cells other than Paneth cells (**Figure 4-12a-b**). Interestingly, the detailed scoring of inflammatory changes in the mucosa showed an interaction of genotype and community type. In general, *Atg16l1*^{ΔIEC} mice and mice with FT community type harbored a significantly higher degree of inflammation as measured by histology score and changes in the length of ileal regions (**Figure 4-12c-d**). Both factors were additive, with *Atg16l1*^{ΔIEC} mice carrying an FT community type showing the highest inflammation scores (**Figure 4-12e**). A similar trend was observed in the colonic mucosa, however the histological changes were more subtle (data not shown). These results suggest that the FT community type promotes a pro-inflammatory phenotype

after antibiotic treatment. Importantly, indicator analysis demonstrated a genotype-dependent negative correlation of *Lachnospiraceae* in the FT community type with *Lachnospiraceae* being significantly depleted in the *Atg16l1*^{ΔIEC}-FT mice (Figure 4-12f).

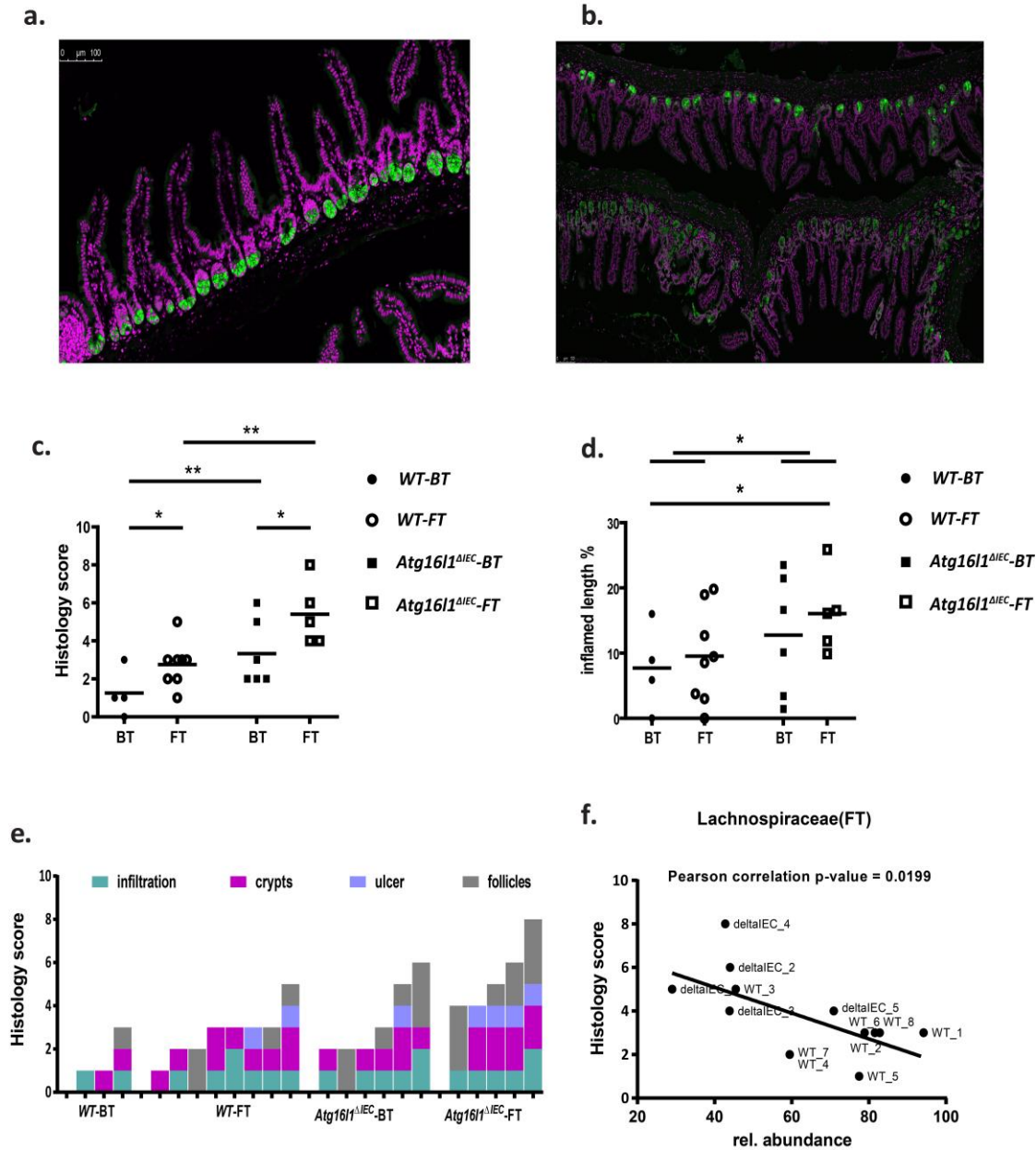


Figure 4-12 Ileal inflammation in *WT-Bacteroidetes* type n=4; *WT-Firmicutes* type n=8; *Atg16l1*^{ΔIEC}-*Bacteroidetes* type n=6; *Atg16l1*^{ΔIEC}-*Firmicutes* type n=5.

(a) Uninflamed (b) and inflamed ileal crypts. FFPE sections of Day56 are immunostained for Lysozyme (green) and nuclei are counterstained with DRAQ5 (magenta). (c) Histology score in the small intestine at Day56. HE stained and lysozyme immunostained ileal FFPE sections were scored. *Atg16l1*^{ΔIEC} mice show more signs of inflammation, which is also true for the *Firmicutes* resilience type subsets of the two genotypes. (One tailed unpaired Student's t-test; * P≤0.05, ** P≤0.01). (d) Length of ileal regions at Day56 with severe changes in lysozyme allocation patterns and crypt villus architecture as depicted in (b), (one tailed unpaired Student's t-test; * P≤0.05). (e) Detailed results of the histology score categories for all individuals. (f) Negative correlation of inflammation score with *Lachnospiraceae* abundance within the *Firmicutes* type *WT* and *Atg16l1*^{ΔIEC} mice (n=8/5), Pearson * P≤0.05).

4.2.8 Colon mucosa transcriptome analysis shows distinct subsets of genes regulated by *Atg16l1* deficiency mice and bacterial community types

RNA-Seq and subsequent Gene Ontology based gene enrichment analysis was performed to describe effects of genotype and community type in host physiology and metabolism. The analysis of genes differentially regulated in both the genotypes revealed that WT animals showed increased expression of genes mainly involved in maintaining homeostasis (DHRS4) (Oszkiel et al., 2014), immune receptors (CEACAM2) (Alshahrani et al., 2014) as well as certain house-keeping genes (RPL4 and RPL7) (Kenny et al., 2012; Loos et al., 2012). Interestingly, the *Atg16l1*^{ΔIEC} samples showed increased expression of antimicrobial peptide (AMP) genes such as CHGB which has been associated with IBD, colon and rectal cancer (Arijs et al., 2009). Besides, increased expression of certain transcripts such as RPS6KB1 encoding the p70 ribosomal S6 kinase. The kinase responds to mTOR signaling and impaired S6P signaling has been found associated with intestinal inflammation and colorectal cancer (Hashimoto et al., 2012) and TPH (polymorphism associated with IBD) suggest a complex link of the observed mucosal expression pattern of *Atg16l1*^{ΔIEC} mice to metabolic processes (Slattery et al., 2011; Xiong et al., 2015) (**Figure 4-14**).

These observations were further supported by KEGG/GO pathway analysis of WT versus *Atg16l1*^{ΔIEC} animals that showed an enrichment of upregulated transcripts in major metabolic pathways such as cellular homeostasis, glucuronate metabolism, and ATP catabolism in the wildtype situation. Interestingly, as the top-ranked category we found an increased expression of transcripts related to ER stress pathways, although it is important to note that the transcripts (e.g. ERN2, CREB3L2, FAM129a, PDIA5) do not reflect a canonical pattern associated with an active unfolded protein response in epithelial cells (Novacek et al., 2010). Conversely, in the *Atg16l1*^{ΔIEC} mice an upregulation of lipid catabolic/peroxisome-related transcripts, as well as a signature related to caspase activation was found (**Figure 4-15**).

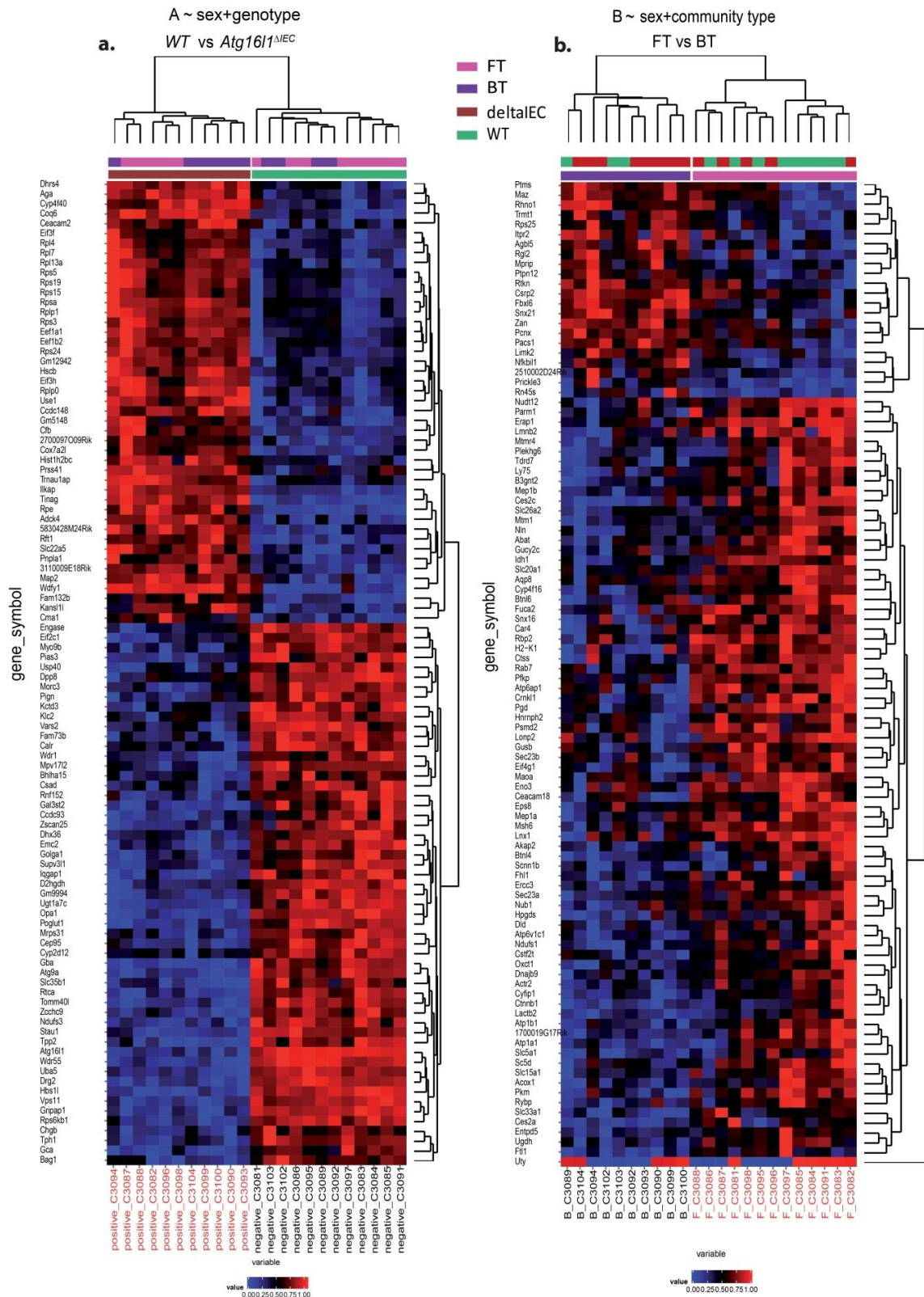


Figure 4-13 Colon mucosa transcriptome.

(a) Heatmap of significant differentially regulated genes between *WT* and *Atg16l1*^{ΔIEC} samples at Day56. (b) Heatmap of significant differentially regulated genes between *Bacteroidetes*-type and *Firmicutes*-type samples at Day56. (Also available as Electronic Copy - Figure 4-13)

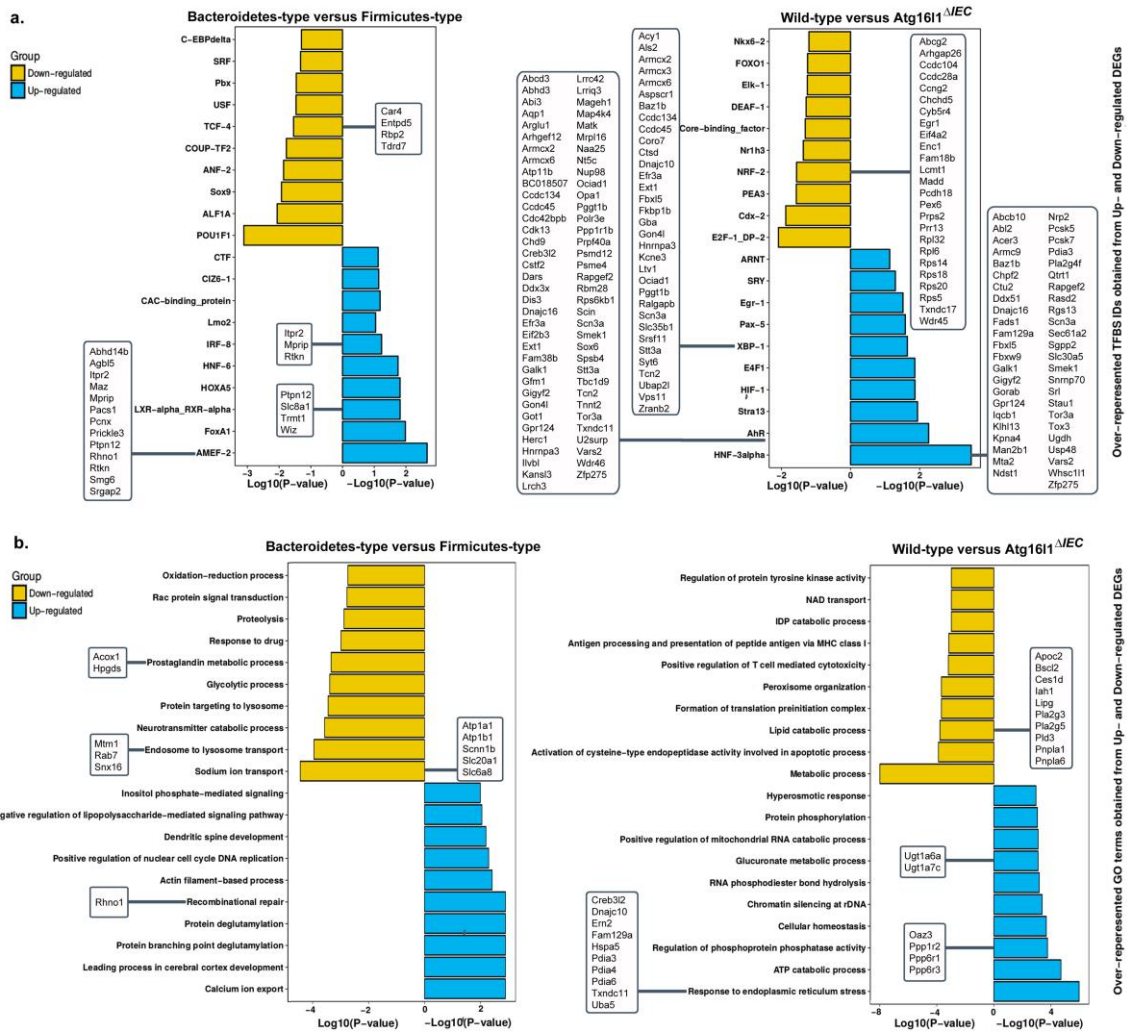


Figure 4-14 Enrichment analysis of differentially expressed genes.

Over-represented up- and down-regulated a) transcription factor binding sites and b) GO term categories defined in differentially expressed genes (P -value<0.05) between sampling groups. A comparison between *WT* versus *Atg1611*^{ΔIEC} and *Bacteroidetes* type versus *Firmicutes* type communities is shown in each case. Each box shows the list of differentially expressed genes involve in the respective categories. The x-axis indicates their associated enrichment P -value for each category. (Also available as Electronic Copy - Figure 4-14)

Further, comparing the two community types indicated that FT mice have pronounced increase in ion channel modulators (*Atp1a1*) and intestinal transporters associated with intake of bacterial products (*SLC15A1*). Interestingly, the intestinal transporter protein PePT1 encoded by *SLC15A1* is a risk gene associated with IBD (Zucchelli et al., 2009) and modulates NOD-like receptor signaling (Dalmasso et al., 2011; Vavricka et al., 2004). Pathway and transcription factor binding analyses revealed an up-regulation of proteolytic and glycolytic processes in FT mice, whereas transcripts with predicted LXR/RXR binding sites were higher in mice with a bacterial community dominated by *Bacteroidetes* (Figure 4-15).

4.3 Discussion

Antibiotic administration is a common event and an average individual until adulthood undergoes several courses of antibiotics, which each time at least transiently affects the normal intestinal microbiota. Several epidemiological and mechanistic studies suggest a direct link between early life exposure to antibiotics and an increased risk of inflammatory and/or metabolic diseases, including obesity, diabetes, allergic disorders and Crohn's disease. However, little is known about host genetic factors determining resilience of intestinal microbial communities to different environmental stressors. The major aim of present work was to investigate the effects of a deficiency of *ATG16L1* in the intestinal epithelium on the dynamics of intestinal microbial communities during and after a single course of antibiotics (AB). For this, a longitudinal study comprising antibiotic treatment for two weeks followed by monitoring microbial recolonization for 6 weeks was carried out.

Surprisingly, there were no significant differences in baseline microbial diversity between the *WT* and *Atg16l1* ^{Δ IEC} mice, despite of reported physiological alterations of Paneth cells and decreased defensin levels (Adolph et al., 2013). Nevertheless, during AB treatment, colitis-like symptoms in both mice groups were evident from observed clinical symptoms (inflammation, rectal bleeding and diarrhea. Histological analysis showed increased distorted crypts, ulcers and lymphoid follicles in the mucosa of *Atg16l1* ^{Δ IEC} mice still present at Day56. This signified that the mutation impedes the recovery phase suggestive of a higher inflammation state in the mutant mice. Interestingly, microbiome alterations in both mice groups were almost similar after 56 days. On dissecting the data at sample level, I found that few samples showed microbial diversity closer to native microbiota (Day00). This observation altogether meant that antibiotic stress led disturbances in microbial recovery actually affected *WT* and *Atg16l1* ^{Δ IEC} mice in a similar fashion. This also meant that an absence of *Atg16l1* doesn't invoke any different response during and after antibiotic administration. Interestingly, on analyzing the major microbial phyla, an anti-correlation in the distribution of *Bacteroidetes* and *Firmicutes* in samples was observed. This led us to categorize the samples into *Bacteroidetes* rich (BT) or *Firmicutes* rich (FT) based on relative abundance. The microbial composition of BT group resembled the native microbial composition (Day00), while the FT group showed similarity with Day21 composition (**Figure 4-10**).

Previous reports have shown that the *Firmicutes/Bacteroidetes* ratio gets influenced by age, obesity, race and other dietary and lifestyle factors (Bervoets et al., 2013; De Filippo et al., 2010; Mariat et al., 2009; Wolf et al., 2014). It was hypothesized that decision for conversion to either community types may be brought about by specific group(s) of bacteria that affect initial re-colonization. These bacterial groups might be present either from the starting point (Day00) or may arise during the initial recovery phases (Day21) helping the community to acquire a positive or negative recovery path. It was found that *Ruminococcaceae* and *Lachnospiraceae* identified as Day21 indicator species for BT and FT type at day 56, respectively. Both these bacterial groups have been previously shown to be actively involved in carbohydrate metabolism and sugar transport mechanisms along with other metabolic pathways (Biddle et al., 2013). Thus, owing to these crucial metabolic functions they plausibly may play a role for “correct” microbial recovery. Besides, *Lachnospiraceae* was also found negatively correlated with inflammation in both *WT* and *Atg16l1^{ΔIEC}* samples. However, on investigating its initial and final distribution, a significantly higher abundance was found in the *WT* samples. Our results may point to a potentially important association, as certain bacteria belonging to *Lachnospiraceae*, e.g. *Roseburia hominis* have been postulated to play a protective role in IBD and human colorectal cancer development (Frank et al., 2007; Machiels et al., 2014; Meehan and Beiko, 2014). Moreover, in human IBD patients the defective *ATG16L1* allele has been associated with decreased numbers of *Bacteroidaceae* and increased *Lachnospiraceae* (Sadaghian Sadabad et al., 2015).

It was further shown that impaired Paneth cell morphology and low epithelial stem cell number were not correlated with any of the two community types, but were only dependent on the mouse genotype. Besides, the absence of any extraneous infection (SFB, Norovirus or *Helicobacter*) further strengthened the opinion that the morphological changes are only genotype driven and do not require any additional trigger.

Prolonged inflammation was found to be associated with both genotype as well as bacterial community type in an additive interaction. Evidently, although the FT community type was linked to higher levels of histological inflammation overall, the *WT*-FT animals showed lower levels of inflammation compared to *Atg16l1^{ΔIEC}* mice. The observations support that functional *Atg16l1* is required for regulating the immune homeostasis in the gut and that dysfunction of the protein interacts with microbial community structure under

stress conditions. The findings corroborate previous reports that Paneth cell distortion and higher inflammatory state have to interact with other factors like ER stress to cause overt intestinal inflammation (Adolph et al., 2013).

The presented data fuel the hypothesis that microbial challenges act together with genetic variants in *ATG16L1* act on the resilience potential of the intestinal ecosystem, thereby influencing the immune balance, and likely, also the manifestation of Crohn's disease. Our study is to the best of our knowledge the first one which links a simple genetic defect to the fate of recovery of host-microbial homeostasis during and after AB treatment. The findings connote a potential therapeutic and/or preventive exploitation in IBD and beyond. It can be postulated that future personalized medical care must involve an integrated diagnostic understanding and therapeutic targeting of longitudinal stability and/or resilience of the intestinal metaorganism.

5 Role of *Atg16/1* in microbial variations during pregnancy and postpartum recovery

5.1 Background

A study on dynamics of intestinal microbiota and its variation during pregnancy is described in this chapter. The study outlines microbial compositional differences during different stages of pregnancy in the presence and absence of genetic susceptibility (*Atg16/1* intestinal knockout) in mice models. Pregnancy represents an altered state of body functions where the metabolism and related immune system of the body is affected. It represents a unique immune condition that is created temporarily (gestation period) to prevent any damage to the fetus and is thus very important for the conservation of species. However, a significant number of pregnancies result in preterm births which eventually lead to death of the newborns. In most of these cases, host microbiota invasion into the amniotic cavity and inflammation have been commonly observed (Blencowe et al., 2013; Gagliardi et al., 2013). Accumulating evidences have shown that maternal gut microbiota is able to translocate to the placenta via the blood stream (Jimenez et al., 2008; Matamoros et al., 2013; Prince et al., 2014; Rautava et al., 2012). During pregnancy, the microbial communities associated with numerous metabolic as well as immune pathways directly impact the fetus. This notion is supported by several previous findings and in particular a recent report where neurodevelopmental disorders associated with fetus are implicated to alterations in microbial ecology (Jasarevic et al., 2015). Thus, the microbiota not just impacts host metabolism but may also influence the healthy status of the fetus (Prince et al., 2014).

In another study, it was shown that vaginal microbial profiles in early pregnancy were significantly associated with an elevated rate of subsequent preterm births (DiGiulio et al., 2015). Thus, it seems that different stages of pregnancy impart different effects on the microbiota composition which in turn impacts the health status of the fetus through altered metabolic and immune responses. This notion is corroborated by a recent report that showed differences in bacterial communities between the first and third trimesters and also that fecal microbiota from third trimester exhibit inflammation and energy loss. Similarly, higher adiposity and insulin insensitivity was seen in animals colonized by 3rd trimester microbiota as compared to first trimester microbiota colonized mice (Collado et al., 2008; Koren et al., 2012). Besides, the host-microbiota cross talk also gets heavily affected due to

several other factors including diet, the immune system and host genetics (Spor et al., 2011). In addition, microbiota is known to play key role in the development of immune response, nutrient assimilation, protection against pathogen invasion and several other physiological and homeostatic functions (Walter and Ley, 2011). The composition of this diverse microbiota can be affected by host genetics and environmental exposures (Spor et al., 2011). Altogether, these reports substantiate critical host-microbiota cross talk that varies during different phases of pregnancy. Nevertheless, a good number of studies have associated host-specific factors with alterations in steady-state levels of the microbiome (association of genotypes or KO models). For example, a genetically deficient mouse (say *Nod2* gene) may lead to an altered acquisition of microbial communities over time early in life. This in turn suggests that recognition of the microbiome is an important part of the host-microbial dynamics (Rehman et al., 2011; Rosenstiel et al., 2007). Accordingly, mutational defects in several genes such as *FUT2*, *NLRPs*, *TLR*, *ATG16L1* are reportedly associated with altered gut microbiota composition (Elinav et al., 2011; Frank et al., 2011; Heno-Mejia et al., 2012; Rausch et al., 2011; Sadaghian Sadabad et al., 2015; Ubeda et al., 2012).

The T300A polymorphism in the coding region of *Atg16/1* gene (autophagy related 16-like 1) has been found associated with increased risk of CD (Hampe et al., 2007; Rioux et al., 2007) in Caucasian populations. This genetic defect in *ATG16L1* gene produces impaired autophagy and Paneth cell defect which in turn is a major source of antimicrobial peptides required for the intestinal homeostasis (Adolph et al., 2013; Cadwell et al., 2008; Clevers and Bevins, 2013). Additionally, *Atg16/1* has been also found associated with cellular clearance mechanisms and assists in eliminating intracellular bacterial pathogens (xenophagy) thus preventing long-term infections (Levine and Deretic, 2007; Schmid and Munz, 2007). However, at this stage, very little is known about dynamics of the microbiome during pregnancy and its variations in genotype dependent manner. Thus, dysbiosis of the microbiota in *Atg16/1* mutated individuals represents a possible link to understand the functional association of *Atg16/1* to pregnancy and associated community variations. Previous few studies on microbiome changes during pregnancy were limited by low sampling points and thus a longitudinal investigation encompassing larger number of time points including pre-pregnancy, pregnancy and post pregnancy is required (Collado et al., 2008; Koren et al., 2012). In the present study described here, a longitudinal approach has been used to investigate microbial and physiological response of pregnancy in *Atg16/1* deficient

(*Atg16l1* ^{Δ IEC}/*KO*) and corresponding wild types controls (*Atg16l1*^{*fl/fl*}/*WT*). The temporal dynamics of microbiota was studied using 16S rRNA genes profiling and shotgun metagenomics to correlate with observed host responses. The data indicates overall stable microbial diversity and richness (alpha-diversity) during and after pregnancy. However, the study also indicates shift in the microbial composition and profile along the pregnancy stages between *Atg16l1*^{*fl/fl*} and *Atg16l1* ^{Δ IEC} animals.

5.2 Results

5.2.1 Study design

In order to understand the effect of gestation on gut microbial communities in genotype dependent manner the fecal samples were collected at weekly time intervals until 3 months from each sampling group (n=7 per genotype). To study the host response, one set (n=5/4, *Atg16l1*^{*fl/fl*}/*Atg16l1* ^{Δ IEC}) of mice from Pregnant, and equal number of mice from nulliparous non-pregnant from day 0 to trimester 3) were sacrificed. Intestinal organs were cut longitudinally and fixed with PFA for analyzing other host/histological parameters from individual mice. The fecal pellets sampled for reported time points were stored at -20 °C until DNA extraction (**Described in methods section 2.2**). Similarly, tissues were stored at -80 °C for RNA isolation (**Described in methods section 2.3**). The primers for 16S rRNA sequencing of the samples are described (**see Annexure-II**) and metagenomics (WGS) DNA isolation in **methods section**. Details of samples collection time points are depicted in (**Figure 5-1a**).

5.2.2 *Atg16l1* deficiency produces marked microbial changes in pregnancy

In order to investigate the temporal changes in microbial diversity during pregnancy, alpha diversity estimations were performed using Chaos richness and Shannon index (**Figure 5-1b-c**). Both diversity indices remained unaltered in nulliparous and pregnant control animals. However, species richness and diversity was slightly reduced in *Atg16l1* ^{Δ IEC} pregnant mice compared to the nulliparous control groups ($P=0.05$) (**Figure 5-1b**). Phylum level bacterial abundance showed the presence of four major phyla at Baseline: *Bacteroidetes* (51.4%), *Firmicutes* (45.5%), *Proteobacteria* (1.7%), *Actinobacteria* (0.9%) for *Atg16l1*^{*fl/fl*} at baseline and *Bacteroidetes* (50.7%), *Firmicutes* (45.2%), *Proteobacteria* (2.7%), *Actinobacteria* (0.5%) for *Atg16l1* ^{Δ IEC} (**Figure 5-8a, S2a**).

Interestingly, a trend of reduced relative abundance of *Bacteroidetes* in *Atg16l1*^{ΔIEC} mice at T3 was observed (**Figure 5-1d**). It was found that the most dominant genus *Bacteroides* in this phylum showed significant reduction at T3 ($P=0.03$) (**Figure 5-1f**). However, such differences were not observed in *Atg16l1*^{fl/fl} animals across all stages of pregnancy. Besides relative abundance of *Proteobacteria* in the *Atg16l1*^{ΔIEC} mice, initially increased slightly and then decreased. But these differences in relative abundances were significant ($P=0.01$) only from T1 to T3 (**Figure 5-1e**). Similarly, *Lactobacillus* showed increase in abundance from day 0 to T3 in both the genotypes but the increase was higher in *Atg16l1*^{fl/fl} as compared to *Atg16l1*^{ΔIEC} mice (**Figure 5-1g, S2b**).

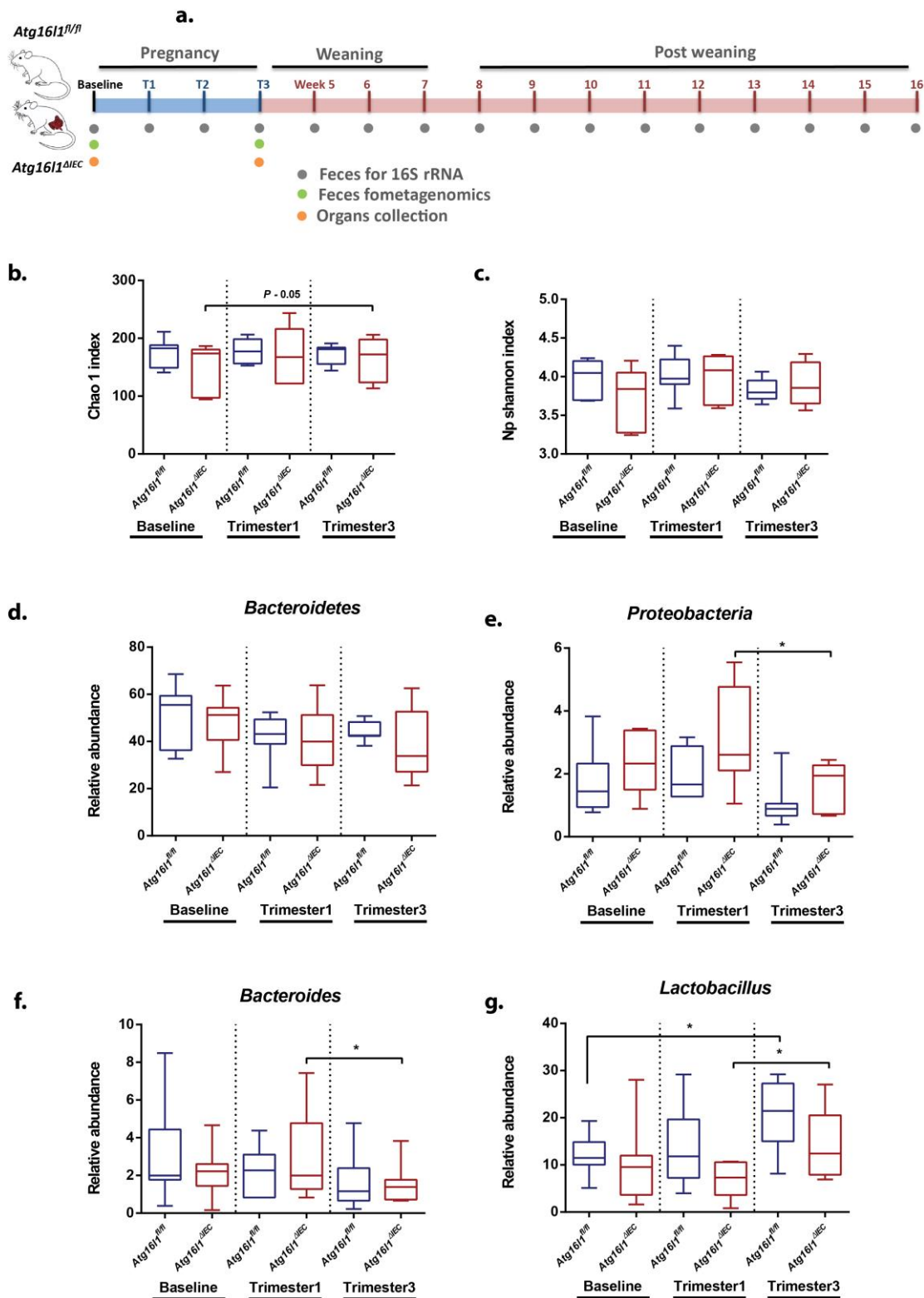


Figure 5-1 Study design, bacterial diversity and abundance estimation.

(a) Study design depicting pregnancy and post-pregnancy timeline and longitudinal fecal sampling points. Fecal samples were collected at weekly time interval to study bacterial phylogenetic (16S rRNA) and functional profiles (WGS). (b) Estimated species richness and (c) Shannon diversity indices. (d-e) phylum and (f-g) genus level bacterial abundance distribution at time points: day 0 (D00), Trimester1 (T1) and Trimester3 (T3) for *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}*. In each case, significance was determined using nonparametric test and expressed as * $P < 0.05$.

5.2.3 Compositional differences in microbiota of *Atg16l1* mutants and control mice

Distance based constrained analyses (CAP) was performed to delineate the microbiota dynamics at different stages of pregnancy (i.e. baseline, T1, T2 and T3) in *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* mice. Regardless of stage of pregnancy, *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* clustered distinctly in abundance based and non-abundance based analysis. Significance of separation was confirmed by PERMANOVA/Adonis (Bray-Curtis; Genotype, $P=0.000999$ ***; Jaccard; Genotype, $P=0.000999$ ***, $R^2=0.05600$) (Figure 5-2a-b, S3). Interestingly at baseline, microbial composition and structure was not significantly different (PERMANOVA; $P=0.1009$ (Bray); $P=0.101$ (Jaccard)) in *Atg16l1^{ΔIEC}*, although the microbiota started to decipher between *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* with the progression of pregnancy (PERMANOVA; T1, $P=0.0679$ (Bray); $P=0.07592$ (Jaccard)) and became significantly different at trimester three (T3) (PERMANOVA; T3, $P=0.006$ ** (Bray); $P=0.002$ ** (Jaccard)).

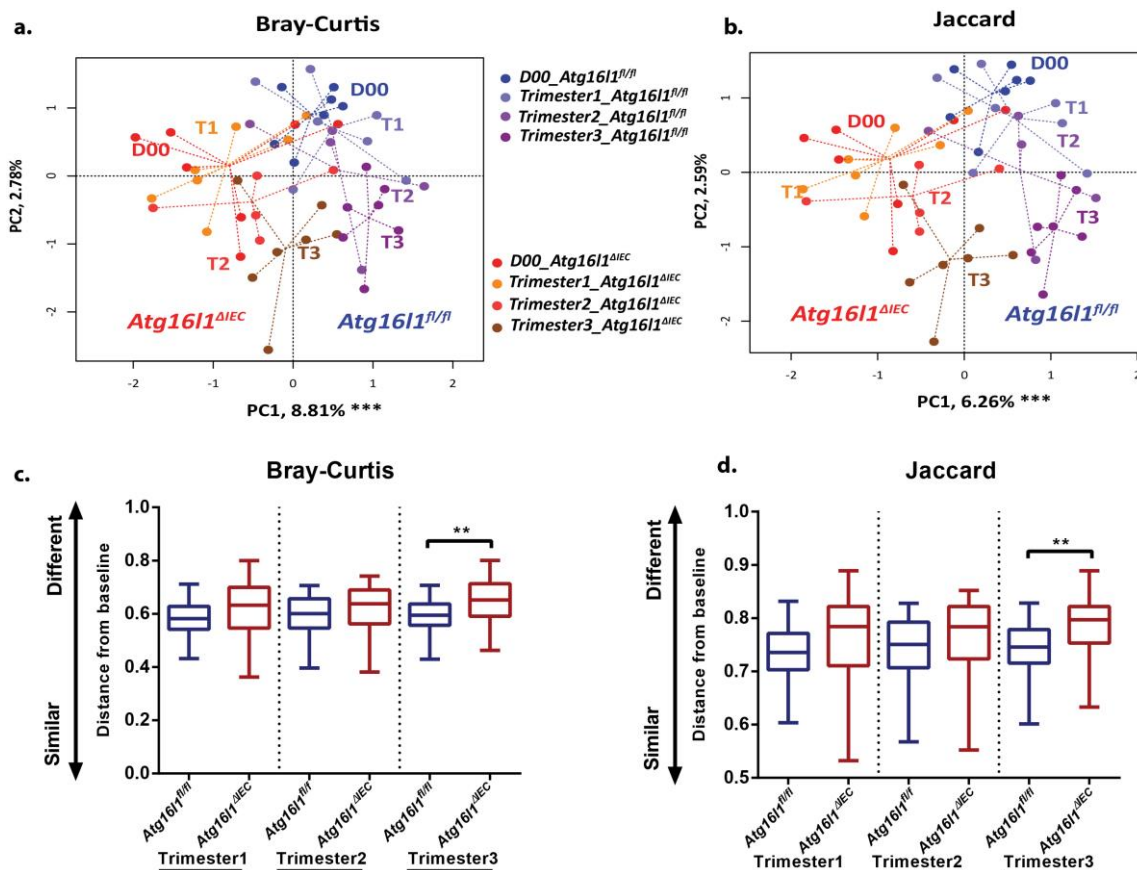


Figure 5-2 Distance based β -diversity analysis.

(a-b) constrained ordination analysis with factors (Trimesters and Genotype) showing pre-pregnancy and pregnancy phases : Baseline (D00), Trimester1 (T1), Trimester1 (T2) and Trimester3 (T3) for *Atg16l1^{fl/fl}* (n=7) and *Atg16l1^{ΔIEC}* (n=7) with (a) Bray-Curtis (b) Jaccard distances (c-d) Boxplot showing (c) Bray and (d) Jaccard distances from baseline (D00) to Trimester1 (T1), Trimester1 (T2) and Trimester3 (T3) for *Atg16l1^{fl/fl}* (n = 7) and *Atg16l1^{ΔIEC}* (n=7).

A combined Bray and Jaccard analysis showed that the beta-diversity between *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* at T1 and T2 were more similar compared to T3. And the beta-diversity at T3 was found significantly higher in *Atg16l1^{ΔIEC}* compared to *Atg16l1^{fl/fl}* (Bray, $P=0.007$; Jaccard, $P=0.008$) (**Figure 5-2c-d**). This indicated that the inter-individual variation of microbiota between *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* animals increased consistently at T1 to T3, however differences reached significance only at T3 between *Atg16l1^{ΔIEC}* and *Atg16l1^{fl/fl}*.

Moreover, longitudinal metastats analysis of OTU distribution at different pregnancy stages showed that bacterial sub-populations at T3 were significantly different (**Figures S4a-b**). The OTU distribution was further compared to identify differences in unique OTUs associated with each group (**Described in methods section**). On comparing the Day 0 and T3 samples for the *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* groups, a general decrement in unique OTUs was observed in *Atg16l1^{ΔIEC}* (**Figures S5a-b**). It was found that a major anaerobe *Clostridia* was significantly reduced in case of *Atg16l1^{ΔIEC}* samples. In addition, a comparison of OTUs at T3 between *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* groups was also done (**Figure 5-3**). It was found that the *Atg16l1^{fl/fl}* group comprised unique OTUs mainly consisting of *Bacteroidia* and *Clostridia*. In contrast, the *Atg16l1^{ΔIEC}* group showed comparatively low *Clostridia* and an altogether absence of OTUs associated with *Bacteroidia*.

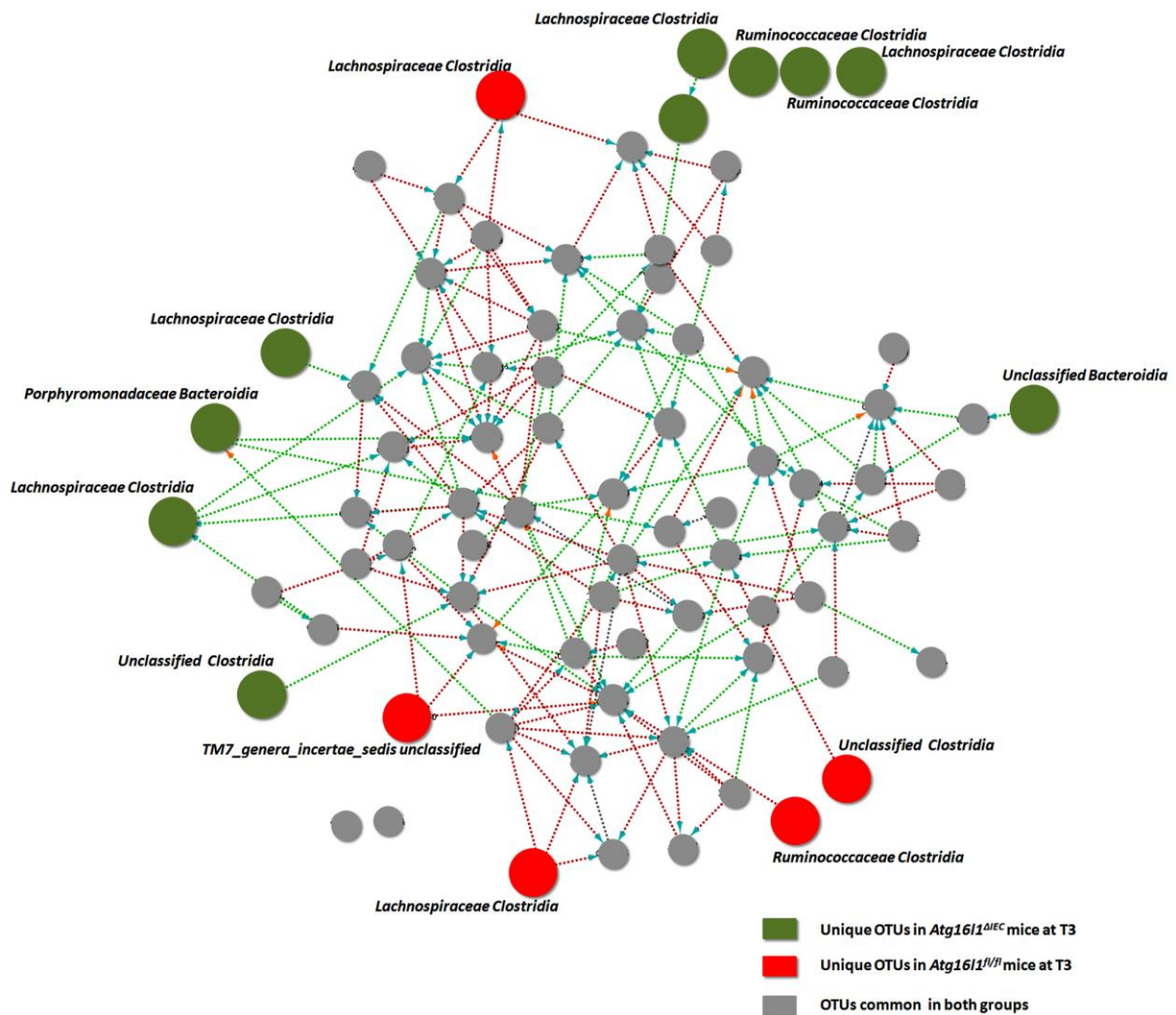


Figure 5-3 Correlation based network showing genotype associated (red and green) unique OTUs and common OTUs (grey) at Trimester 3 (T3).

The unique OTUs at T3 associated with each group are shown (green; *Atg16l1*^{fl/fl}, Red; *Atg16l1*^{ΔIEC}). The arrows represent correlation between the OTUs (Blue; positive, Red; negative). A *P*-value cutoff of 0.01 and a correlation coefficient of ± 0.6 were used for identifying unique OTUs and generating the network respectively.

Further, to identify development of particular enterotypes/community type due to the effect of pregnancy and genotype, DMM (Dirichlet multinomial mixtures) probabilistic modeling was used to cluster microbial communities into meta communities **described in methods section 2.5.1.5** (Holmes et al., 2012). In this study, data from day 0 to T3, the minimum Laplace is $K=1$ that means there were essentially single community type during the entire period. Similarly on analyzing day 0 to month 3 (M3), absence of any new enterotype development was noted (**Table 5a-b**).

Table 5. Metacommunities estimation based on DMM model for pregnancy study.**(a) D00 to T3**

K	NLE	logDet	BIC	AIC	Laplace
1	15835.72	-3299.28	18809.94	18089.72	12114.79
2	22069.5	-7780	28019.26	26578.5	14036.01
3	28270.59	-12481.9	37195.88	35034.59	15813.93
4	34486.78	-17139	46387.61	43505.78	17629.36
5	40663.36	-21906.4	55539.72	51937.36	19350.03

(b) D00 to M3

K	NLE	logDet	BIC	AIC	Laplace
1	59415.96	-1185.4	67055.78	63362.96	55196.22
2	68740.42	-8739.56	84021.99	76635.42	57115.62
3	79384.85	-16641.5	102308.2	91227.85	60181.14
4	90365.08	-24619.3	120930.1	106156.1	63544.48
5	100129.5	-33179.4	138336.3	119868.5	65400.84

5.2.4 Microbial compositional dynamics from pregnancy to post-pregnancy

After demonstrating that pregnancy associated microbiota is different between *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* at T3, it was further questioned if these changes are persistent in postpartum or not. Inter-individual distances of microbial communities started to shrink as early as first analyzed sample at week 1 and the difference became non-significant at month 3 (M3) (**Figure 5-4a-b, S6a-d**). Furthermore to identify specific OTUs which might be associated with genotype differences at postpartum metastats analysis was performed from baseline (D00) to M3. On analyzing the bacterial groups/ OTUs, at day 0 in *Atg16l1^{fl/fl}* mice, the major population included mainly anaerobes such as *Eubacterium* (otu0364), *Porphyromonadaceae* (otu0174), *Lachnospiraceae* (otu0193) and *Ruminococcaceae* (otu0287).

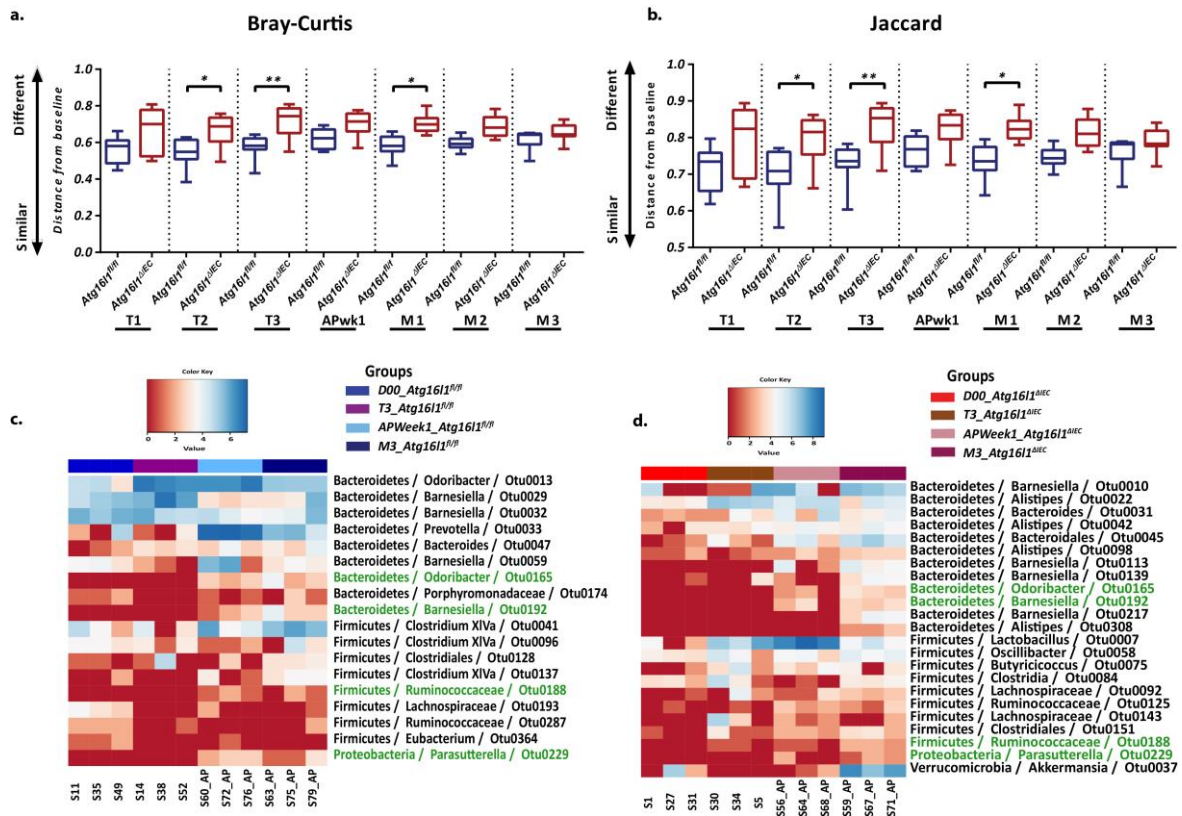


Figure 5-4 Distance based diversity and metastats analysis.

(a-b) Bray and Jaccard distances from baseline (D00) to Trimester1 (T1), Trimester1 (T2), Trimester3 (T3) and post-pregnancy period (APwk1, M1, M2, M3) for both *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}*. (c-d) OTU reported from pairwise comparison from baseline to T3, APwk1 and M3 for both *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}*. Green color labels represent common OTUs in both *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}*.

Besides, another important butyrate producer, *Odoribacter* (otu013) was abundant in (T3) *Atg16l1^{fl/fl}* mice. In addition, *Prevotella* (otu033) (mucin degrading bacteria) was found higher at week 1 and 12th week (M3) after pregnancy. Interestingly, the control mice also showed the presence of important commensal bacteria, *Barnesiella* (otu029, otu032), at both D00 and M3 indicating a better microbial recovery (**Figure 5-4c**). In both the groups (*Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}*) only 4 OTUs (otu0165, otu0188, otu0192, and otu0229) were overlapping over the course of pregnancy and were mostly expressing after pregnancy. Overall, it was evident that the major differences in the bacterial community composition were present during the pregnancy trimesters in both the genotype/groups. Hence, further analysis of the differences in the histological parameters and immune response was studied in the pregnancy period for both groups.

5.2.5 Physiological features of nulliparous and pregnant mice

At baseline, nulliparous *Atg16l1*^{ΔIEC} animals showed a marginally low body weight ($P=0.01$) and spleen weight ($P=0.04$) as compared to control animals (*Atg16l1*^{fl/fl}) (Figure 5-5a, 5-5c). Further, colon length and liver weight of *Atg16l1*^{fl/fl} and *Atg16l1*^{ΔIEC} animals were comparable at baseline as well as at T3. These organs mostly showed the dynamical changes associated with pregnancy where the trend of increase in colon length and liver was observed from baseline to T3 in both *Atg16l1*^{fl/fl} and *Atg16l1*^{ΔIEC} (Figure 5-5b, 5-5e). Nonetheless, at third trimester (T3), caecum weight in *Atg16l1*^{ΔIEC} mice were significantly higher than *Atg16l1*^{fl/fl} animals ($P=0.03$) (Figure 5-1d). Besides, no considerable differences in embryo length were observed (Figure 5-5f). Further, the histological examination of the organ sections corroborated the observed differences in each mice group. In addition, a comparison of the puppy weight and length in each group showed significantly higher weight of *Atg16l1*^{ΔIEC} puppies ($P>0.01$) (Figure S7).

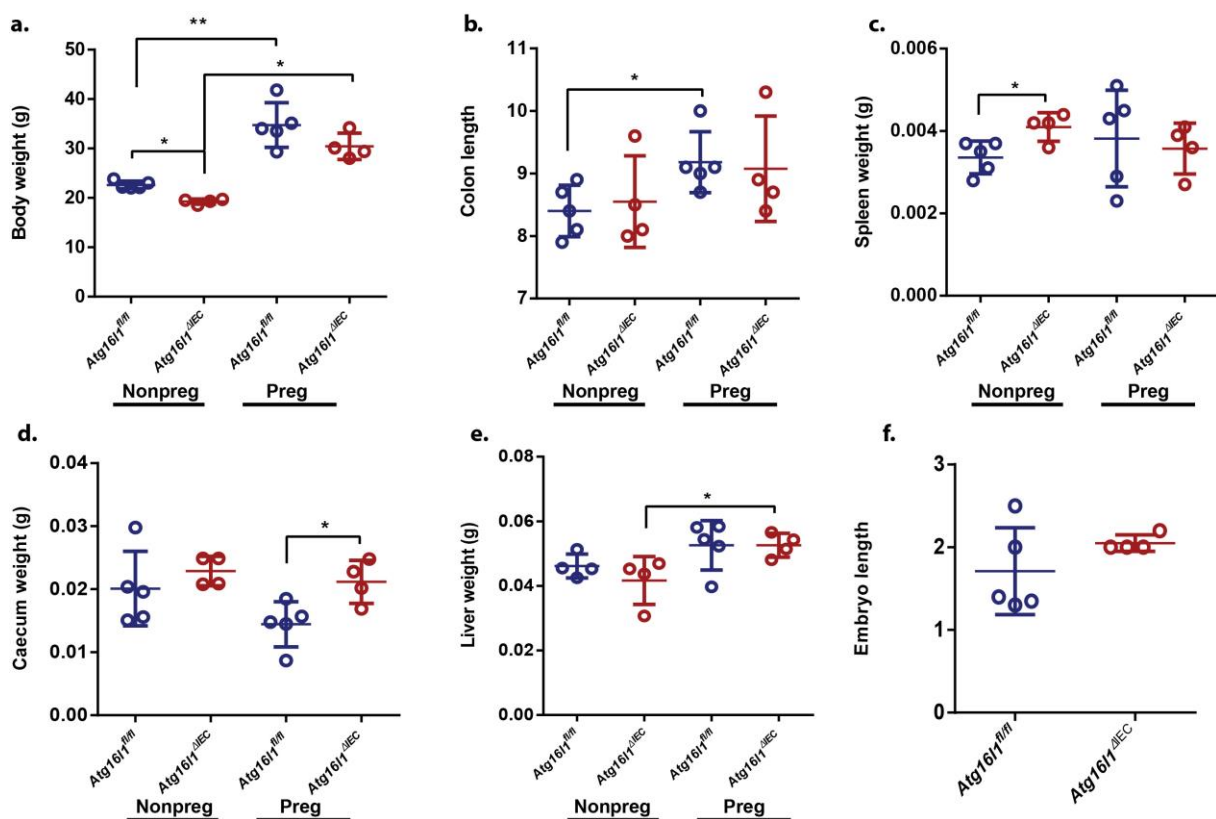


Figure 5-5 Histological phenotypes for nulliparous and pregnant mice.

Histological phenotypes for nulliparous *Atg16l1*^{fl/fl}, nulliparous *Atg16l1*^{ΔIEC}, pregnant *Atg16l1*^{fl/fl} and pregnant *Atg16l1*^{ΔIEC} showing (a) Body weight (b) Colon length (c) Spleen weight (d) Caecum weight (e) Liver weight (f) Embryo length. Significance of differences was determined using nonparametric Mann Whitney test * $P<0.05$ and expressed as the mean \pm SD.

5.2.6 Analysis of inflammation states of pregnant and non-pregnant mice

After observing pregnancy associated differences in major body organs at the edge (last trimester, T3) of pregnancy in a genotype dependent manner, I further investigated if pro- and anti-inflammatory cytokine levels were also altered. Interestingly, significantly increased levels of pro-inflammatory cytokine TNF α ($P=0.01$) and CXCL1 ($P=0.03$) in *Atg16l1* ^{Δ IEC} animals at T3 as compared to the pregnant control animals was observed. The rise in CXCL1 ($P=0.02$) and TNF α ($P=0.057$) expression was also significantly higher to nulliparous *Atg16l1* ^{Δ IEC} animals (**Figure 5-6a-b**). Besides, anti-inflammatory cytokine IL-10 level were also found higher in both genotypes at T3, albeit differences were significant only between nulliparous and pregnant *Atg16l1* ^{Δ IEC} mice ($P=0.02$) (**Figure 5-6c**).

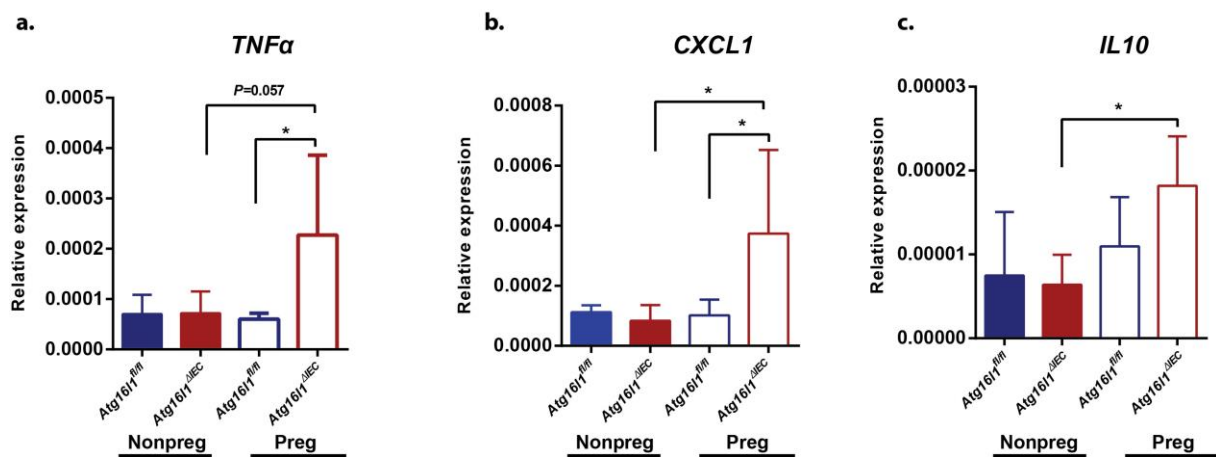


Figure 5-6 Expression levels of the cytokines in nulliparous and pregnant *Atg16l1*^{fl/fl} and *Atg16l1* ^{Δ IEC} mice. (a-c) showing (a) TNF α (b) CXCL1 and (c) IL10 expression by Taqman assay. Groups are shown as nulliparous *Atg16l1*^{fl/fl}, nulliparous *Atg16l1* ^{Δ IEC}, Pregnant *Atg16l1*^{fl/fl} and Pregnant *Atg16l1* ^{Δ IEC}. Significance determined using nonparametric Mann Whitney test and expressed as the mean +/- SD. * $P<0.05$.

5.2.7 Quality assessment and bacterial profiling of metagenomics data

To examine longitudinal changes in the functional hierarchy of microbiome in *WT* and *Atg16l1* ^{Δ IEC} pregnant mice, metagenomics analysis was carried out. The quality check of resulting sequences showed approximately ~70% reads coverage for all metagenomics samples (**Figure 5-7a**) (**Annexure-V Table 1**). Next, the assessment of level of coverage in the metagenomics samples was done by rarefaction that showed saturation in species count (**Figure 5-7b**) (**Annexure-V Table 2**). This signified satisfactory sampling in all cases in the metagenomics data set. Further, the reads were analyzed through representative hits in the MG-RAST that showed close to 99% reads belonging to bacterial communities (**Figure 5-8a**).

Additionally, these representative hits were also used to quantify bacterial phylotypes in both *WT* and *Atg16l1*^{ΔIEC} samples. The analysis of phylotypes supported the results observed in the 16S rRNA gene profiling. At phylum level, reduction in *Bacteroidetes* population at T3 was noted in all samples. However, the reduction was found higher in case of *Atg16l1*^{ΔIEC} pregnant mice corroborating the 16S rRNA data (**Figure 5-8b**). At the genus level, a significant increase in *Lactobacillus* population at T3 in all samples was observed. But, the increase in *Lactobacilli* population was found significantly higher in the *WT* mice (~ 35%) as compared to the *Atg16l1*^{ΔIEC} mice (~16%).

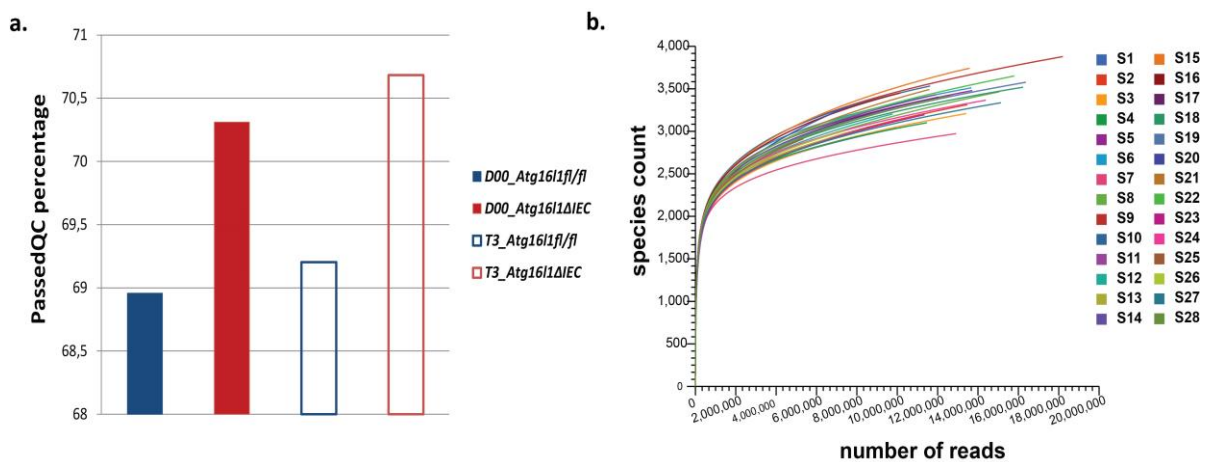


Figure 5-7 Quality check and rarefaction curves for metagenomics reads.

Metagenomics reads for *Atg16l1*^{fl/fl} and *Atg16l1*^{ΔIEC} at D00 and T3 showing (a) Bar graph showing percentage reads passed the quality control (b) rarefaction curve depicting satisfactory sampling.

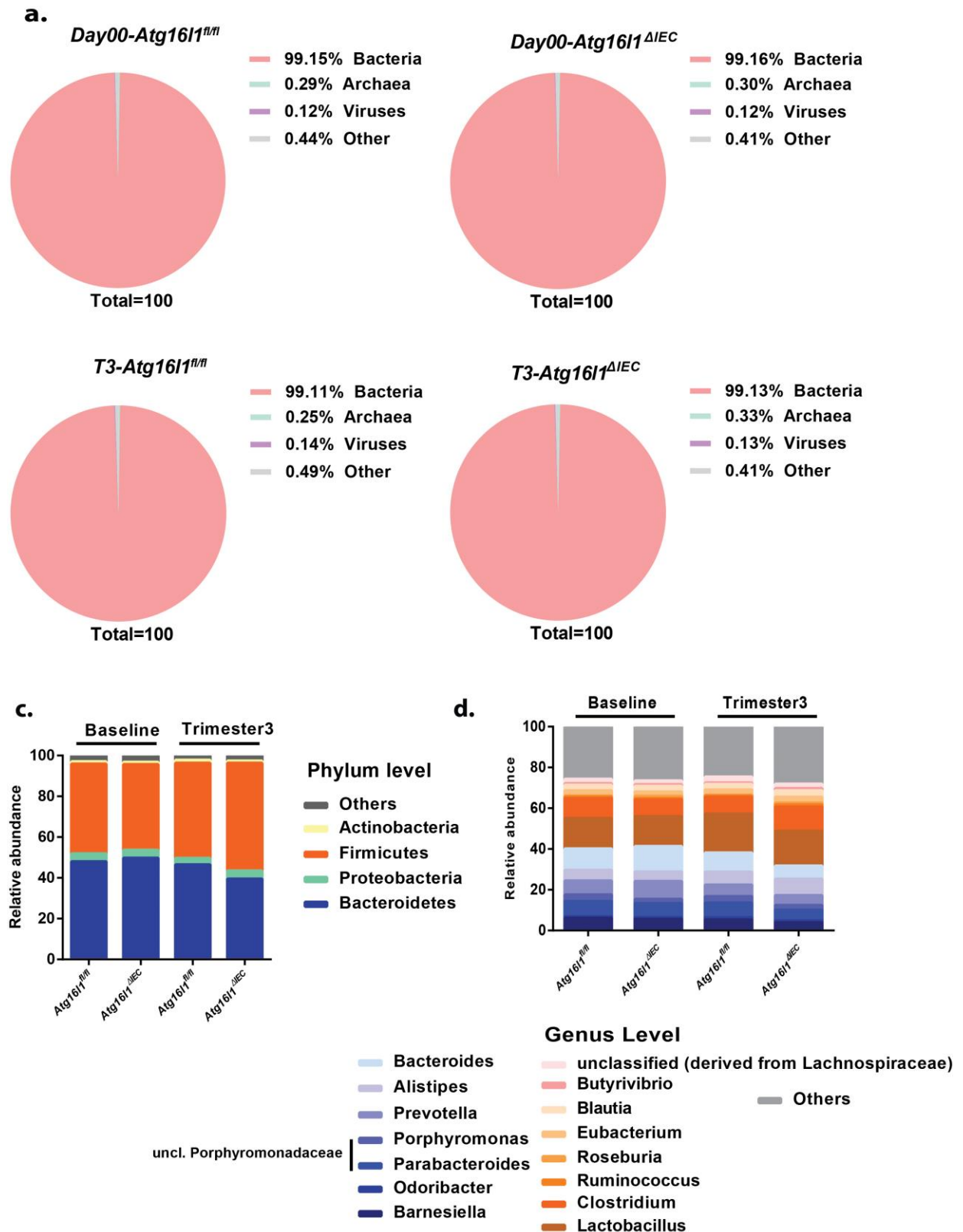


Figure 5-8 Representative hit based bacterial profile of metagenomics reads.

(a) Pie chart representation at domain level (b-c) Phylum and genus level profiles in *Atg16l1*^{fl/fl} and *Atg16l1*^{ΔIEC} samples at D00 and T3.

In addition, a gradual increase in butyrate producing *Odoribacter* from day 0 to T3 was also noted in *WT* mice samples which was conspicuously found reduced in mutant mice. The mutant mice also harbored significantly reduced population of small chain fatty acid (SCFA) producer *Bacteroides* at T3 (compared to day 0) as compared to control mice. Besides, approximately 50% reduction in the mucin degrading bacteria, *Prevotella* along with *Barnesiella* spp. at T3 was noted in the mutant mice. The later has been shown to play important role in controlling the intestinal infection of antibiotic resistant pathogenic bacteria such as *Enterococcus* (Ubeda et al., 2013).

On the contrary, the mutant mice at T3 also showed almost doubled proportions of another important SCFA producer, *Alistipes*. Besides, the *Clostridium* population was also found significantly higher at T3 in the mutant mice. *Clostridium* are classified as important butyrate producers and have been implicated in regulation of gut homeostasis (**Figure 5-8c**) (Lopetuso et al., 2013).

5.2.8 Clustering and functional profiling of metagenomics samples.

The longitudinal analysis of diversity in the metagenomics reads from both *WT* (*Atg16l1^{fl/fl}*) and *Atg16l1^{ΔIEC}* samples was done to identify any influence during the course of pregnancy. For this, a two-way clustering of metagenomics reads *i.e.* species level and function level was done. The *WT* mice showed similar overlaps for basal (day 0) and T3 samples indicating lower variations between the two time points at both levels (**Figure 5-9a**). However, the mutant mice samples showed a large gap in clusters of two time points at both levels, indicating major alterations.

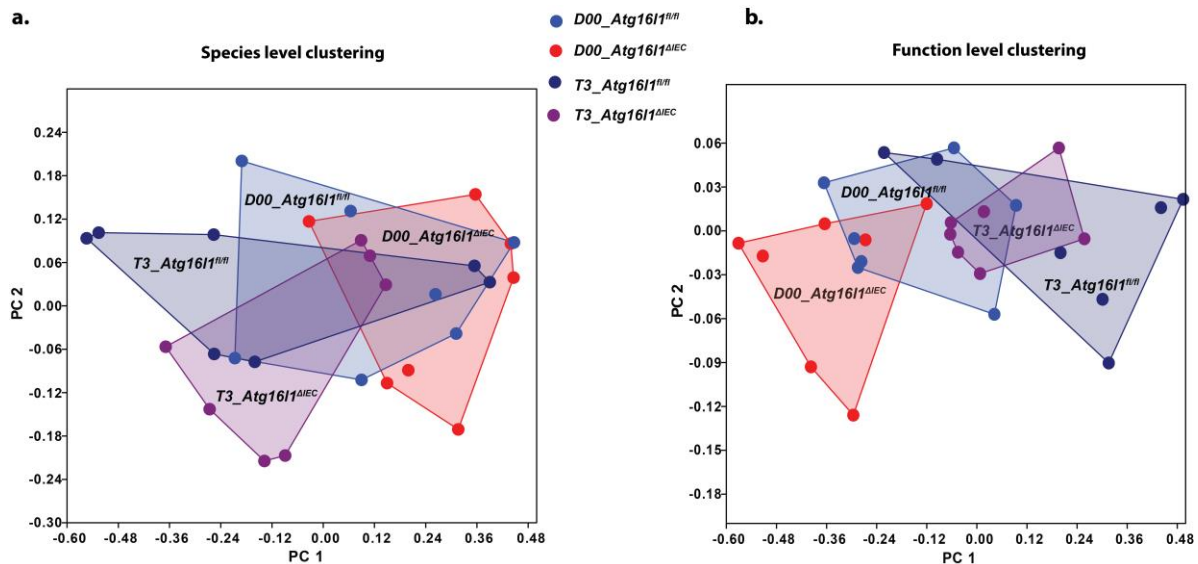


Figure 5-9 Distance based clustering of metagenomics.

The metagenomics reads were subjected to Canonical correspondence analysis (CCA) where both (a) Species level and (b) Function level clustering is shown for *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* samples at D00 and T3.

The difference in the function level clustering was more prominent suggesting major changes in the metabolic functions and pathways at T3. To discern these functional changes, the metagenomics reads were analyzed using four levels of functional categories with increasing specificity based on the mapping to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (**Annexure-V Table 3 Electronic copy**). At level 1, genes involved in metabolism and genetic information processing were found to be the largest contributors in all samples without a clear difference (**Figure 5-10**).

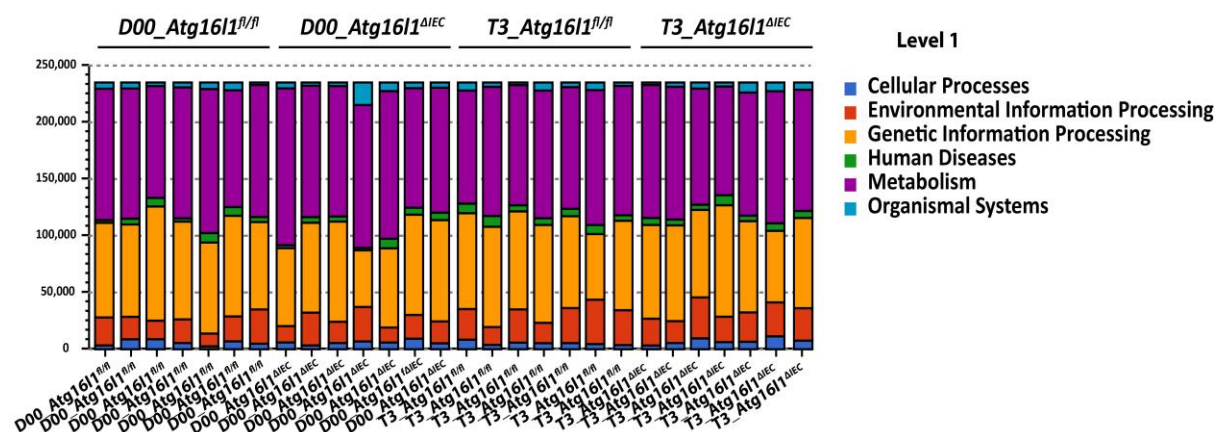


Figure 5-10 Level 1 functional categories distribution in metagenomics samples.

The Level 1 KEGG functional categories distribution for all samples corresponding to D00 and T3 for *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}*.

Similarly, on analyzing level 2, the genes associated with signal transduction, carbohydrate metabolism, and membrane transport were found prominently present in all samples (**Figure 5-11**). In all, at both the levels, not many discernible pathways were found significantly altered in any sampling group.

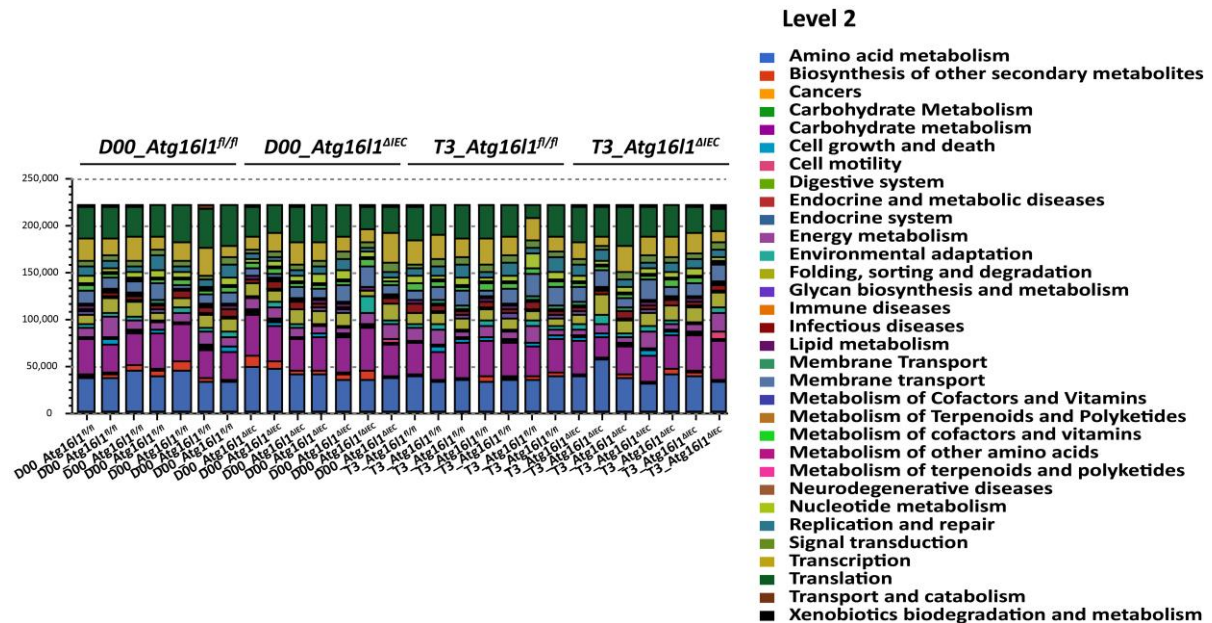


Figure 5-11 Level 2 functional categories distribution in metagenomics samples.

The Level 2 KEGG functional categories distribution for all samples corresponding to D00 and T3 for *Atg16l1*^{fl/fl} and *Atg16l1*^{ΔIEC}.

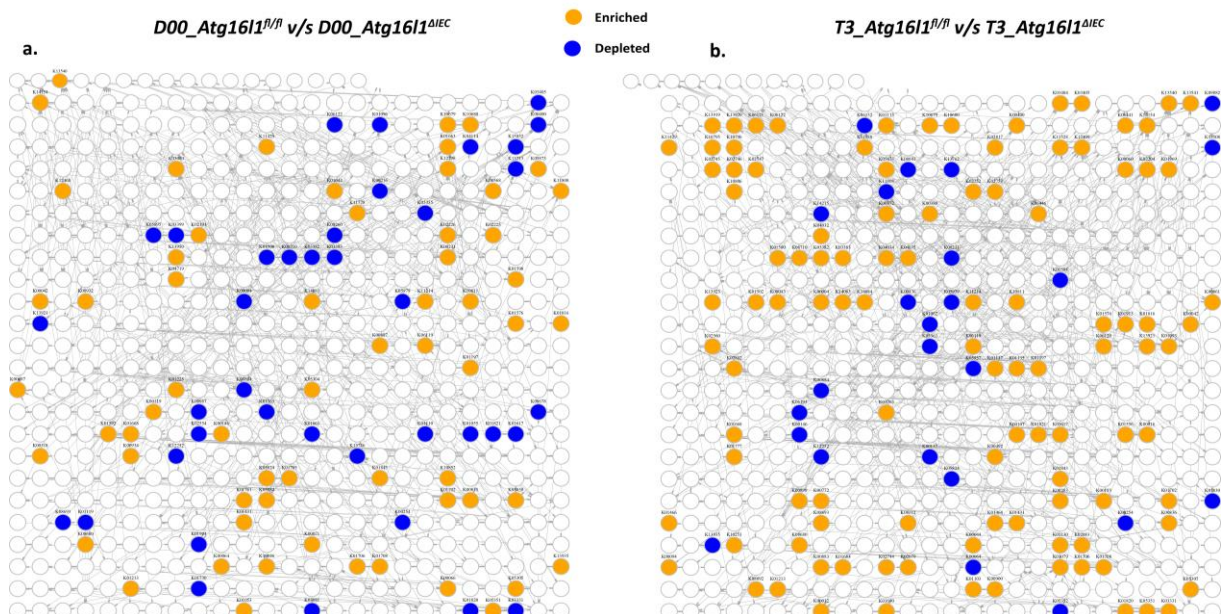


Figure 5-12 Level 4 functional enrichment analysis in metagenomics samples.

The Level 4 functional categories distribution as 'enriched' (orange) and 'depleted' (blue) metabolites for all samples corresponding to (a) D00 *Atg16l1*^{fl/fl} versus *Atg16l1*^{ΔIEC} and (b) T3 *Atg16l1*^{fl/fl} versus *Atg16l1*^{ΔIEC} with degree=20.

Following this, the level 4 functional analysis was done where the 'enrichment' and 'depletion' of metabolites listed in KEGG were investigated using "mmnet" Bioconductor package for metagenomics analysis (**Figure 5-12**). In order to avoid the complexity of network degree 20 was considered (excluding nodes having edges more than 20). Interestingly it was found that the number of enriched and depleted pathways for day 0 between *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* were 62 and 39 (**Figure 5-12a**) whereas at T3 enriched pathways increased to 107 and depleted pathways were reduced to 25 (**Figure 5-12b**). Four were common between depleted and 23 were common between enriched group of D00 and T3. In all, the WT and *Atg16l1^{ΔIEC}* samples at basal level (day 0) and T3 were compared to obtain a list of KEGG orthologs (**Annexure-V Table 4 Electronic copy**). As seen in all comparisons, the samples showed a variation in distribution of certain metabolites. This indicated that the samples differed in certain metabolic pathways which were first catalogued to compare their presence or absence in different samples. The obtained results were first represented by LEfSe (Linear discriminant analysis Effect Size) analysis with effect size ≥ 2 with P value of significance $p \leq 0.05$ (**Figure 5-13**). The LEfSe analysis showed major alterations in the *Atg16l1^{ΔIEC}* samples both at basal level as well as T3. The mutant samples showed alterations majorly in carbohydrate metabolism, fatty acid biosynthesis, amino acid biosynthesis and metabolism of cofactors and vitamins. It was important here to understand that metabolic pathways involve consortium of bacterial species that produce different metabolic intermediates. Henceforth, pathways influenced in the mutant samples were compared in detail with the WT samples to identify critical differences in presence/absence of metabolites.



Figure 5-13 Linear discriminant Effect Size (LEfSe) analysis at level 4 KEGG functional categories.

Explaining the significant different between *Atg16l1*^{fl/fl} and *Atg16l1*^{ΔIEC} at D00 and T3 due to functional alterations. Labels are colored according the level1 functional categories (Green=Environmental information processing, Orange=Metabolism, Light green=Genetic information processing).

For comparing the carbohydrate metabolism, the core metabolic network of glycolysis/gluconeogenesis pathway was analyzed. It was found that the metagenomics hits distribution of two crucial enzymes *i.e.* acetyl-CoA synthetase (EC 6.2.1.1) were upregulated in *WT* mice at day 0 and T3 respectively whereas pyruvate kinase (EC 2.7.1.40) were enriched only in T3 (**Figure 5-14**). Both these enzymes play important role in butyrate degradation and associated energy metabolism in the gut which was found altered in case of mutant mice (Donohoe et al., 2011; Gao et al., 2014; Turnbaugh et al., 2008; Wolfe, 2015). In case of fatty acid metabolism, the mutant mice at basal level showed increased abundance of FabF that later at T3 got equally distributed in both *Atg16l1*^{fl/fl} and *Atg16l1*^{ΔIEC} samples (**Figure 5-15**). This protein comes from a cluster of bacterial genes involved in the biosynthesis of saturated fatty acids by commensal *Lactobacillus* (Chen et al., 2015). Interestingly, our earlier observation showed increase in *Lactobacilli* at T3 to be significantly higher in the control mice (~35%) as compared to the mutants (~16%). Next, the analysis of amino acid metabolism showed that at T3 an important enzyme aspartate kinase (EC 2.7.2.4) associated with lysine biosynthesis was found upregulated in *WT* mice only. On the contrary, increased abundance of another enzyme, argininosuccinate lyase (EC 4.3.2.1) was observed in case of Alanine/aspartate/glutamate metabolism pathway only in mutant mice.

Interestingly, the mutant mice harbored dysregulations mostly in Glycine-serine-threonine metabolism, where crucial enzymes such as phosphatidylserine synthase (EC 2.7.8.8) and serine deaminase (EC 4.3.1.17) at day 0 and phosphoserine transaminase (EC

2.6.1.52) at T3 were found upregulated (**Figure 5-16**). Further in case of metabolism of cofactors and vitamins, it was found that another important enzyme nicotinate-nucleotide adenyltransferase (EC 2.7.7.18) was upregulated in WT mice. All these evidences, suggested significant alterations in the metabolic pathways due to changes in microbial compositions during pregnancy. Although the basal microbiota composition looked similar in both sampling group, the *Atg16/1* ^{Δ IEC} samples showed differences in metabolites associated with major pathways. All the associated pathways are provided in **supplementary figure as Electronic Copy – S8**.

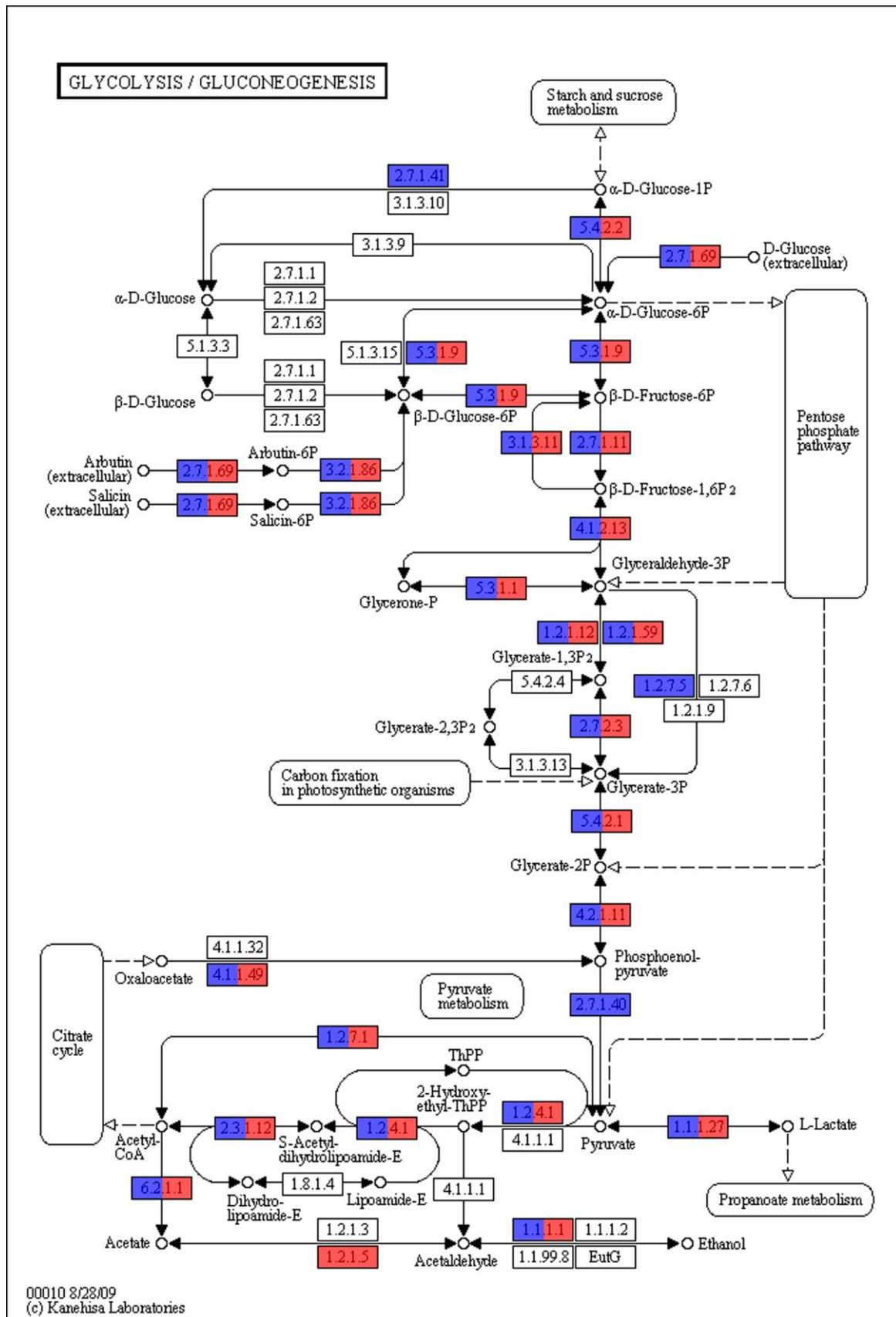


Figure 5-14 KEGG pathways for Glycolysis and Gluconeogenesis.

Color represents the number of genes mapped on pathway for each group where blue represents *T3_Atg16l1^{fl/fl}* and red represents *T3_Atg16l1^{ΔEC}*.

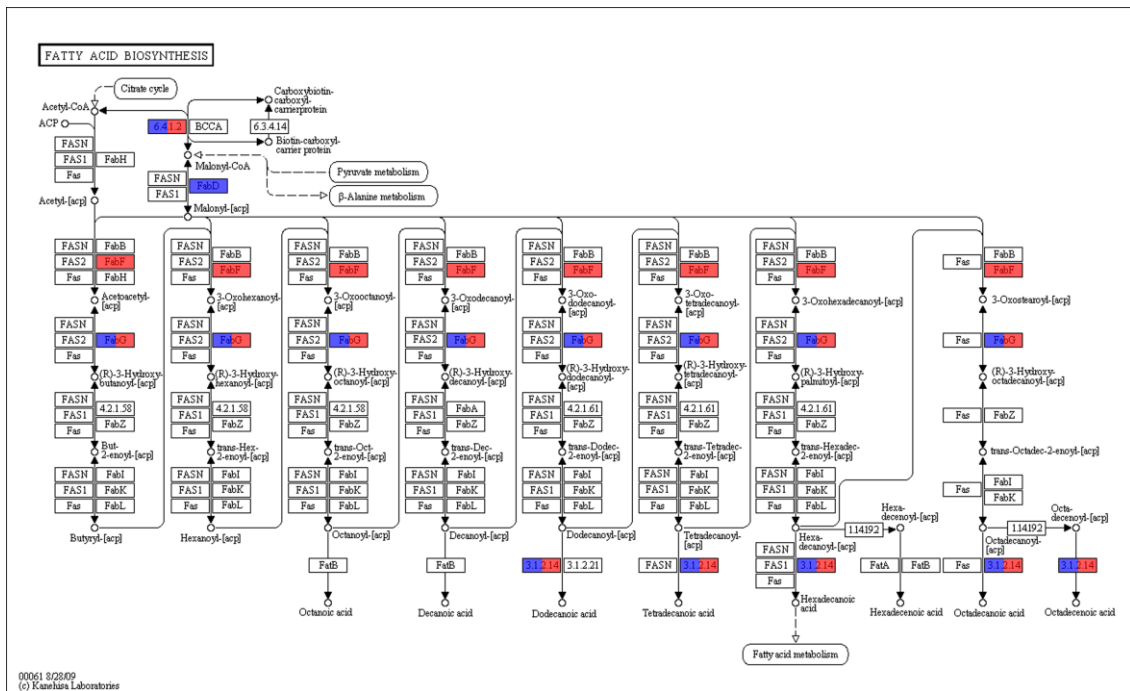


Figure 5-15 KEGG pathways showing Fatty acid biosynthesis.

Color represents the number of genes mapped on pathway for each group where blue represents *D00_Atg16l1^{fl/fl}* and red represents *D00_Atg16l1^{ΔIEC}*.

(Cadwell et al. 2008). These abnormalities further affect the intestinal homeostasis regulated by antimicrobial peptides and enzymes produced from Paneth cells. The present work attempts to study the long term effects of disrupted secretion of antimicrobial peptides in mice genetically deficient in *Atg16l1* during pregnancy phases. A comparative analysis with normal controls was done to catalogue the microbial components that get altered and their incurring outcomes on the inflammatory response during different phases of pregnancy were studied.

At basal level, no significant differences in the physiological and histological features of control and *Atg16l1*^{ΔIEC} mice were noted. However, during the course of pregnancy, a significant increase in the body weight was noted along with major changes in the organ sizes to accommodate the growing fetus. This process is associated with maintenance of energy efficiency and the metabolic fluctuations during the stages of pregnancy. The energy metabolism is precisely adjusted to conserve the essential nutrients required by the fetus (Whitaker et al., 2012). At later stages, the metabolism of nitrogenous waste is regulated to allow positive nitrogen retention (King, 2000). Interestingly, in case of *Atg16l1*^{ΔIEC} mice, a significantly higher caecum and liver weights were observed. In general, an increased caecum weight is associated with digestive disturbances leading to higher digesta retention (Kuo et al., 2013). This closely relates to corresponding increase in liver weight indicating metabolic disturbances. In pregnancy, an increase in colon length is generally observed that helps in higher waste elimination processes associated with pregnancy. However, this observation was only found in case of control mice and not in the *Atg16l1*^{ΔIEC} mice indicating higher digesta retention in latter.

All these observation also supported the previously reported *Atg16l1* associated transmural inflammation in the small intestine through the colon and caecum that may have associated symptoms in form of digestive and metabolic disturbances (Cadwell et al., 2009; Lassen et al., 2014). Further on analyzing the inflammation states of mice during the different stages of pregnancy, subtle changes in the level of circulating cytokines in the pregnant *Atg16l1*^{ΔIEC} mice were observed. An overall increase in levels of Cxcl1, Il10 and TNFα were noted for the mutant mice at T3. It has been previously shown that mice lacking the ΔCCD fragment of *Atg16l1* (residues 70-212) result in an increased secretion of CXCL1 in response to pathogenic bacterial infection (Sorbara et al., 2013). This was implicated in a

domain specific control of cytokine expression and autophagy by *Atg16l1* gene. On the other hand, increase in IL10 expression has been shown to be a genotype effect of *Atg16l1* deletion (Glubb et al., 2011). Besides, it was also interesting to find an increase in TNF α expression which was previously found to be at normal levels during *Atg16l1* deficiency (Saitoh et al., 2008). All these evidences indicated an altered physiological and inflammatory state of the *Atg16l1* deficient pregnant mice. *Atg16l1* associated impairment of immune response and inflammation has been recently found associated with alterations in the intestinal microbiome composition (Adolph et al., 2013).

Next, the time lapse changes in the intestinal microbiome were investigated during the course of pregnancy in the control and *Atg16l1* deficient mice. Significant differences were found in microbial richness of *Atg16l1*^{AIEC} samples between day 0 and T3 as compared to control mice. Further, a combined 16S rRNA profiling and metagenomics approach was taken to decipher the compositional differences in microbiome during and after pregnancy. During the past several years, high-through put sequencing efforts have established the role of certain bacterial product and substrates as crucial components that regulate immune and inflammatory response in body (Belkaid and Hand, 2014; Kuo, 2013). For a healthy gut consortium, production of lactate is followed by its dissimilation into either butyrate or other short chain fatty acids (SCFA) such as acetate, succinate etc. (Duncan et al., 2004). These byproducts subsequently are involved in several physiological and immunogenic functions inside the gut. Butyrate is utilized for production of mucin that helps in maintaining the tight junctions between cells (Bourriaud et al., 2005).

Several other bacteria such as *Akkermansia* and *Prevotella* are involved in mucin degradation and metabolism (Derrien et al., 2010; Rho et al., 2005). Both the 16S rRNA and metagenomics analysis revealed reduction in major butyrate producer *Odoribacter* in the *Atg16l1* mutants from T1 to T3. Interestingly, a major SCFA producer *Bacteroides* and *Alistipes* were found significantly reduced and doubled respectively in case of mutant mice at T3. In addition, the mucin degrading *Prevotella spp.* was also found reduced by more than 50% in case of mutant mice at T3. The process of mucin sustentation is important for the barrier function of the epithelial cells that sustains the gut integrity and determines the defense and immunological responses inside the gut. Besides, other SCFAs such as acetate and propionate act as substrates for lipogenesis, cholesterol synthesis and also help in

gluconeogenesis by liver (Belkaid and Hand, 2014; Comalada et al., 2006). In addition, SCFAs also act on leukocytes and endothelial cells and regulate several leukocyte functions including production of cytokines (Dalal and Chang, 2014; Vinolo et al., 2011). Further, Several studies have highlighted molecular mechanisms of gut microbiome associated metabolic pathways such as catabolism of xenobiotics, drugs and assimilation of dietary components and exclusion of toxic components (Spanogiannopoulos et al., 2016). This observation strongly emphasized that alteration in *Atg16l1* gene hampers the normal functioning of microbial pathways plausibly due to alterations in the microbial compositions during the phases of pregnancy. The comparative analysis suggests that the microbial composition in the mutant mice gets directed towards an inflammatory and low immune reactive state. In this case, not only the mediators of energy metabolism (lowered butyrate producer) are affected but also the inflammatory indicators (in form of SCFA) and gut barrier regulatory mechanism (reduction in mucin degrading *Prevotella*) is also altered. In addition, few other variations in mutant mice such as reduction in *Lactobacillus* and *Barnesiella*, that have been shown to restrict colonization of bacteria and mediating the body's immune system (Ubeda et al., 2013). This altered state of microbiota is further confirmed by an overall increase in *Clostridium* in the mutant mice. The commensal *Clostridia* has been previously shown to contribute in the maintenance of overall gut function (effect on caecum size, butyrate production and lowering of inflammatory response) during unfavorable alteration of microbiota composition or dysbiosis (Lopetuso et al., 2013).

The dysbiosis event was further substantiated by the unconstrained and constrained analysis of T3 composition that showed clear deviations between the control and mutant mice. Further, the metagenomics data was longitudinally analyzed by a two-way clustering of metagenomics reads at species level and function level was done. Here, the *Atg16l1* mutant samples showed a large gap in clusters of two time points at both levels, indicating major alterations. However, on further analyzing functional level changes based on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, no significant alterations were observed. Thus, the analysis of metabolites involved in pathways and produced by different microbiome communities was assessed to investigate the differences. The 'enrichment' and 'depletion' map of KEGG listed metabolites were compared that showed critical variations in their distributions. The *Atg16l1*^{ΔIEC} samples showed alterations majorly in carbohydrate metabolism, fatty acid biosynthesis, amino acid biosynthesis and metabolism of cofactors

and vitamins. In case of carbohydrate metabolism, two important enzymes i.e. acetyl-CoA synthetase (EC 6.2.1.1) and pyruvate kinase (EC 2.7.1.40) were found upregulated in WT mice at day 0 and T3 respectively. Both these enzymes play important role in butyrate degradation and associated energy metabolism in the gut which was found altered in case of mutant mice (Donohoe et al., 2011; Gao et al., 2014; Turnbaugh et al., 2008; Wolfe, 2015). This suggested that the adjustment of normal energy metabolism to conserve essential nutrients required by the fetus was compromised in the *Atg16l1*^{ΔIEC}. Similarly, in case of fatty acid metabolism, the mutant mice showed upregulated FabF which is a major *Lactobacillus* protein involved in the biosynthesis of saturated fatty acids. Its high basal level presence could have plausibly deviated the general increase in protective *Lactobacilli* that showed comparatively higher amounts in control mice at T3.

In addition, the analysis of amino acid metabolism also showed critical differences in certain enzymes in both the control and mutant mice. The *Atg16l1*^{ΔIEC} mice showed alterations mainly in the Glycine-serine-threonine metabolism, where crucial enzymes such as phosphatidylserine synthase (EC 2.7.8.8) and serine deaminase (EC 4.3.1.17) at basal level and phosphoserine transaminase (EC 2.6.1.52) at T3 were found upregulated. This is interesting since these set of amino acids are associated with the production of Glutathione (GSH) by liver. The GSH production plays a key role in reducing oxidative stress and has been shown to get influenced by intestinal microbiota (Mardinoglu et al., 2015). All these evidences, suggested significant alterations in the metabolites associated with critical pathways due to changes in microbial compositions during pregnancy.

As a whole, the community composition in the mutant mice showed lowering of energy metabolism and inflammatory indicators. A closure look corroborated this change in composition through major alterations in metabolites associated with carbohydrate and fatty acid metabolism. This also indicates that Paneth cell defect may lead to formation of faulty microbial communities that lead to major changes in energy metabolism and inflammatory responses during different phases of pregnancy. The present study highlights the role of *Atg16l1* and other IBD susceptibility genes in deviating correct and protective microbial components during pregnancy which may have serious pathophysiological implications.

6 Conclusion and outlook

6.1 Conclusion

The diverse and complex microbial community residing inside the human gut plays a central role in human health. The composition of this microbiota is often specific to an individual, evolves throughout the lifetime and susceptible to several environmental as well as host specific factors. Several lines of evidence point towards the role of microbiota in development and modulation of immune response as well as maintaining physical barrier to incoming pathogens and production of antimicrobial substances. Besides, the microbiota is also involved in numerous metabolic processes such as nutrient acquisition and xenobiotic processing and contributes immensely in maintaining homeostasis. Thus, the microbiota composition varies in presence of environmental stress such as antibiotics, alcoholism, pathogen invasion and disease. However, the effect of these compositional variations and their association with disease or inflammatory state of the body is still debated. In the first section of my thesis (**Chapter 3**), the intestinal microbiota changes during acute appendicitis are studied. A comparative analysis of patients (inflamed appendicitis) and healthy human subjects (non-inflamed appendicitis) has been done using 16S rRNA sequencing. The study revealed an overall decrease in bacterial abundance and diversity in the inflamed appendices. Further analysis showed that inflamed samples had high expression of anti-microbial peptides (AMPs) as well as high expression of bacterial sensor gene NOD2 and TLR1. Overall, these data significantly correlated with overall decrease in bacterial diversity, accumulation of certain bacteria (e.g. *Fusobacterium*) and suggested its important role in development of inflammatory response in acute appendicitis.

Another important aspect of microbiota is its intrinsic ability to recapture after a catastrophic event such as pathogenic invasion, disease or even antibiotic administration. The microbial composition remains resilient towards perturbations over lifetime and different nonspecific and specific fluctuations in its composition are indicative of acute or long term pathological states. The Paneth cells present in the crypts of intestine are associated with the maintenance microbial homeostasis mainly by the release and modulation of AMPs. Besides, it has been also established that defect in Paneth cells may initiate an inflammatory response during catastrophic events. Thus, in the next part of my work (**Chapter 4**), role of functional Paneth cells during the recovery after a catastrophic

event and its link to *Atg16l1* gene is explored under antibiotic stress. Here, mice with conditional knockout of the *Atg16l1* gene in their intestinal epithelium (*Atg16l1*^{ΔIEC}) were studied during and after antibiotic administration. The study showed that antibiotic induced disturbances in microbial recovery affects normal (wild-type) and *Atg16l1*^{ΔIEC} mice in a similar way. On analyzing the bacterial phyla in both groups, a dichotomous distribution of *Bacteroidetes* and *Firmicutes* in samples was and thus two community types *Bacteroidetes* rich (BT; similar to native, Day 0 composition) or *Firmicutes* rich (FT; similar to post-Antibiotic composition) based on relative abundance were identified. The study showed that impaired Paneth cells were not correlated with development of these community types and is solely a genotype (*Atg16l1*^{ΔIEC}) based effect. The data presented in this chapter demonstrated that dysbiotic events act together with genetic variations such as *Atg16l1*^{ΔIEC} and affect the resilience potential of the microbiota leading to alterations in immune balance and may thus induce disease like symptoms.

Accumulating evidences suggest that the gut microbiome also plays a crucial role in maternal and child health outcomes. The state of maternal microbiome during pregnancy is now known to impact neonatal and infant health and its alterations has been also linked to preterm births, metabolic complications and gestational weight gain. However, how the microbiome components actually exert their effect during and after pregnancy is still not known. Thus, the last part of this work (**Chapter 5**) was focused on understanding compositional changes in microbiome during different phases of pregnancy. To study the effects of Paneth cells abnormalities which in turn lead to disturbed microbial homeostasis, mice genetically deficient in *Atg16l1* during pregnancy phases were also studied. The presented work takes into account a longitudinal analysis of microbiome variations using a comprehensive 16S rRNA profiling and metagenomics analysis during different phases of pregnancy. The results show that the microbiome composition gets influenced in *Atg16l1* deficient mice during different phases of pregnancy. The changes in the microbial composition were accompanied by changes in secretion of inflammatory cytokines in the mutant mice. The analysis of metagenomics data showed that although no global alterations in the metabolic pathway were noted, there were significant variations in the expression of crucial metabolites associated with carbohydrates and fatty acid metabolism. This suggested that the adjustment of normal energy metabolism to conserve essential nutrients required by the fetus was compromised in the *Atg16l1* deficient mice. These observations were also

supported by the presence of clinical symptoms such as significantly higher caecum and liver weights in the mutant mice that has been linked previously with digestive and metabolic disturbances. As a whole, the present work provides important clues on the critical role of bacterial species in maintaining conducive environment for the growth of fetus. And any compositional alterations resulting due to the presence of IBD susceptible genes affecting the microbial homeostasis may affect the energy metabolism as well as the inflammatory response during pregnancy.

6.2 Outlook

The variations in gut microbiome have been linked to development of metabolic and inflammatory diseases. Besides, the role of host genetics in determining disease susceptibility has been long known. Despite these shared effects, any relationship between host genetic variations and changes in microbiome composition in gut remains largely undiscovered. The modern day next generation sequencing (NGS) approaches along with genome wide association studies (GWAS) have revealed a number of genes that interact with gut bacteria and modulate physiological responses. However, elaborate details of the functional aspects of these genes and their association to microbiome components is still not available. This has created a barrier in understanding the etiology of complex gut associated diseases such as IBDs, various cancers and other pathogenic infestations. The observations in the present study provide a comprehensive overview on the critical role of *Atg16/1* in regulating inflammatory responses and creating conducive environment for the maintenance of healthy gut microbiome. The work provides a preliminary background for identifying functional groups of gut bacteria that get affected by *Atg16/1* and their role in mitigating inflammatory responses against a variety of internal (host-specific) and environmental stress. The compositional alterations in microbiome in *Atg16/1* deficient mice and resulting changes in metabolites of crucial pathways during antibiotic treatment and pregnancy phases demonstrated the critical role of this gene in maintaining microbial homeostasis inside gut. It is apparent that the role of *Atg16/1* in other dysbiotic events such as alcohol administration, pathogen invasion and alterations in food habits must be explored. Further, the effect of its disease variants/mutants (reported in IBD) on intestinal microbiota may be studied in mice models to establish their specific or general effect on the microbiome

composition. As a whole, this would immensely increase the information on genetic variants of *Atg16/1* and their long term effect on gut microbiota that leads to development of disease.

7 Summary (English)

Human gut hosts a large microbial community or microbiome that plays a critical role in immune maturation, endocrine functions, energy biogenesis and synthesis of vitamins besides numerous other functions. These functional characteristics of the microbiome are closely coordinated by number of host associated genes and gene products. Thus, along with host genetic factors, several other hosts associated exogenous factors such as food habits; alcoholism and antibiotic usage influence the functioning of microbiome. The advent of high throughput sequencing has highlighted the link between alterations in microbiome composition and development of several complex disorders such as inflammatory bowel diseases (IBD), diabetes, atherosclerosis and several cancers. Although several reports have linked dysbiosis of gut bacteria with the etiology of diseases, their actual roles and association with risk genes are still debated. Hence, longitudinal analysis of the microbiome composition under disease mimicking and genetically susceptible conditions could help decipher the underlying mechanisms. This would also help in understanding the functional activities of the microbial component that ultimately affects host physiology and inflammatory response.

The first part of the work was focused on studying the changes in microbial compositions during appendicitis associated inflammation. In the following parts the genetic susceptibility associated with functional defect in Paneth cells during different conditions was studied. Paneth cells present in the crypts of intestine are associated with the maintenance of microbial homeostasis. It has been shown that mutation in *Atg16l1* gene produces functional defect in Paneth cells and is a cause of initial inflammatory response during catastrophic events. Thus, in the later parts of this work, mice models with conditional knockout of the *Atg16l1* gene in their intestinal epithelium (*Atg16l1^{ΔIEC}*) were studied during different host-stress conditions. To monitor the microbiota changes, 16S rRNA profiling and metagenomics based approach was undertaken in the present study. The study design included: 1) analysis of human inflamed and non-inflamed appendicitis samples for identifying variations in the intestinal microbiota compositions; 2) Microbiota variations during antibiotic mediated dysbiotic event in control and *Atg16l1* intestinal knock down (*Atg16l1^{ΔIEC}*) mice models; 3) Microbial and functional variations during pregnancy in control and *Atg16l1^{ΔIEC}* mice models.

The analysis of inflamed and non-inflamed appendices showed compositional differences in bacterial flora in both samples with an overall decrease in bacterial abundance and diversity in the inflamed appendices. An increase in *Fusobacteria spp.* in the inflamed appendicitis along with a significant decrease in *Actinobacteria* and *Lentisphaerae* indicated their role in pathophysiology of appendicitis. In the next part of the work, it was found that *Atg16/1*^{ΔIEC} mice showed slower intestinal bacterial recovery after antibiotic administration. Besides, two mutually exclusive, genotype independent microbial community types were identified during the recovery state i.e. the *Bacteroidetes*-dominated type (BT), resembling the community structure before antibiotics and a novel *Firmicutes*-dominated type (FT). The work demonstrated that *Atg16/1* plays a subtle, yet distinct role in determining recovery of intestinal microbiota after a dysbiotic event. In the final part of the work, a comprehensive, longitudinal study of intestinal microbial compositions during the phases of pregnancy and 3 months post-parturition is presented. The work demonstrated that *Atg16/1* gene is crucial for determining correct microbial composition and may help in maintaining healthy gut microbiome during pregnancy. This in a way may affect the host associated metabolic pathways and immune response during pregnancy. A combined analysis of 16S rRNA profiling and metagenomics showed that absence of *Atg16/1* directly influences the development of bacterial enterotypes during pregnancy. The absence of *Atg16/1* leads to an alteration in microbial composition which is characterized by a decrease in *Lactobacillus* and an increase in diverse anaerobes such as *Bacteroides*, *Clostridiales* and *Lachnospiracea*. The metagenomics data indicated that although no global alterations in the metabolic pathway were noted, there were significant variations in the expression of crucial metabolites associated with carbohydrates and fatty acid metabolism. This suggested that the adjustment of normal energy metabolism to conserve essential nutrients required by the fetus was compromised in the *Atg16/1* deficient mice. Thus, the results show that *Atg16/1* is linked to healthy pregnancy through the regulation of protective microbiota and its genetic mutations may have important implications in preterm births or termination of pregnancy. Taken together, the presented results give a comprehensive outlook of the role of microbiota variations in normal and diseased states. The work highlights gene-microbiome crosstalk with emphasis on *Atg16/1* and its role in maintaining correct and healthy microbiota that functionally regulate metabolic pathways and inflammatory responses in the body.

8 Zusammenfassung (Deutsch)

Der menschliche Darm ist von einer großen Anzahl an Mikroorganismen, auch intestinale Mikrobiota genannt, besiedelt. Diese spielen unter anderem eine kritische Rolle in der Reifung des Immunsystems, bei endokrinen Vorgängen, der Biogenese von Energie und der Synthese von Vitaminen. Diese funktionellen Eigenschaften des Mikrobioms sind eng mit einigen wirtsassoziierten Genen und Genprodukten assoziiert. Neben genetischer Variation des Wirtes beeinflussen andere wirtsassoziierte Faktoren, wie Essgewohnheiten, Hygiene oder die Einnahme von Antibiotika die Zusammensetzung der intestinalen Bakteriengemeinschaften. Das Aufkommen von Hochdurchsatzsequenzierung vereinfacht die Analyse des komplexen Zusammenspiels zwischen Mikrobiota und Krankheitsbildern wie chronisch entzündliche Darmerkrankungen (CED), Diabetes, Arteriosklerose und einigen Krebsarten. Obwohl es verschiedene Studien über das Wechselspiel zwischen Dysbiose und Ätiologie gibt, ist der genaue Zusammenhang bisher nicht geklärt. Infolgedessen können Längsschnittstudien über die Mikrobiotazusammensetzung in Krankheitsmodellen und genetischen prädisponierten Risikovarianten dabei helfen, die zugrundeliegenden Mechanismen zu entschlüsseln. Neue Einblicke in die funktionellen Eigenschaften der mikrobiellen Komponente, die die Physiologie und die Entzündungsantwort des Wirtes beeinflussen, sind zu erwarten. Der erste Teil dieser Arbeit konzentriert sich auf die Veränderungen der Mikrobiotakomposition während einer Blinddarmentzündung. In weiteren Teilen der Arbeit wurde eine mit einem funktionellen Defekt in Paneth-Zellen verknüpfte genetische Prädisposition unter verschiedenen Bedingungen untersucht. Die in den Krypten des Darms gelegenen Paneth-Zellen werden mit der Aufrechterhaltung des mikrobiellen Gleichgewichts in Verbindung gebracht. In der Vergangenheit wurde bereits gezeigt, dass eine Mutation des *Atg16l1*-Gens einen funktionellen Defekt in Paneth-Zellen hervorruft und damit ursächlich für die initiale Entzündungsreaktion ist. Daher wurden in späteren Teilen dieser Arbeit Mausmodelle mit einem konditionalen Knockout des *Atg16l1*-Gens im intestinalen Epithel (*Atg16l1^{ΔIEC}*) in verschiedenen Stresssituationen des Wirtes untersucht. Um die Veränderungen der Mikrobiota zu beobachten, wurde 16S rRNA Profiling und ein Metagenomik-basiertes Vorgehen in dieser Studie angewandt. Das Studiendesign enthält: 1) Eine Analyse humaner entzündeter und nicht-entzündeter Appendizitis-Proben für die Identifizierung von Veränderungen der intestinalen Mikrobiotakomposition; 2) Mikrobiotaveränderungen während einer durch Antibiotika vermittelten Dysbiose in

Kontroll- und *Atg16/1*^{ΔIEC}-Mäusen; 3) Mikrobielle und funktionelle Veränderungen während und nach einer Schwangerschaft in Kontroll- und *Atg16/1*^{ΔIEC}-Mäusen.

Die Analyse entzündeter und nicht-entzündeter Appendizitisproben zeigte Veränderungen in der Zusammensetzung der Bakterienflora in beiden Gruppen mit einer Abnahme der bakteriellen Abundanz und Diversität bei Entzündung. Ein Anstieg der *Fusobacteria* spp. in der entzündeten Gruppe weist, zusammen mit einer signifikanten Reduktion der *Actinobacteria* und *Lentisphaerae*, auf eine Rolle dieser Bakterienspezies in der Pathophysiologie der Appendizitis hin. Im nächsten Teil der Arbeit wurde eine verlangsamte intestinale Erholung nach Antibiotikagabe in *Atg16/1*^{ΔIEC}-Mäusen gezeigt. Es wurden außerdem in der Erholungsphase zwei sich gegenseitig ausschließende, Genotyp-unabhängige mikrobielle Gemeinschaftstypen identifiziert, z.B der *Bacteroidetes*-dominierte Typ (BT), ähnlich der Mikrobiotazusammensetzung vor der antibiotischen Intervention, und ein neuer *Firmicutes*-dominierter Typ (FT). Die Arbeit zeigt auf, dass *Atg16/1* in der prädisponierten Zusammensetzung der intestinalen Mikrobiota nach einem dysbiotischen Vorfall eine entscheidende Rolle spielt. Im letzten Teil dieser Arbeit wird eine umfassende Längsschnittstudie der intestinalen Mikrobiotazusammensetzung während der Trächtigkeit und drei Monate post partum gezeigt. Die Arbeit demonstriert den Einfluss des *Atg16/1*-Gens auf die mikrobielle Zusammensetzung und Aufrechterhaltung einer gesunden Darmflora während der Trächtigkeit. *Atg16/1* könnte zudem über das Mikrobiom wirtsassoziierte Stoffwechselforgänge und Immunantworten während einer Trächtigkeit beeinflussen. Eine kombinierte Analyse aus 16S rRNA Profiling und einem Metagenomik-basierten Vorgehen zeigte, dass bei Abwesenheit von *Atg16/1* die Entwicklung des bakteriellen Enterotyps während der Trächtigkeit beeinflusst wird. Die Abwesenheit von *Atg16/1* führt zu Veränderungen der Mikrobiotakomposition, die durch eine Reduktion von *Lactobacillus* und einem Anstieg von diversen Anaerobiern, wie *Bacteroides*, *Clostridiales* und *Lachnospiracea*, gekennzeichnet ist. Obwohl keine umfassenden Veränderungen in den Stoffwechselfsignalwegen gefunden werden konnten, deuten die Metagenomikdaten auf eine signifikante Veränderung der Expression von Schlüsselenzymen des Kohlenhydrat- und Fettsäurestoffwechsels hin. In *Atg16/1* defizienten Mäusen scheint daher die Umstellung des normalen Energiestoffwechsels auf eine für den Fötus wichtige, Nährstoffe einsparende Stoffwechselart beeinträchtigt zu sein. Die Ergebnisse zeigen eine Verbindung zwischen *Atg16/1* und einer physiologischen Schwangerschaft durch die Regulation protektiver

Mikrobiota auf. Diese genetische Mutation könnte daher eine Rolle bei Frühgeburten und Schwangerschaftsabbrüchen spielen. Zusammenfassend geben die hier aufgezeigten Ergebnisse einen umfassenden Ausblick auf die Rolle der Mikrobiotavariationen in physiologischen und pathologischen Zuständen. Die Arbeit unterstreicht das Wechselspiel von genetischer Variation und intestinalem Mikrobiotabesatz und zeigt eine wichtige Rolle von *Atg16l1* fuer die Erhaltung der intestinalen Homoeostase.

9 References

- Aden, K., Rehman, A., Falk-Paulsen, M., Secher, T., Kuiper, J., Tran, F., Pfeuffer, S., Sheibani-Tezerji, R., Breuer, A., Luzius, A., *et al.* (2016). Epithelial IL-23R Signaling Licenses Protective IL-22 Responses in Intestinal Inflammation. *Cell reports*.
- Adolph, T.E., Tomczak, M.F., Niederreiter, L., Ko, H.J., Bock, J., Martinez-Naves, E., Glickman, J.N., Tschurtschenthaler, M., Hartwig, J., Hosomi, S., *et al.* (2013). Paneth cells as a site of origin for intestinal inflammation. *Nature* *503*, 272-276.
- Alekshun, M.N., and Levy, S.B. (2007). Molecular mechanisms of antibacterial multidrug resistance. *Cell* *128*, 1037-1050.
- Alshahrani, M.M., Yang, E., Yip, J., Ghanem, S.S., Abdallah, S.L., deAngelis, A.M., O'Malley, C.J., Moheimani, F., Najjar, S.M., and Jackson, D.E. (2014). CEACAM2 negatively regulates hemi (ITAM-bearing) GPVI and CLEC-2 pathways and thrombus growth in vitro and in vivo. *Blood* *124*, 2431-2441.
- Arijs, I., De Hertogh, G., Lemaire, K., Quintens, R., Van Lommel, L., Van Steen, K., Leemans, P., Cleynen, I., Van Assche, G., Vermeire, S., *et al.* (2009). Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. *PloS one* *4*, e7984.
- Arlt, A., Bharti, R., Ilves, I., Hasler, R., Miettinen, P., Paajanen, H., Brunke, G., Ellrichmann, M., Rehman, A., Hauser, C., *et al.* (2015). Characteristic changes in microbial community composition and expression of innate immune genes in acute appendicitis. *Innate Immun* *21*, 30-41.
- Arrieta, M.C., Bistritz, L., and Meddings, J.B. (2006). Alterations in intestinal permeability. *Gut* *55*, 1512-1520.
- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., *et al.* (2011). Enterotypes of the human gut microbiome. *Nature* *473*, 174-180.
- Azad, M.B., Bridgman, S.L., Becker, A.B., and Kozyrskyj, A.L. (2014). Infant antibiotic exposure and the development of childhood overweight and central adiposity. *International journal of obesity* *38*, 1290-1298.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., *et al.* (2008). The RAST server: Rapid annotations using subsystems technology. *Bmc Genomics* *9*.
- Bandzar, S., Gupta, S., and Platt, M.O. (2013). Crohn's disease: A review of treatment options and current research. *Cellular Immunology* *286*, 45-52.
- Bass, J.A., Goldman, J., Jackson, M.A., Gasior, A.C., Sharp, S.W., Drews, A.A., Saunders, C.J., and St Peter, S.D. (2012). Pediatric Crohn disease presenting as appendicitis: differentiating features from typical appendicitis. *Eur J Pediatr Surg* *22*, 274-278.
- Baumgart, D.C., and Dignass, A.U. (2002). Intestinal barrier function. *Current Opinion in Clinical Nutrition and Metabolic Care* *5*, 685-694.
- Baumgart, D.C., and Sandborn, W.J. (2012). Crohn's disease. *Lancet* *380*, 1590-1605.
- Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. *Cell* *157*, 121-141.

References

- Berglund, E.C., Kiialainen, A., and Syvanen, A.C. (2011). Next-generation sequencing technologies and applications for human genetic history and forensics. *Investig Genet* 2, 23.
- Bervoets, L., Van Hoorenbeeck, K., Kortleven, I., Van Noten, C., Hens, N., Vael, C., Goossens, H., Desager, K.N., and Vankerckhoven, V. (2013). Differences in gut microbiota composition between obese and lean children: a cross-sectional study. *Gut Pathog* 5, 10.
- Bevins, C.L., and Salzman, N.H. (2011). Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat Rev Microbiol* 9, 356-368.
- Bhangu, A., Soreide, K., Di Saverio, S., Assarsson, J.H., and Drake, F.T. (2015). Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *Lancet* 386, 1278-1287.
- Biagi, E., Candela, M., Turroni, S., Garagnani, P., Franceschi, C., and Brigidi, P. (2013). Ageing and gut microbes: perspectives for health maintenance and longevity. *Pharmacol Res* 69, 11-20.
- Biancheri, P., Powell, N., Monteleone, G., Lord, G., and MacDonald, T.T. (2013). The challenges of stratifying patients for trials in inflammatory bowel disease. *Trends in Immunology* 34, 564-571.
- Biddle, A., Stewart, L., Blanchard, J., and Leschine, S. (2013). Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. *Diversity* 5, 627-640.
- Biedermann, L., and Rogler, G. (2015). The intestinal microbiota: its role in health and disease. *European journal of pediatrics* 174, 151-167.
- Blencowe, H., Cousens, S., Chou, D., Oestergaard, M., Say, L., Moller, A.B., Kinney, M., Lawn, J., and Born Too Soon Preterm Birth Action, G. (2013). Born too soon: the global epidemiology of 15 million preterm births. *Reprod Health* 10 Suppl 1, S2.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
- Bollinger, R.R., Barbas, A.S., Bush, E.L., Lin, S.S., and Parker, W. (2007). Biofilms in the normal human large bowel: fact rather than fiction. *Gut* 56, 1481-1482.
- Boon, E., Meehan, C.J., Whidden, C., Wong, D.H.J., Langille, M.G.I., and Beiko, R.G. (2014). Interactions in the microbiome: communities of organisms and communities of genes. *Fems Microbiology Reviews* 38, 90-118.
- Booth, C., and Potten, C.S. (2000). Gut instincts: thoughts on intestinal epithelial stem cells. *Journal of Clinical Investigation* 105, 1493-1499.
- Bourriaud, C., Robins, R.J., Martin, L., Kozlowski, F., Tenailleau, E., Cherbut, C., and Michel, C. (2005). Lactate is mainly fermented to butyrate by human intestinal microfloras but inter-individual variation is evident. *J Appl Microbiol* 99, 201-212.
- Boursi, B., Mamtani, R., Haynes, K., and Yang, Y.X. (2015). The effect of past antibiotic exposure on diabetes risk. *European journal of endocrinology* 172, 639-648.
- Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S., and Lynn, D.J. (2013). InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic acids research* 41, D1228-1233.
- Bry, L., Falk, P.G., Midtvedt, T., and Gordon, J.I. (1996). A model of host-microbial interactions in an open mammalian ecosystem. *Science* 273, 1380-1383.

References

- Buffie, C.G., Jarchum, I., Equinda, M., Lipuma, L., Gobourne, A., Viale, A., Ubeda, C., Xavier, J., and Pamer, E.G. (2012). Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infection and immunity* *80*, 62-73.
- Cadwell, K., Liu, J.Y., Brown, S.L., Miyoshi, H., Loh, J., Lennerz, J.K., Kishi, C., Kc, W., Carrero, J.A., Hunt, S., *et al.* (2008). A key role for autophagy and the autophagy gene *Atg16l1* in mouse and human intestinal Paneth cells. *Nature* *456*, 259-263.
- Cadwell, K., Patel, K.K., Komatsu, M., Virgin, H.W.t., and Stappenbeck, T.S. (2009). A common role for *Atg16L1*, *Atg5* and *Atg7* in small intestinal Paneth cells and Crohn disease. *Autophagy* *5*, 250-252.
- Cani, P.D., and Delzenne, N.M. (2007). Gut microflora as a target for energy and metabolic homeostasis. *Current Opinion in Clinical Nutrition and Metabolic Care* *10*, 729-734.
- Cao, Y., Zheng, X., Li, F., and Bo, X. (2015). mmnet: An R Package for Metagenomics Systems Biology Analysis. *BioMed research international* *2015*, 167249.
- Caporaso, J.G., Lauber, C.L., Costello, E.K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., Gajer, P., Ravel, J., Fierer, N., *et al.* (2011). Moving pictures of the human microbiome. *Genome biology* *12*, R50.
- Cesta, M.F. (2006). Normal structure, function, and histology of mucosa-associated lymphoid tissue. *Toxicol Pathol* *34*, 599-608.
- Chan, C.K.K., Hsu, A.L., Halgamuge, S.K., and Tang, S.L. (2008). Binning sequences using very sparse labels within a metagenome. *Bmc Bioinformatics* *9*.
- Chen, P., Torralba, M., Tan, J., Embree, M., Zengler, K., Starkel, P., van Pijkeren, J.P., DePew, J., Loomba, R., Ho, S.B., *et al.* (2015). Supplementation of saturated long-chain fatty acids maintains intestinal eubiosis and reduces ethanol-induced liver injury in mice. *Gastroenterology* *148*, 203-214 e216.
- Chevreur, B., Pfisterer, T., Drescher, B., Driesel, A.J., Muller, W.E.G., Wetter, T., and Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* *14*, 1147-1159.
- Clarridge, J.E. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews* *17*, 840-+.
- Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell* *148*, 1258-1270.
- Clevers, H.C., and Bevins, C.L. (2013). Paneth cells: maestros of the small intestinal crypts. *Annu Rev Physiol* *75*, 289-311.
- Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., Chandra, S., McGarrell, D.M., Schmidt, T.M., Garrity, G.M., *et al.* (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic acids research* *31*, 442-443.
- Collado, M.C., Isolauri, E., Laitinen, K., and Salminen, S. (2008). Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women. *Am J Clin Nutr* *88*, 894-899.

- Comalada, M., Bailon, E., de Haro, O., Lara-Villoslada, F., Xaus, J., Zarzuelo, A., and Galvez, J. (2006). The effects of short-chain fatty acids on colon epithelial proliferation and survival depend on the cellular phenotype. *J Cancer Res Clin Oncol* *132*, 487-497.
- Cooney, R., Baker, J., Brain, O., Danis, B., Pichulik, T., Allan, P., Ferguson, D.J., Campbell, B.J., Jewell, D., and Simmons, A. (2010). NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nat Med* *16*, 90-97.
- Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science* *326*, 1694-1697.
- Dalal, S.R., and Chang, E.B. (2014). The microbial basis of inflammatory bowel diseases. *Journal of Clinical Investigation* *124*, 4190-4196.
- Dalmasso, G., Nguyen, H.T., Ingersoll, S.A., Ayyadurai, S., Laroui, H., Charania, M.A., Yan, Y., Sitaraman, S.V., and Merlin, D. (2011). The PepT1-NOD2 signaling pathway aggravates induced colitis in mice. *Gastroenterology* *141*, 1334-1345.
- Danese, S., and Fiocchi, C. (2011). Medical Progress Ulcerative Colitis. *New England Journal of Medicine* *365*, 1713-1725.
- David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., *et al.* (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* *505*, 559-563.
- David R. Flum, M.D., M.P.H. (2015). Acute appendicitis--appendectomy or the "antibiotics first" strategy. *N Engl J Med* *372*, 2274.
- De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poulet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 14691-14696.
- Derrien, M., van Passel, M.W., van de Bovenkamp, J.H., Schipper, R.G., de Vos, W.M., and Dekker, J. (2010). Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut microbes* *1*, 254-268.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* *72*, 5069-5072.
- Diaz, N.N., Krause, L., Goesmann, A., Niehaus, K., and Nattkemper, T.W. (2009). TACO - Taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. *Bmc Bioinformatics* *10*.
- DiGiulio, D.B., Callahan, B.J., McMurdie, P.J., Costello, E.K., Lyell, D.J., Robaczewska, A., Sun, C.L., Goltsman, D.S., Wong, R.J., Shaw, G., *et al.* (2015). Temporal and spatial variation of the human microbiota during pregnancy. *Proceedings of the National Academy of Sciences of the United States of America* *112*, 11060-11065.
- Donohoe, D.R., Garge, N., Zhang, X., Sun, W., O'Connell, T.M., Bunger, M.K., and Bultman, S.J. (2011). The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell metabolism* *13*, 517-526.
- Dudgeon, L.S. (1926). A Study of the Intestinal Flora under normal and abnormal Conditions. *The Journal of hygiene* *25*, 119-141.

- Dufrene M, L.P. (1997). SPECIES ASSEMBLAGES AND INDICATOR SPECIES:THE NEED FOR A FLEXIBLE ASYMMETRICAL APPROACH. *ECOLOGICAL SOCIETY OF AMERICA Volume 67*.
- Duncan, S.H., Louis, P., and Flint, H.J. (2004). Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Applied and environmental microbiology* 70, 5810-5817.
- Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635-1638.
- Elinav, E., Strowig, T., Kau, A.L., Henao-Mejia, J., Thaiss, C.A., Booth, C.J., Peaper, D.R., Bertin, J., Eisenbarth, S.C., Gordon, J.I., *et al.* (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* 145, 745-757.
- Fasano, A., and Shea-Donohue, T. (2005). Mechanisms of disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. *Nature Clinical Practice Gastroenterology & Hepatology* 2, 416-422.
- Fierro, J.L., Prasad, P.A., Localio, A.R., Grundmeier, R.W., Wasserman, R.C., Zaoutis, T.E., and Gerber, J.S. (2014). Variability in the diagnosis and treatment of group a streptococcal pharyngitis by primary care pediatricians. *Infection control and hospital epidemiology* 35 *Suppl 3*, S79-85.
- Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., *et al.* (2010). The Pfam protein families database. *Nucleic acids research* 38, D211-D222.
- Frank, D.N., Robertson, C.E., Hamm, C.M., Kpadeh, Z., Zhang, T., Chen, H., Zhu, W., Sartor, R.B., Boedeker, E.C., Harpaz, N., *et al.* (2011). Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* 17, 179-184.
- Frank, D.N., St Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences of the United States of America* 104, 13780-13785.
- Friedman, J., and Alm, E.J. (2012). Inferring correlation networks from genomic survey data. *PLoS computational biology* 8, e1002687.
- Gagliardi, L., Rusconi, F., Da Fre, M., Mello, G., Carnielli, V., Di Lallo, D., Macagno, F., Miniaci, S., Corchia, C., and Cuttini, M. (2013). Pregnancy disorders leading to very preterm birth influence neonatal outcomes: results of the population-based ACTION cohort study. *Pediatr Res* 73, 794-801.
- Galan, J.E., and Bliska, J.B. (1996). Cross-talk between bacterial pathogens and their host cells. *Annual Review of Cell and Developmental Biology* 12, 221-255.
- Gao, Y.D., Zhao, Y., and Huang, J. (2014). Metabolic modeling of common *Escherichia coli* strains in human gut microbiome. *BioMed research international* 2014, 694967.
- Gibson, M.K., Crofts, T.S., and Dantas, G. (2015). Antibiotics and the developing infant gut microbiota and resistome. *Curr Opin Microbiol* 27, 51-56.
- Gibson, P.R. (2004). Increased gut permeability in Crohn's disease: is TNF the link? *Gut* 53, 1724-1725.

References

- Glubb, D.M., Geary, R.B., Barclay, M.L., Roberts, R.L., Pearson, J., Keenan, J.I., McKenzie, J., and Bentley, R.W. (2011). NOD2 and ATG16L1 polymorphisms affect monocyte responses in Crohn's disease. *World J Gastroenterol* *17*, 2829-2837.
- Goenawan, I.H., Bryan, K., and Lynn, D.J. (2016). DyNet: visualization and analysis of dynamic molecular interaction networks. *Bioinformatics* *32*, 2713-2715.
- Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W.F., and Veldhuyzen van Zanten, S.J. (2006). Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *Journal of clinical microbiology* *44*, 4136-4141.
- Groschwitz, K.R., and Hogan, S.P. (2009). Intestinal barrier function: Molecular regulation and disease pathogenesis. *Journal of Allergy and Clinical Immunology* *124*, 3-20.
- Guinane, C.M., Tadrous, A., Fouhy, F., Ryan, C.A., Dempsey, E.M., Murphy, B., Andrews, E., Cotter, P.D., Stanton, C., and Ross, R.P. (2013). Microbial composition of human appendices from patients following appendectomy. *MBio* *4*.
- Guo, M., Huang, K., Chen, S., Qi, X., He, X., Cheng, W.H., Luo, Y., Xia, K., and Xu, W. (2014). Combination of metagenomics and culture-based methods to study the interaction between ochratoxin a and gut microbiota. *Toxicol Sci* *141*, 314-323.
- Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Albrecht, M., Mayr, G., De La Vega, F.M., Briggs, J., *et al.* (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genetics* *39*, 207-211.
- Haque, M.M., Ghosh, T.S., Komanduri, D., and Mande, S.S. (2009). SOrt-ITEMS: Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics* *25*, 1722-1730.
- Hashimoto, T., Perlot, T., Rehman, A., Trichereau, J., Ishiguro, H., Paolino, M., Sigl, V., Hanada, T., Hanada, R., Lipinski, S., *et al.* (2012). ACE2 links amino acid malnutrition to microbial ecology and intestinal inflammation. *Nature* *487*, 477-481.
- Hawrelak, J.A., and Myers, S.P. (2004). The causes of intestinal dysbiosis: a review. *Altern Med Rev* *9*, 180-197.
- Heinsen, F.A., Knecht, H., Neulinger, S.C., Schmitz, R.A., Knecht, C., Kuhbacher, T., Rosenstiel, P.C., Schreiber, S., Friedrichs, A.K., and Ott, S.J. (2015). Dynamic changes of the luminal and mucosa-associated gut microbiota during and after antibiotic therapy with paromomycin. *Gut microbes* *6*, 243-254.
- Henao-Mejia, J., Elinav, E., Jin, C., Hao, L., Mehal, W.Z., Strowig, T., Thaiss, C.A., Kau, A.L., Eisenbarth, S.C., Jurczak, M.J., *et al.* (2012). Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* *482*, 179-185.
- Henderson, P., and Stevens, C. (2012). The role of autophagy in Crohn's disease. *Cells* *1*, 492-519.
- Hoff, K.J., Lingner, T., Meinicke, P., and Tech, M. (2009). Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic acids research* *37*, W101-W105.
- Holmes, E., Li, J.V., Athanasiou, T., Ashrafian, H., and Nicholson, J.K. (2011). Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol* *19*, 349-359.
- Holmes, I., Harris, K., and Quince, C. (2012). Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS one* *7*, e30126.

- Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* *486*, 207-214.
- Hunt, R.H., Camilleri, M., Crowe, S.E., El-Omar, E.M., Fox, J.G., Kuipers, E.J., Malfertheiner, P., McColl, K.E., Pritchard, D.M., Rugge, M., *et al.* (2015). The stomach in health and disease. *Gut* *64*, 1650-1668.
- Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. *Genome Research* *17*, 377-386.
- Ivanov, II, Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V., *et al.* (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* *139*, 485-498.
- Ivanov, II, and Littman, D.R. (2010). Segmented filamentous bacteria take the stage. *Mucosal Immunol* *3*, 209-212.
- Janda, J.M., and Abbott, S.L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of clinical microbiology* *45*, 2761-2764.
- Jasarevic, E., Howerton, C.L., Howard, C.D., and Bale, T.L. (2015). Alterations in the Vaginal Microbiome by Maternal Stress Are Associated With Metabolic Reprogramming of the Offspring Gut and Brain. *Endocrinology* *156*, 3265-3276.
- Jernberg, C., Lofmark, S., Edlund, C., and Jansson, J.K. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *The ISME journal* *1*, 56-66.
- Jimenez, E., Marin, M.L., Martin, R., Odriozola, J.M., Olivares, M., Xaus, J., Fernandez, L., and Rodriguez, J.M. (2008). Is meconium from healthy newborns actually sterile? *Res Microbiol* *159*, 187-193.
- Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., Vandamme, P., and Vermeire, S. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* *60*, 631-637.
- Kamada, N., Seo, S.U., Chen, G.Y., and Nunez, G. (2013). Role of the gut microbiota in immunity and inflammatory disease. *Nature Reviews Immunology* *13*, 321-335.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* *28*, 27-30.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic acids research* *32*, D277-D280.
- Kararli, T.T. (1995). Comparison of the gastrointestinal anatomy, physiology, and biochemistry of humans and commonly used laboratory animals. *Biopharm Drug Dispos* *16*, 351-380.
- Kaser, A., Zeissig, S., and Blumberg, R.S. (2010). Inflammatory Bowel Disease. *Annual Review of Immunology*, Vol 28 *28*, 573-621.
- Kelly, C.P., Okeane, J.C., Orellana, J., Schroy, P.C., Yang, S., Lamont, J.T., and Brady, H.R. (1992). Human Colon Cancer-Cells Express Icam-1 In vivo and Support Lfa-1-Dependent Lymphocyte Adhesion In vitro. *American Journal of Physiology* *263*, G864-G870.
- Kelly, J.R., Kennedy, P.J., Cryan, J.F., Dinan, T.G., Clarke, G., and Hyland, N.P. (2015). Breaking down the barriers: the gut microbiome, intestinal permeability and stress-related psychiatric disorders. *Front Cell Neurosci* *9*, 392.

References

- Kenny, E.E., Pe'er, I., Karban, A., Ozelius, L., Mitchell, A.A., Ng, S.M., Erazo, M., Ostrer, H., Abraham, C., Abreu, M.T., *et al.* (2012). A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS Genet* 8, e1002559.
- Kim, S.C., Tonkonogy, S.L., Albright, C.A., Tsang, J., Balish, E.J., Braun, J., Huycke, M.M., and Sartor, R.B. (2005). Variable phenotypes of enterocolitis in interleukin 10-deficient mice monoassociated with two different commensal bacteria. *Gastroenterology* 128, 891-906.
- King, J.C. (2000). Physiology of pregnancy and nutrient metabolism. *Am J Clin Nutr* 71, 1218S-1225S.
- Komuro, T. (2006). Structure and organization of interstitial cells of Cajal in the gastrointestinal tract. *J Physiol* 576, 653-658.
- Korecka, A., and Arulampalam, V. (2012). The gut microbiome: scourge, sentinel or spectator? *J Oral Microbiol* 4.
- Koren, O., Goodrich, J.K., Cullender, T.C., Spor, A., Laitinen, K., Backhed, H.K., Gonzalez, A., Werner, J.J., Angenent, L.T., Knight, R., *et al.* (2012). Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* 150, 470-480.
- Kostic, A.D., Xavier, R.J., and Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146, 1489-1499.
- Krause, L., Diaz, N.N., Goesmann, A., Kelley, S., Nattkemper, T.W., Rohwer, F., Edwards, R.A., and Stoye, J. (2008). Phylogenetic classification of short environmental DNA fragments. *Nucleic acids research* 36, 2230-2239.
- Kronman, M.P., Zaoutis, T.E., Haynes, K., Feng, R., and Coffin, S.E. (2012). Antibiotic exposure and IBD development among children: a population-based cohort study. *Pediatrics* 130, e794-803.
- Kuo, S.M. (2013). The interplay between fiber and the intestinal microbiome in the inflammatory response. *Adv Nutr* 4, 16-28.
- Kuo, S.M., Merhige, P.M., and Hagey, L.R. (2013). The effect of dietary prebiotics and probiotics on body weight, large intestine indices, and fecal bile acid profile in wild type and IL10^{-/-} mice. *PLoS one* 8, e60270.
- Lamps, L.W. (2004). Appendicitis and infections of the appendix. *Semin Diagn Pathol* 21, 86-97.
- Lasken, R.S. (2009). Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochemical Society Transactions* 37, 450-453.
- Lassen, K.G., Kuballa, P., Conway, K.L., Patel, K.K., Becker, C.E., Peloquin, J.M., Villablanca, E.J., Norman, J.M., Liu, T.C., Heath, R.J., *et al.* (2014). Atg16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. *Proceedings of the National Academy of Sciences of the United States of America* 111, 7741-7746.
- Laurin, M., Everett, M.L., and Parker, W. (2011). The cecal appendix: one more immune component with a function disturbed by post-industrial culture. *Anat Rec (Hoboken)* 294, 567-579.
- Legendre P, L.L. (1998). Numerical ecology. ELSEVIER.
- Levine, B., and Deretic, V. (2007). Unveiling the roles of autophagy in innate and adaptive immunity. *Nat Rev Immunol* 7, 767-777.
- Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837-848.

- Li, R.Q., Li, Y.R., Kristiansen, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics* 24, 713-714.
- Liu, B., Gibbons, T., Ghodsi, M., Treangen, T., and Pop, M. (2011). Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *Bmc Genomics* 12.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012a). Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012, 251364.
- Liu, L., Li, Y.H., Li, S.L., Hu, N., He, Y.M., Pong, R., Lin, D.N., Lu, L.H., and Law, M. (2012b). Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*.
- Loos, M., Geens, M., Schauvliege, S., Gasthuys, F., van der Meulen, J., Dubreuil, J.D., Goddeeris, B.M., Niewold, T., and Cox, E. (2012). Role of heat-stable enterotoxins in the induction of early immune responses in piglets after infection with enterotoxigenic *Escherichia coli*. *PloS one* 7, e41041.
- Lopetuso, L.R., Scaldaferri, F., Petito, V., and Gasbarrini, A. (2013). Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog* 5, 23.
- Louis, P., Hold, G.L., and Flint, H.J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nature Reviews Microbiology* 12, 661-672.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15, 550.
- Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and environmental microbiology* 73, 1576-1585.
- Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220-230.
- Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijis, I., Eeckhaut, V., Ballet, V., Claes, K., Van Immerseel, F., Verbeke, K., *et al.* (2014). A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* 63, 1275-1283.
- Maranduba, C.M.D., De Castro, S.B.R., de Souza, G.T., Rossato, C., da Guia, F.C., Valente, M.A.S., Rettore, J.V.P., Maranduba, C.P., de Souza, C.M., do Carmo, A.M.R., *et al.* (2015). Intestinal Microbiota as Modulators of the Immune System and Neuroimmune System: Impact on the Host Health and Homeostasis. *Journal of Immunology Research*.
- Marchiando, A.M., Graham, W.V., and Turner, J.R. (2010). Epithelial Barriers in Homeostasis and Disease. *Annual Review of Pathology-Mechanisms of Disease* 5, 119-144.
- Mardinoglu, A., Shoaie, S., Bergentall, M., Ghaffari, P., Zhang, C., Larsson, E., Backhed, F., and Nielsen, J. (2015). The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Molecular systems biology* 11, 834.
- Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet* 24, 133-141.
- Mariat, D., Firmesse, O., Levenez, F., Guimaraes, V., Sokol, H., Dore, J., Corthier, G., and Furet, J.P. (2009). The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol* 9, 123.

- Markowitz, V.M., Mavromatis, K., Ivanova, N.N., Chen, I.M.A., Chu, K., and Kyripides, N.C. (2009). IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25, 2271-2278.
- Martinez, C., Antolin, M., Santos, J., Torrejon, A., Casellas, F., Borrueal, N., Guarner, F., and Malagelada, J.R. (2008). Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. *The American journal of gastroenterology* 103, 643-648.
- Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., and de La Cochetiere, M.F. (2013). Development of intestinal microbiota in infants and its impact on health. *Trends Microbiol* 21, 167-173.
- McHardy, A.C., Martin, H.G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007). Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods* 4, 63-72.
- Meehan, C.J., and Beiko, R.G. (2014). A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome biology and evolution* 6, 703-713.
- Merchant, R., Mower, W.R., Ourian, A., Abrahamian, F.M., Moran, G.J., Krishnadasan, A., and Talan, D.A. (2012). Association Between Appendectomy and Clostridium difficile Infection. *J Clin Med Res* 4, 17-19.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., *et al.* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *Bmc Bioinformatics* 9.
- Miettinen, P., Pasanen, P., Lahtinen, J., and Alhava, E. (1996). Acute abdominal pain in adults. *Ann Chir Gynaecol* 85, 5-9.
- Miller, J.R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics* 95, 315-327.
- Molodecky, N.A., Soon, I.S., Rabi, D.M., Ghali, W.A., Ferris, M., Chernoff, G., Benchimol, E.I., Panaccione, R., Ghosh, S., Barkema, H.W., *et al.* (2012). Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases With Time, Based on Systematic Review. *Gastroenterology* 142, 46-54.
- Moreno-Indias, I., Cardona, F., Tinahones, F.J., and Queipo-Ortuno, M.I. (2014a). Impact of the gut microbiota on the development of obesity and type 2 diabetes mellitus. *Frontiers in microbiology* 5.
- Moreno-Indias, I., Cardona, F., Tinahones, F.J., and Queipo-Ortuno, M.I. (2014b). Impact of the gut microbiota on the development of obesity and type 2 diabetes mellitus. *Front Microbiol* 5, 190.
- Muller, J., Szklarczyk, D., Julien, P., Letunic, I., Roth, A., Kuhn, M., Powell, S., von Mering, C., Doerks, T., Jensen, L.J., *et al.* (2010). eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic acids research* 38, D190-D195.
- Murthy, A., Li, Y., Peng, I., Reichelt, M., Katakam, A.K., Noubade, R., Roose-Girma, M., DeVoss, J., Diehl, L., Graham, R.R., *et al.* (2014). A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. *Nature* 506, 456-462.
- Neel, J.V., Major, E.O., Awa, A.A., Glover, T., Burgess, A., Traub, R., Curfman, B., and Satoh, C. (1996). Hypothesis: "Rogue cell"-type chromosomal damage in lymphocytes is associated with infection with the JC human polyoma virus and has implications for oncogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 93, 2690-2695.

References

- Neish, A.S. (2009). Microbes in gastrointestinal health and disease. *Gastroenterology* *136*, 65-80.
- Niederreiter, L., and Kaser, A. (2011). Endoplasmic reticulum stress and inflammatory bowel disease. *Acta Gastro-Enterologica Belgica* *74*, 330-333.
- Noguchi, H., Taniguchi, T., and Itoh, T. (2008). MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. *DNA Research* *15*, 387-396.
- Novacek, G., Weltermann, A., Sobala, A., Tilg, H., Petritsch, W., Reinisch, W., Mayer, A., Haas, T., Kaser, A., Feichtenschlager, T., *et al.* (2010). Inflammatory Bowel Disease Is a Risk Factor for Recurrent Venous Thromboembolism. *Gastroenterology* *139*, 779-U114.
- Oszkiel, H., Wilczak, J., and Jank, M. (2014). Biologically active substances-enriched diet regulates gonadotrope cell activation pathway in liver of adult and old rats. *Genes Nutr* *9*, 427.
- Ott, S.J., Plamondon, S., Hart, A., Begun, A., Rehman, A., Kamm, M.A., and Schreiber, S. (2008). Dynamics of the mucosa-associated flora in ulcerative colitis patients during remission and clinical relapse. *Journal of clinical microbiology* *46*, 3510-3513.
- Otte, J.M., Kiehne, K., and Herzig, K.H. (2003). Antimicrobial peptides in innate immunity of the human intestine. *J Gastroenterol* *38*, 717-726.
- Øyvind Hammer, D.A.T.H., and Paul D. Ryan (2001). PAST: PALEONTOLOGICAL STATISTICS SOFTWARE PACKAGE FOR EDUCATION AND DATA ANALYSIS. *Palaeontological Association vol. 4*.
- Perez-Cobas, A.E., Gosalbes, M.J., Friedrichs, A., Knecht, H., Artacho, A., Eismann, K., Otto, W., Rojo, D., Bargiela, R., von Bergen, M., *et al.* (2013). Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut* *62*, 1591-1601.
- Peterfi, Z., Kovacs, K., Antal, A., and Kocsis, B. (2006). Anti-lipopolysaccharide antibodies in acute appendicitis detected by ELISA. *APMIS* *114*, 265-269.
- Peterson, L.W., and Artis, D. (2014). Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat Rev Immunol* *14*, 141-153.
- Prince, A.L., Antony, K.M., Ma, J., and Aagaard, K.M. (2014). The microbiome and development: a mother's perspective. *Semin Reprod Med* *32*, 14-22.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J., and Glockner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research* *35*, 7188-7196.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* *464*, 59-65.
- Rath, H.C., Wilson, K.H., and Sartor, R.B. (1999). Differential induction of colitis and gastritis in HLA-B27 transgenic rats selectively colonized with *Bacteroides vulgatus* or *Escherichia coli*. *Infection and immunity* *67*, 2969-2974.
- Rausch, P., Rehman, A., Kunzel, S., Hasler, R., Ott, S.J., Schreiber, S., Rosenstiel, P., Franke, A., and Baines, J.F. (2011). Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proceedings of the National Academy of Sciences of the United States of America* *108*, 19030-19035.

- Rautava, S., Collado, M.C., Salminen, S., and Isolauri, E. (2012). Probiotics modulate host-microbe interaction in the placenta and fetal gut: a randomized, double-blind, placebo-controlled trial. *Neonatology* *102*, 178-184.
- Rehman, A., Rausch, P., Wang, J., Skieceviciene, J., Kiudelis, G., Bhagalia, K., Amarapurkar, D., Kupcinskis, L., Schreiber, S., Rosenstiel, P., *et al.* (2016). Geographical patterns of the standing and active human gut microbiome in health and IBD. *Gut* *65*, 238-248.
- Rehman, A., Sina, C., Gavriloiva, O., Hasler, R., Ott, S., Baines, J.F., Schreiber, S., and Rosenstiel, P. (2011). Nod2 is essential for temporal development of intestinal microbial communities. *Gut* *60*, 1354-1362.
- Rho, J.H., Wright, D.P., Christie, D.L., Clinch, K., Furneaux, R.H., and Robertson, A.M. (2005). A novel mechanism for desulfation of mucin: identification and cloning of a mucin-desulfating glycosidase (sulfoglycosidase) from *Prevotella* strain RS2. *J Bacteriol* *187*, 1543-1551.
- Rho, M.N., Tang, H.X., and Ye, Y.Z. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic acids research* *38*.
- Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004). Metagenomics: Genomic analysis of microbial communities. *Annual Review of Genetics* *38*, 525-552.
- Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., *et al.* (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* *39*, 596-604.
- Risnes, K.R., Belanger, K., Murk, W., and Bracken, M.B. (2011). Antibiotic exposure by 6 months and asthma and allergy at 6 years: Findings in a cohort of 1,401 US children. *American journal of epidemiology* *173*, 310-318.
- Rivera-Chavez, F.A., Peters-Hybki, D.L., Barber, R.C., Lindberg, G.M., Jialal, I., Munford, R.S., and O'Keefe, G.E. (2004). Innate immunity genes influence the severity of acute appendicitis. *Ann Surg* *240*, 269-277.
- Rivera-Chavez, F.A., Wheeler, H., Lindberg, G., Munford, R.S., and O'Keefe, G.E. (2003). Regional and systemic cytokine responses to acute inflammation of the vermiform appendix. *Ann Surg* *237*, 408-416.
- Roberts, J.P. (1988). Quantitative bacterial flora of acute appendicitis. *Arch Dis Child* *63*, 536-540.
- Robertson, G., Bilenky, M., Lin, K., He, A., Yuen, W., Dagpinar, M., Varhol, R., Teague, K., Griffith, O.L., Zhang, X., *et al.* (2006). cisRED: a database system for genome-scale computational discovery of regulatory elements. *Nucleic Acids Res* *34*, D68-73.
- Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* *281*, 363, 365.
- Rosenstiel, P., Till, A., and Schreiber, S. (2007). NOD-like receptors and human diseases. *Microbes Infect* *9*, 648-657.
- Round, J.L., Lee, S.M., Li, J., Tran, G., Jabri, B., Chatila, T.A., and Mazmanian, S.K. (2011). The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* *332*, 974-977.

References

- Sadaghian Sadabad, M., Regeling, A., de Goffau, M.C., Blokzijl, T., Weersma, R.K., Penders, J., Faber, K.N., Harmsen, H.J., and Dijkstra, G. (2015). The ATG16L1-T300A allele impairs clearance of pathosymbionts in the inflamed ileal mucosa of Crohn's disease patients. *Gut* *64*, 1546-1552.
- Saitoh, T., Fujita, N., Jang, M.H., Uematsu, S., Yang, B.G., Satoh, T., Omori, H., Noda, T., Yamamoto, N., Komatsu, M., *et al.* (2008). Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1beta production. *Nature* *456*, 264-268.
- Salem, M., Nielsen, O.H., Nys, K., Yazdanyar, S., and Seidelin, J.B. (2015). Impact of T300A Variant of ATG16L1 on Antibacterial Response, Risk of Culture Positive Infections, and Clinical Course of Crohn's Disease. *Clin Transl Gastroenterol* *6*, e122.
- Salminen, S., Bouley, C., Boutron-Ruault, M.C., Cummings, J.H., Franck, A., Gibson, G.R., Isolauri, E., Moreau, M.C., Roberfroid, M., and Rowland, I. (1998). Functional food science and gastrointestinal physiology and function. *Br J Nutr* *80 Suppl 1*, S147-171.
- Salzman, N.H., Hung, K., Haribhai, D., Chu, H., Karlsson-Sjoberg, J., Amir, E., Tegatz, P., Barman, M., Hayward, M., Eastwood, D., *et al.* (2010). Enteric defensins are essential regulators of intestinal microbial ecology. *Nat Immunol* *11*, 76-83.
- Sanger, F., and Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* *94*, 441-448.
- Sartor, R.B. (2006). Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nature Clinical Practice Gastroenterology & Hepatology* *3*, 390-407.
- Sartor, R.B. (2008). Microbial influences in inflammatory bowel diseases. *Gastroenterology* *134*, 577-594.
- Scanlan, P.D., Shanahan, F., O'Mahony, C., and Marchesi, J.R. (2006). Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *Journal of clinical microbiology* *44*, 3980-3988.
- Schloss, J.A. (2008). How to get genomes at one ten-thousandth the cost. *Nature Biotechnology* *26*, 1113-1115.
- Schloss, P.D., and Handelsman, J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and environmental microbiology* *71*, 1501-1506.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology* *75*, 7537-7541.
- Schmid, D., and Munz, C. (2007). Innate and adaptive immunity through autophagy. *Immunity* *27*, 11-21.
- Schneeman, B.O. (2002). Gastrointestinal physiology and functions. *British Journal of Nutrition* *88*, S159-S163.
- Schultz, M., Tonkonogy, S.L., Sellon, R.K., Veltkamp, C., Godfrey, V.L., Kwon, J., Grenther, W.B., Balish, E., Horak, I., and Sartor, R.B. (1999). IL-2-deficient mice raised under germfree conditions develop delayed mild focal intestinal inflammation. *Am J Physiol* *276*, G1461-1472.

- Schumann, A., Nutten, S., Donnicola, D., Comelli, E.M., Mansourian, R., Cherbut, C., Cortesy-Theulaz, I., and Garcia-Rodenas, C. (2005). Neonatal antibiotic treatment alters gastrointestinal tract developmental gene expression and intestinal barrier transcriptome. *Physiological genomics* 23, 235-245.
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome biology* 12, R60.
- Selengut, J.D., Haft, D.H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W.C., Richter, A.R., and White, O. (2007). TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic acids research* 35, D260-D264.
- Shaw, S.Y., Blanchard, J.F., and Bernstein, C.N. (2010). Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease. *The American journal of gastroenterology* 105, 2687-2692.
- Shimizu, T., Ishizuka, M., and Kubota, K. (2016). A lower neutrophil to lymphocyte ratio is closely associated with catarrhal appendicitis versus severe appendicitis. *Surg Today* 46, 84-89.
- Shreiner, A.B., Kao, J.Y., and Young, V.B. (2015). The gut microbiome in health and in disease. *Current opinion in gastroenterology* 31, 69-75.
- Slattery, M.L., Lundgreen, A., Herrick, J.S., and Wolff, R.K. (2011). Genetic variation in RPS6KA1, RPS6KA2, RPS6KB1, RPS6KB2, and PDK1 and risk of colon or rectal cancer. *Mutat Res* 706, 13-20.
- Smith, H.F., Fisher, R.E., Everett, M.L., Thomas, A.D., Bollinger, R.R., and Parker, W. (2009). Comparative anatomy and phylogenetic distribution of the mammalian cecal appendix. *J Evol Biol* 22, 1984-1999.
- Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L.G., Gratadoux, J.J., Blugeon, S., Bridonneau, C., Furet, J.P., Corthier, G., *et al.* (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences of the United States of America* 105, 16731-16736.
- Sorbara, M.T., Ellison, L.K., Ramjeet, M., Travassos, L.H., Jones, N.L., Girardin, S.E., and Philpott, D.J. (2013). The protein ATG16L1 suppresses inflammatory cytokines induced by the intracellular sensors Nod1 and Nod2 in an autophagy-independent manner. *Immunity* 39, 858-873.
- Spanogiannopoulos, P., Bess, E.N., Carmody, R.N., and Turnbaugh, P.J. (2016). The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. *Nat Rev Microbiol* 14, 273-287.
- Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* 9, 279-290.
- Steubesand, N., Kiehne, K., Brunke, G., Pahl, R., Reiss, K., Herzig, K.H., Schubert, S., Schreiber, S., Folsch, U.R., Rosenstiel, P., *et al.* (2009). The expression of the beta-defensins hBD-2 and hBD-3 is differentially regulated by NF-kappaB and MAPK/AP-1 pathways in an in vitro model of Candida esophagitis. *BMC Immunol* 10, 36.
- Sugawara, K., Olson, T.S., Moskaluk, C.A., Stevens, B.K., Hoang, S., Kozaiwa, K., Cominelli, F., Ley, K.F., and McDuffie, M. (2005). Linkage to peroxisome proliferator-activated receptor-gamma in SAMP1/YitFc mice and in human Crohn's disease. *Gastroenterology* 128, 351-360.

- Swidsinski, A., Dorffel, Y., Loening-Baucke, V., Theissig, F., Ruckert, J.C., Ismail, M., Rau, W.A., Gaschler, D., Weizenegger, M., Kuhn, S., *et al.* (2011). Acute appendicitis is characterised by local invasion with *Fusobacterium nucleatum/necrophorum*. *Gut* *60*, 34-40.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic acids research* *28*, 33-36.
- Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. *Nat Genet* *22*, 281-285.
- Thome, M. (2008). Multifunctional roles for MALT1 in T-cell activation. *Nat Rev Immunol* *8*, 495-500.
- Tong, M., Li, X., Wegener Parfrey, L., Roth, B., Ippoliti, A., Wei, B., Borneman, J., McGovern, D.P., Frank, D.N., Li, E., *et al.* (2013). A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. *PLoS one* *8*, e80702.
- Travassos, L.H., Carneiro, L.A., Ramjeet, M., Hussey, S., Kim, Y.G., Magalhaes, J.G., Yuan, L., Soares, F., Chea, E., Le Bourhis, L., *et al.* (2010). Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nat Immunol* *11*, 55-62.
- Tschurtschenthaler, M., Adolph, T.E., Ashcroft, J.W., Niederreiter, L., Bharti, R., Saveljeva, S., Bhattacharyya, J., Flak, M.B., Shih, D.Q., Fuhler, G.M., *et al.* (2017). Defective ATG16L1-mediated removal of IRE1alpha drives Crohn's disease-like ileitis. *The Journal of experimental medicine*.
- Turnbaugh, P.J., Backhed, F., Fulton, L., and Gordon, J.I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell host & microbe* *3*, 213-223.
- Ubeda, C., Bucci, V., Caballero, S., Djukovic, A., Toussaint, N.C., Equinda, M., Lipuma, L., Ling, L., Gobourne, A., No, D., *et al.* (2013). Intestinal microbiota containing *Barnesiella* species cures vancomycin-resistant *Enterococcus faecium* colonization. *Infection and immunity* *81*, 965-973.
- Ubeda, C., Lipuma, L., Gobourne, A., Viale, A., Leiner, I., Equinda, M., Khanin, R., and Pamer, E.G. (2012). Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *The Journal of experimental medicine* *209*, 1445-1456.
- Ubeda, C., Taur, Y., Jenq, R.R., Equinda, M.J., Son, T., Samstein, M., Viale, A., Socci, N.D., van den Brink, M.R., Kamboj, M., *et al.* (2010). Vancomycin-resistant *Enterococcus* domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *The Journal of clinical investigation* *120*, 4332-4341.
- Van Limbergen, J., Radford-Smith, G., and Satsangi, J. (2014). Advances in IBD genetics. *Nat Rev Gastroenterol Hepatol* *11*, 372-385.
- Vavricka, S.R., Musch, M.W., Chang, J.E., Nakagawa, Y., Phanvijhitsiri, K., Waypa, T.S., Merlin, D., Schneewind, O., and Chang, E.B. (2004). hPepT1 transports muramyl dipeptide, activating NF-kappaB and stimulating IL-8 secretion in human colonic Caco2/bbe cells. *Gastroenterology* *127*, 1401-1409.
- Vinolo, M.A., Rodrigues, H.G., Nachbar, R.T., and Curi, R. (2011). Regulation of inflammation by short chain fatty acids. *Nutrients* *3*, 858-876.
- Waidmann, M., Bechtold, O., Frick, J.S., Lehr, H.A., Schubert, S., Dobrindt, U., Loeffler, J., Bohn, E., and Autenrieth, I.B. (2003). *Bacteroides vulgatus* protects against *Escherichia coli*-induced colitis in gnotobiotic interleukin-2-deficient mice. *Gastroenterology* *125*, 162-177.

- Walsh, C. (2000). Molecular mechanisms that confer antibacterial drug resistance. *Nature* *406*, 775-781.
- Walter, J., and Ley, R. (2011). The human gut microbiome: ecology and recent evolutionary changes. *Annu Rev Microbiol* *65*, 411-429.
- Wang, J., Linnenbrink, M., Kunzel, S., Fernandes, R., Nadeau, M.J., Rosenstiel, P., and Baines, J.F. (2014). Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. *Proceedings of the National Academy of Sciences of the United States of America* *111*, E2703-2710.
- Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology* *73*, 5261-5267.
- Wehkamp, J., Salzman, N.H., Porter, E., Nuding, S., Weichenthal, M., Petras, R.E., Shen, B., Schaeffeler, E., Schwab, M., Linzmeier, R., *et al.* (2005). Reduced Paneth cell alpha-defensins in ileal Crohn's disease. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 18129-18134.
- Whitaker, K.W., Totoki, K., and Reyes, T.M. (2012). Metabolic adaptations to early life protein restriction differ by offspring sex and post-weaning diet in the mouse. *Nutr Metab Cardiovasc Dis* *22*, 1067-1074.
- White, J.R., Nagarajan, N., and Pop, M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS computational biology* *5*, e1000352.
- Windey, K., De Preter, V., and Verbeke, K. (2012). Relevance of protein fermentation to gut health. *Mol Nutr Food Res* *56*, 184-196.
- Wolf, K.J., Daft, J.G., Tanner, S.M., Hartmann, R., Khafipour, E., and Lorenz, R.G. (2014). Consumption of acidic water alters the gut microbiome and decreases the risk of diabetes in NOD mice. *J Histochem Cytochem* *62*, 237-250.
- Wolfe, A.J. (2015). Glycolysis for Microbiome Generation. *Microbiology spectrum* *3*.
- Wong, J.M.W., de Souza, R., Kendall, C.W.C., Emam, A., and Jenkins, D.J.A. (2006). Colonic health: Fermentation and short chain fatty acids. *Journal of Clinical Gastroenterology* *40*, 235-243.
- Woo, P.C.Y., Lau, S.K.P., Teng, J.L.L., Tse, H., and Yuen, K.Y. (2008). Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical Microbiology and Infection* *14*, 908-934.
- Woo, P.C.Y., Leung, K.W., Tsoi, H.W., Wong, S.S.Y., Teng, J.L.L., and Yuen, K.Y. (2002). Thermo-tolerant *Campylobacter fetus* bacteraemia identified by 16S ribosomal RNA gene sequencing: an emerging pathogen in immunocompromised patients. *Journal of Medical Microbiology* *51*, 740-746.
- Xiong, Y.J., Chen, D.P., Yu, C.C., Lv, B.C., Peng, J.Y., Wang, J.Y., and Lin, Y. (2015). Citrus nobilitein ameliorates experimental colitis by reducing inflammation and restoring impaired intestinal barrier function. *Molecular Nutrition & Food Research* *59*, 829-842.
- Yamamoto, T., Umegae, S., Kitagawa, T., and Matsumoto, K. (2005). Intraperitoneal cytokine productions and their relationship to peritoneal sepsis and systemic inflammatory markers in patients with inflammatory bowel disease. *Dis Colon Rectum* *48*, 1005-1015.

References

Yatsunenکو, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., *et al.* (2012). Human gut microbiome viewed across age and geography. *Nature* *486*, 222-227.

Young, V.B., and Schmidt, T.M. (2004). Antibiotic-associated diarrhea accompanied by large-scale alterations in the composition of the fecal microbiota. *Journal of clinical microbiology* *42*, 1203-1206.

Zerbino, D.R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* *18*, 821-829.

Zheng, H., and Wu, H. (2010). Short prokaryotic DNA fragment binning using a hierarchical classifier based on linear discriminant analysis and principal component analysis. *J Bioinform Comput Biol* *8*, 995-1011.

Zoetendal, E.G., Collier, C.T., Koike, S., Mackie, R.I., and Gaskins, H.R. (2004). Molecular ecological analysis of the gastrointestinal microbiota: a review. *J Nutr* *134*, 465-472.

Zucchelli, M., Torkvist, L., Bresso, F., Halfvarson, J., Hellquist, A., Anedda, F., Assadi, G., Lindgren, G.B., Svanfeldt, M., Janson, M., *et al.* (2009). PepT1 oligopeptide transporter (SLC15A1) gene polymorphism in inflammatory bowel disease. *Inflamm Bowel Dis* *15*, 1562-1569.

10 Annexure

Annexure I : Genetics of mice studied *Atg16l1*^{ΔIEC} mice

The conditional mouse model for *Atg16l1* knock out in epithelial cells (Δ IEC) was designed as a villin-promoter driven cre-recombinase excision. Exon 1 of the *Atg16l1* cDNA sequence AB087881 was flanked by loxP-sites and excision of exon 1 containing the transcription start sequence targets all major splice variants of *Atg16l1*. The details and the model are described elsewhere (Cadwell et al., 2008). Efficiency of the conditional knock-out was controlled by immunofluorescence staining of *Atg16l1* in the intestines. Control wild-type mice (*WT*) were bred as littermates to the LoxP/cre homozygous Δ IEC mice (*Atg16l1*^{ΔIEC}/KO) to minimize disturbing effects of genetic background.

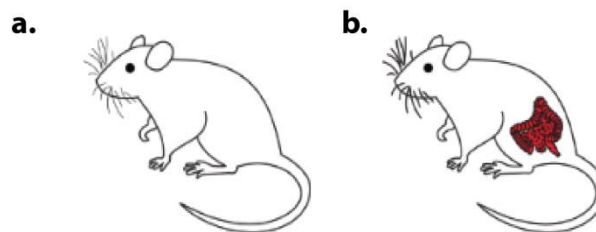


Figure showing a) *WT* and b) Conditional knockout of *Atg16l1* in intestinal epithelial cells

Annexure II : Table showing Primer/Barcode used in the study

PD = Post-delivery; AD = after delivery

Target gene/Region	Name	DNA-Sequence (5'-3')
16SrRNA gene/V1-V2	Pyro_27F	5'-TATGCGCCTTGCCAGCCCCTCAGTCAGAG TTTGATCCTGGCTCAG-3'
16SrRNA gene/V3-V4	(319F/806R)	5'-ACTCCTACGGGAGGCAGCAG-3' 5'-GGACTACHVGGGTWTCTAAT-3'
for Pyrosequencing	MIDX_338R	5'-CGTATCGCCTCCCTCGGCCATCAG(MID) CATGCTGCCTCCCGTAGGAGT-3'
Study (Chapter 1)	(MID/barcode)	
acute Appendectomy	MID33_338R	ATAGAGTACT
acute Appendectomy	MID34_348R	CACGCTACGT
acute Appendectomy	MID35_348R	CAGTAGACGT
acute Appendectomy	MID36_348R	CGACGTGACT
elective Appendectomy	MID49_338R	ACGCGATCGA
elective Appendectomy	MID50_338R	ACTAGCAGTA

elective Appendectomy	MID51_338R	AGCTCACGTA
elective Appendectomy	MID52_338R	AGTATACATA
Study (Chapter 2)	(MID/barcode)	
<i>Atg16/1</i> AB study	MID1_338R	ACGAGTGCGT
<i>Atg16/1</i> AB study	MID2_338R	ACGCTCGACA
<i>Atg16/1</i> AB study	MID3_338R	AGACGCACTC
<i>Atg16/1</i> AB study	MID4_338R	AGCACTGTAG
<i>Atg16/1</i> AB study	MID5_338R	ATCAGACACG
<i>Atg16/1</i> AB study	MID6_338R	ATATCGCGAG
<i>Atg16/1</i> AB study	MID7_338R	CGTGTCTCTA
<i>Atg16/1</i> AB study	MID8_338R	CTCGCGTGTC
<i>Atg16/1</i> AB study	MID10_338R	TCTCTATGCG
<i>Atg16/1</i> AB study	MID11_338R	TGATACGTCT
<i>Atg16/1</i> AB study	MID13_338R	CATAGTAGTG
<i>Atg16/1</i> AB study	MID14_338R	CGAGAGATAC
<i>Atg16/1</i> AB study	MID15_338R	ATACGACGTA
<i>Atg16/1</i> AB study	MID16_338R	TCACGTACTA
<i>Atg16/1</i> AB study	MID17_338R	CGTCTAGTAC
<i>Atg16/1</i> AB study	MID18_338R	TCTACGTAGC
<i>Atg16/1</i> AB study	MID19_338R	TGTACTACTC
<i>Atg16/1</i> AB study	MID20_338R	ACGACTACAG
<i>Atg16/1</i> AB study	MID21_338R	CGTAGACTAG
<i>Atg16/1</i> AB study	MID22_338R	TACGAGTATG
<i>Atg16/1</i> AB study	MID23_338R	TACTCTCGTG
<i>Atg16/1</i> AB study	MID24_338R	TAGAGACGAG
<i>Atg16/1</i> AB study	MID25_338R	TCGTCGCTCG
Study (Chapter 3)	(MID/barcode)	
<i>Atg16/1</i> Pregnancy study(PD)/S1	Jac 319F_10/Jac 806R_2	ACAGCCACCCAT CGA/TGCAGATCCAAC
<i>Atg16/1</i> Pregnancy study(PD)/S3	Jac 319F_10/Jac 806R_4	ACAGCCACCCAT CGA/GTGGTATGGGAGA
<i>Atg16/1</i> Pregnancy study(PD)/S4	Jac 319F_10/Jac 806R_5	ACAGCCACCCAT CGA/ACTTTAAGGGTG A
<i>Atg16/1</i> Pregnancy study(PD)/S5	Jac 319F_10/Jac 806R_7	ACAGCCACCCAT CGA/TGTTGCGTTTCT TC
<i>Atg16/1</i> Pregnancy study(PD)/S6	Jac 319F_10/Jac 806R_8	ACAGCCACCCAT CGA/ATGTCCGACCAA TC
<i>Atg16/1</i> Pregnancy study(PD)/S8	Jac 319F_10/Jac 806R_10	ACAGCCACCCAT CGA/ACAGCCACCCAT CTA
<i>Atg16/1</i> Pregnancy study(PD)/S9	Jac 319F_10/Jac 806R_10	ACAGCCACCCAT CGA/TGTCTCGCAAGC CTA

Annexure

<i>Atg16/1</i> Pregnancy study(PD)/S10	Jac 319F_10/Jac 806R_12	ACAGCCACCCAT CGA/GAGGAGTAAAGC CTA
<i>Atg16/1</i> Pregnancy study(PD)/S11	Jac 319F_10/Jac 806R_13	ACAGCCACCCAT CGA/GTTACGTGGTTG GATA
<i>Atg16/1</i> Pregnancy study(PD)/S12	Jac 319F_10/Jac 806R_14	ACAGCCACCCAT CGA/TACCGCCTCGGA GATA
<i>Atg16/1</i> Pregnancy study(PD)/S13	Jac 319F_10/Jac 806R_16	ACAGCCACCCAT CGA/TACCGGCTTGCA ACTCA
<i>Atg16/1</i> Pregnancy study(PD)/S14	Jac 319F_10/Jac 806R_17	ACAGCCACCCAT CGA/ATCTAGTGGCAA ACTCA
<i>Atg16/1</i> Pregnancy study(PD)/S15	Jac 319F_10/Jac 806R_19	ACAGCCACCCAT CGA/CACCTTACCTTA TTCTCT
<i>Atg16/1</i> Pregnancy study(PD)/S17	Jac 319F_10/Jac 806R_22	ACAGCCACCCATCGA/TTAACTGGAAGCCACTTCT
<i>Atg16/1</i> Pregnancy study(PD)/S18	Jac 319F_10/Jac 806R_23	ACAGCCACCCATCGA/CGCGGTTACTAACACTTCT
<i>Atg16/1</i> Pregnancy study(PD)/S19	Jac 319F_10/Jac 806R_24	ACAGCCACCCATCGA/GAGACTATATGCCACTTCT
<i>Atg16/1</i> Pregnancy study(PD)/S20	Jac 319F_11/Jac 806R_1	TGTCTCGCAAGC CGA/CCTAAACTACGG
<i>Atg16/1</i> Pregnancy study(PD)/S21	Jac 319F_11/Jac 806R_2	TGTCTCGCAAGC CGA/TGCAGATCCAAC
<i>Atg16/1</i> Pregnancy study(PD)/S22	Jac 319F_11/Jac 806R_3	TGTCTCGCAAGC CGA/CCATCACATAGG
<i>Atg16/1</i> Pregnancy study(PD)/S23	Jac 319F_11/Jac 806R_4	TGTCTCGCAAGC CGA/GTGGTATGGGAGA
<i>Atg16/1</i> Pregnancy study(PD)/S24	Jac 319F_11/Jac 806R_5	TGTCTCGCAAGC CGA/ACTTTAAGGGTG A
<i>Atg16/1</i> Pregnancy study(PD)/S25	Jac 319F_11/Jac 806R_7	TGTCTCGCAAGC CGA/TGTTGCGTTTCT TC
<i>Atg16/1</i> Pregnancy study(PD)/S26	Jac 319F_11/Jac 806R_8	TGTCTCGCAAGC CGA/ATGTCCGACCAA TC
<i>Atg16/1</i> Pregnancy study(PD)/S27	Jac 319F_11/Jac 806R_9	TGTCTCGCAAGC CGA/AGGTACGCAATT TC
<i>Atg16/1</i> Pregnancy study(PD)/S28	Jac 319F_11/Jac 806R_10	TGTCTCGCAAGC CGA/ACAGCCACCCAT CTA
<i>Atg16/1</i> Pregnancy study(PD)/S29	Jac 319F_11/Jac 806R_11	TGTCTCGCAAGC CGA/TGTCTCGCAAGC CTA
<i>Atg16/1</i> Pregnancy study(PD)/S30	Jac 319F_11/Jac 806R_12	TGTCTCGCAAGC CGA/GAGGAGTAAAGC CTA
<i>Atg16/1</i> Pregnancy study(PD)/S31	Jac 319F_11/Jac 806R_13	TGTCTCGCAAGC CGA/GTTACGTGGTTG GATA
<i>Atg16/1</i> Pregnancy study(PD)/S32	Jac 319F_11/Jac 806R_14	TGTCTCGCAAGC CGA/TACCGCCTCGGA GATA
<i>Atg16/1</i> Pregnancy study(PD)/S33	Jac 319F_11/Jac 806R_16	TGTCTCGCAAGC CGA/TACCGGCTTGCA ACTCA
<i>Atg16/1</i> Pregnancy study(PD)/S34	Jac 319F_11/Jac 806R_17	TGTCTCGCAAGC CGA/ATCTAGTGGCAA ACTCA
<i>Atg16/1</i> Pregnancy study(PD)/S35	Jac 319F_11/Jac 806R_19	TGTCTCGCAAGC CGA/CACCTTACCTTA TTCTCT
<i>Atg16/1</i> Pregnancy study(PD)/S36	Jac 319F_11/Jac 806R_20	TGTCTCGCAAGC CGA/ATAGTTAGGGCT TTCTCT
<i>Atg16/1</i> Pregnancy study(PD)/S37	Jac 319F_11/Jac 806R_22	TGTCTCGCAAGCCGA/TTAACTGGAAGCCACTTCT
<i>Atg16/1</i> Pregnancy study(PD)/S38	Jac 319F_11/Jac 806R_23	TGTCTCGCAAGCCGA/CGCGGTTACTAACACTTCT

Annexure

<i>Atg16/1</i> Pregnancy study(PD)/S39	Jac 319F_11/Jac 806R_24	TGTCTCGCAAGCCGA/GAGACTATATGCCACTTCT
<i>Atg16/1</i> Pregnancy study(PD)/S40	Jac 319F_12/Jac 806R_1	GAGGAGTAAAGC CGA/CCTAAACTACGG
<i>Atg16/1</i> Pregnancy study(PD)/S41	Jac 319F_12/Jac 806R_2	GAGGAGTAAAGC CGA/TGCAGATCCAAC
<i>Atg16/1</i> Pregnancy study(PD)/S42	Jac 319F_12/Jac 806R_3	GAGGAGTAAAGC CGA/CCATCACATAGG
<i>Atg16/1</i> Pregnancy study(PD)/S43	Jac 319F_12/Jac 806R_4	GAGGAGTAAAGC CGA/GTGGTATGGGAGA
<i>Atg16/1</i> Pregnancy study(PD)/S44	Jac 319F_12/Jac 806R_5	GAGGAGTAAAGC CGA/ACTTTAAGGGTG A
<i>Atg16/1</i> Pregnancy study(PD)/S45	Jac 319F_12/Jac 806R_7	GAGGAGTAAAGC CGA/TGTTGCGTTTCT TC
<i>Atg16/1</i> Pregnancy study(PD)/S46	Jac 319F_12/Jac 806R_8	GAGGAGTAAAGC CGA/ATGTCCGACCAA TC
<i>Atg16/1</i> Pregnancy study(PD)/S47	Jac 319F_12/Jac 806R_9	GAGGAGTAAAGC CGA/AGGTACGCAATT TC
<i>Atg16/1</i> Pregnancy study(PD)/S48	Jac 319F_12/Jac 806R_10	GAGGAGTAAAGC CGA/ACAGCCACCCAT CTA
<i>Atg16/1</i> Pregnancy study(PD)/S49	Jac 319F_12/Jac 806R_11	GAGGAGTAAAGC CGA/TGTCTCGCAAGC CTA
<i>Atg16/1</i> Pregnancy study(PD)/S50	Jac 319F_12/Jac 806R_12	GAGGAGTAAAGC CGA/GAGGAGTAAAGC CTA
<i>Atg16/1</i> Pregnancy study(PD)/S51	Jac 319F_12/Jac 806R_13	GAGGAGTAAAGC CGA/GTTACGTGGTTG GATA
<i>Atg16/1</i> Pregnancy study(PD)/S52	Jac 319F_12/Jac 806R_14	GAGGAGTAAAGC CGA/TACCGCCTCGGA GATA
<i>Atg16/1</i> Pregnancy study(PD)/S53	Jac 319F_12/Jac 806R_16	GAGGAGTAAAGC CGA/TACCGGCTTGCA ACTCA
<i>Atg16/1</i> Pregnancy study(PD)/S54	Jac 319F_12/Jac 806R_17	GAGGAGTAAAGC CGA/ATCTAGTGGCAACTCA
<i>Atg16/1</i> Pregnancy study(PD)/S55	Jac 319F_12/Jac 806R_19	GAGGAGTAAAGC CGA/CACCTTACCTTA TTCTCT
<i>Atg16/1</i> Pregnancy study(AD)/S56_AP	Jac 319F_1/Jac 806R_9	CCTAAACTACGG/AGGTACGCAATTC
<i>Atg16/1</i> Pregnancy study(AD)/S57_AP	Jac 319F_1/Jac 806R_10	CCTAAACTACGG/ACAGCCACCCATCTA
<i>Atg16/1</i> Pregnancy study(AD)/S58_AP	Jac 319F_1/Jac 806R_11	CCTAAACTACGG/TGTCTCGCAAGCCTA
<i>Atg16/1</i> Pregnancy study(AD)/S59_AP	Jac 319F_1/Jac 806R_12	CCTAAACTACGG/GAGGAGTAAAGCCTA
<i>Atg16/1</i> Pregnancy study(AD)/S60_AP	Jac 319F_1/Jac 806R_13	CCTAAACTACGG/GTTACGTGGTTGGATA
<i>Atg16/1</i> Pregnancy study(AD)/S61_AP	Jac 319F_1/Jac 806R_14	CCTAAACTACGG/TACCGCCTCGGAGATA
<i>Atg16/1</i> Pregnancy study(AD)/S62_AP	Jac 319F_1/Jac 806R_16	CCTAAACTACGG/TACCGGCTTGCAACTCA
<i>Atg16/1</i> Pregnancy study(AD)/S63_AP	Jac 319F_1/Jac 806R_17	CCTAAACTACGG/ATCTAGTGGCAACTCA
<i>Atg16/1</i> Pregnancy study(AD)/S64_AP	Jac 319F_1/Jac 806R_19	CCTAAACTACGG/CACCTTACCTTATTCTCT
<i>Atg16/1</i> Pregnancy study(AD)/S65_AP	Jac 319F_1/Jac 806R_20	CCTAAACTACGG/ATAGTTAGGGCTTTCTCT
<i>Atg16/1</i> Pregnancy study(AD)/S66_AP	Jac 319F_1/Jac 806R_22	CCTAAACTACGG/TTAACTGGAAGCCACTTCT

Annexure

<i>Atg16/1</i> Pregnancy study(AD)/S67_AP	Jac 319F_1/Jac 806R_23	CCTAAACTACGG/CGCGGTTACTAACACTTCT
<i>Atg16/1</i> Pregnancy study(AD)/S68_AP	Jac 319F_1/Jac 806R_24	CCTAAACTACGG/GAGACTATATGCCACTTCT
<i>Atg16/1</i> Pregnancy study(AD)/S69_AP	Jac 319F_2/Jac 806R_1	TGCAGATCCAAC/CCTAAACTACGG
<i>Atg16/1</i> Pregnancy study(AD)/S70_AP	Jac 319F_2/Jac 806R_2	TGCAGATCCAAC/TGCAGATCCAAC
<i>Atg16/1</i> Pregnancy study(AD)/S71_AP	Jac 319F_2/Jac 806R_3	TGCAGATCCAAC/CCATCACATAGG
<i>Atg16/1</i> Pregnancy study(AD)/S72_AP	Jac 319F_2/Jac 806R_4	TGCAGATCCAAC/GTGGTATGGGAGA
<i>Atg16/1</i> Pregnancy study(AD)/S73_AP	Jac 319F_2/Jac 806R_5	TGCAGATCCAAC/ACTTTAAGGGTGA
<i>Atg16/1</i> Pregnancy study(AD)/S74_AP	Jac 319F_2/Jac 806R_7	TGCAGATCCAAC/TGTTGCGTTTCTTC
<i>Atg16/1</i> Pregnancy study(AD)/S75_AP	Jac 319F_2/Jac 806R_8	TGCAGATCCAAC/ATGTCCGACCAATC
<i>Atg16/1</i> Pregnancy study(AD)/S76_AP	Jac 319F_2/Jac 806R_9	TGCAGATCCAAC/AGGTACGCAATTC
<i>Atg16/1</i> Pregnancy study(AD)/S77_AP	Jac 319F_2/Jac 806R_10	TGCAGATCCAAC/ACAGCCACCCATCTA
<i>Atg16/1</i> Pregnancy study(AD)/S78_AP	Jac 319F_2/Jac 806R_11	TGCAGATCCAAC/TGTCTCGCAAGCCTA
<i>Atg16/1</i> Pregnancy study(AD)/S79_AP	Jac 319F_2/Jac 806R_12	TGCAGATCCAAC/GAGGAGTAAAGCCTA

Annexure III: Taxonomic classification for Appendicitis study showing mean relative abundance

Phylum level	Inflamed	Non-Inflamed
<i>p__Actinobacteria</i>	0.12	1.4
<i>p__Bacteroidetes</i>	32.96	34.32
<i>p__Firmicutes</i>	50.32	49.68
<i>p__Fusobacteria</i>	5.48	0.68
<i>p__Lentisphaerae</i>	0	0.12
<i>p__Proteobacteria</i>	7.4	10
<i>p__Tenericutes</i>	2.36	3.24
unclassified	1.36	0.56
Class level	Inflamed	Non-Inflamed
<i>c__Actinobacteria</i>	0.12	1.4
<i>c__Bacilli</i>	1.4	1.88
<i>c__Bacteroidia</i>	32.92	33.72
<i>c__Betaproteobacteria</i>	5.84	3.6
<i>c__Clostridia</i>	48.68	47.72
<i>c__Erysipelotrichi</i>	2.16	3.08
<i>c__Fusobacteria</i>	5.48	0.68
<i>c__Gammaproteobacteria</i>	1.08	5.8

unclassified	1.6	0.64
Others	0.72	1.48
Order level	Inflamed	Non-Inflamed
<i>o__Bacteroidales</i>	32.92	33.72
<i>o__Burkholderiales</i>	5.2	3.24
<i>o__Clostridiales</i>	48.68	47.72
<i>o__Coriobacteriales</i>	0.12	0.64
<i>o__Enterobacteriales</i>	1.04	4.2
<i>o__Erysipelotrichales</i>	2.16	3.08
<i>o__Fusobacteriales</i>	5.48	0.68
unclassified	2.04	0.76
Others	2.36	5.96
Family level	Inflamed	Non-Inflamed
<i>f__Alcaligenaceae</i>	4.96	1.92
<i>f__Bacteroidaceae</i>	21.36	18.88
<i>f__Enterobacteriaceae</i>	1.04	4.2
<i>f__Erysipelotrichaceae</i>	2.16	3.08
<i>f__Fusobacteriaceae</i>	5.48	0.68
<i>f__Lachnospiraceae</i>	27.24	23.36
<i>f__Porphyromonadaceae</i>	7.8	3.92
<i>f__Prevotellaceae</i>	2.2	8.16
<i>f__Rikenellaceae</i>	1.32	1.08
<i>f__Ruminococcaceae</i>	11.48	20.8
<i>f__Veillonellaceae</i>	7.96	1.28
unclassified	2.76	3.6
Others	4.24	9.04
Genus level	Inflamed	Non-Inflamed
<i>g__Alistipes</i>	1.32	1.08
<i>g__Bacteroides</i>	22.6	19.32
<i>g__Blautia</i>	0.52	1.8
<i>g__Clostridium</i>	15.64	7.4
<i>g__Coprococcus</i>	0.6	1.2
<i>g__Enterobacter</i>	1.04	3.72
<i>g__Faecalibacterium</i>	1.52	13.28
<i>g__Fusobacterium</i>	5.44	0.68
<i>g__Lachnobacterium</i>	3.28	1.44
<i>g__Odoribacter</i>	1.6	0.36

<i>g__Oscillospira</i>	4.68	3.08
<i>g__Parabacteroides</i>	0.44	2.64
<i>g__Prevotella</i>	2.2	8.16
<i>g__Ruminococcus</i>	1.48	3.88
<i>g__Selenomonas</i>	7.16	0.2
<i>g__Subdoligranulum</i>	0.28	1.12
<i>g__Sutterella</i>	4.96	1.8
unclassified	18.76	16.44
Others	6.48	12.4
Species level	Inflamed	Non-Inflamed
<i>s__Bacteroidescapillosu</i>	1.24	0.44
<i>s__Bacteroidescrophilu</i>	0	1.36
<i>s__Bacteroidesdore</i>	9.76	0.96
<i>s__Bacteroidesfragili</i>	3.08	0.04
<i>s__Bacteroidesmassiliensi</i>	0	3.56
<i>s__Bacteroidesuniformi</i>	1.12	1.04
<i>s__Clostridiumboltea</i>	4.6	1.6
<i>s__Clostridiumorbiscinden</i>	1.12	0.32
<i>s__Enterobacterhormaeche</i>	1.04	3.72
<i>s__Faecalibacteriumprausnitzi</i>	1.48	13.08
<i>s__Parabacteroidesdistasoni</i>	0.44	2.48
<i>s__Prevotellacopr</i>	0.08	2
<i>s__Prevotellanigrescen</i>	2.12	0.08
unclassified	66.52	54.48
Others	7.4	14.84

Annexure-IV Table 1: Pyrosequencing reads remaining after quality filtering, screening unaligned reads and chimera removal for Wildtype and *Atg16l1*^{ΔIEC} at all-time points.

Reads after Quality control

Wildtype (Day-00)		<i>Atg16l1</i> ^{ΔIEC} (Day-00)		Wildtype (Day-21)		<i>Atg16l1</i> ^{ΔIEC} (Day-21)		Wildtype (Day-56)		<i>Atg16l1</i> ^{ΔIEC} (Day-56)	
Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count
01_d00	3972	07_d00	3791	01_d21	1434	07_d21	3060	01_d56	5781	07_d56	4716
02_d00	2696	08_d00	4436	02_d21	2402	08_d21	2049	02_d56	7748	08_d56	8282
03_d00	2058	10_d00	3594	03_d21	1785	10_d21	4164	03_d56	7034	10_d56	8320

04_d00	1643	11_d00	4389	04_d21	2196	11_d21	3389	04_d56	6061	11_d56	8244
05_d00	5508	13_d00	3903	05_d21	1395	13_d21	2122	05_d56	6052	13_d56	7467
06_d00	6958	14_d00	3439	06_d21	1499	14_d21	2363	06_d56	4724	14_d56	10290
15_d00	12880	21_d00	1528	15_d21	2062	21_d21	5110	15_d56	7646	21_d56	3062
16_d00	3070	22_d00	7865	16_d21	2701	22_d21	5119	16_d56	8388	22_d56	18096
17_d00	7084	23_d00	4768	17_d21	1890	23_d21	3309	17_d56	3747	23_d56	6512
18_d00	4686	24_d00	2917	18_d21	1666	24_d21	4017	18_d56	4297	24_d56	2720
19_d00	3741	25_d00	2208	19_d21	5871	25_d21	3820	19_d56	5012	25_d56	5284
20_d00	1837			20_d21	1235			20_d56	3656		
Total	56133	Total	42838	Total	26136	Total	38522	Total	70146	Total	82993

Reads remaining after screening unaligned reads

Wildtype (Day-00)		<i>Atg16l1</i>^{ΔIEC} (Day-00)		Wildtype (Day-21)		<i>Atg16l1</i>^{ΔIEC} (Day-21)		Wildtype (Day-56)		<i>Atg16l1</i>^{ΔIEC} (Day-56)	
Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count
01_d00	3824	21_d00	1430	01_d21	1393	07_d21	3021	01_d56	5372	07_d56	4438
02_d00	2559	22_d00	7352	02_d21	2315	08_d21	2009	02_d56	7595	08_d56	7835
03_d00	1938	23_d00	4461	03_d21	1694	10_d21	4024	03_d56	6860	10_d56	8052
04_d00	1581	24_d00	2796	04_d21	2125	11_d21	3245	04_d56	5790	11_d56	7966
15_d00	12368	25_d00	2054	05_d21	1322	13_d21	2022	05_d56	5828	13_d56	7123
16_d00	2884	10_d00	3479	06_d21	1433	14_d21	2309	06_d56	4505	14_d56	9653
17_d00	6630	11_d00	4311	15_d21	1922	21_d21	4625	15_d56	6999	21_d56	2804
18_d00	4421	13_d00	3691	16_d21	2419	22_d21	4863	16_d56	7741	22_d56	17685
19_d00	3561	14_d00	3230	17_d21	1729	23_d21	3121	17_d56	3558	23_d56	6400
20_d00	1737	07_d00	3622	18_d21	1559	24_d21	3899	18_d56	4117	24_d56	2643
05_d00	5312	08_d00	4241	19_d21	5626	25_d21	3644	19_d56	4740	25_d56	5236
06_d00	6793			20_d21	1159			20_d56	3537		
Total	53608	Total	40667	Total	24696	Total	36782	Total	66642	Total	79835

Reads remaining after chimera removal

Wildtype (Day-00)		<i>Atg16l1</i> ^{ΔIEC} (Day-00)		Wildtype (Day-21)		<i>Atg16l1</i> ^{ΔIEC} (Day-21)		Wildtype (Day-56)		<i>Atg16l1</i> ^{ΔIEC} (Day-56)	
Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count	Sample	Read Count
01_d00	2738	13_d00	2163	01_d21	1331	07_d21	2996	01_d56	3348	07_d56	3519
02_d00	1816	14_d00	1714	02_d21	1232	08_d21	1985	02_d56	2675	08_d56	5821
03_d00	1247	21_d00	714	03_d21	1248	10_d21	2331	03_d56	1045	10_d56	3786
04_d00	881	22_d00	3968	04_d21	1792	11_d21	1216	04_d56	3785	11_d56	3700
15_d00	7248	23_d00	2425	05_d21	1192	13_d21	1551	05_d56	3678	13_d56	2583
16_d00	1800	24_d00	1990	06_d21	1301	14_d21	2120	06_d56	2721	14_d56	6906
17_d00	3536	25_d00	1152	15_d21	1097	21_d21	3209	15_d56	4109	21_d56	1491
18_d00	2150	10_d00	2401	16_d21	1660	22_d21	3642	16_d56	5267	22_d56	12386
19_d00	2159	11_d00	3049	17_d21	1483	23_d21	2413	17_d56	2253	23_d56	5145
20_d00	1195	07_d00	1943	18_d21	1428	24_d21	3498	18_d56	3134	24_d56	1966
05_d00	3086	08_d00	2325	20_d21	1031	25_d21	2970	19_d56	3478	25_d56	4368
06_d00	4016			19_d21	4255			20_d56	2434		
Total	31872		23844	Total	19050	Total	27931	Total	37927	Total	51671

Annexure-IV Table 2: Representative average values for major diversity indices (richness and evenness) for assessing diversity for Day 00, Day 21 and Day 56 sampling groups.

Sample	Chao1 richness estimate	Shannon diversity index	Good's estimate coverage
Day 00	287.86	2.88	0.93
Day 21	96.27	1.81	0.97
Day 56	241.34	2.44	0.94

Annexure-IV Table 3: Analysis of Deviance % based on comparison of full and null LME models.

	Df	Sum Sq	Mean Sq	F value	upper. den.df	upper. p.val	lower.den.d f	lower. p.val	expl.dev .(%)
Day	2	2806.671	1403.336	27.8065	57	0	34	0	34.1638
Mutant	1	0.261	0.261	0.0052	57	0.9429	34	0.9431	0.0032
Gender	1	333.4984	333.4984	6.6081	57	0.0128	34	0.0147	4.0595
Day:Mutant	2	250.4822	125.2411	2.4816	57	0.0926	34	0.0986	3.049
Day:Gender	2	578.5182	289.2591	5.7315	57	0.0054	34	0.0072	7.0419
Mutant:Gender	1	3.4349	3.4349	0.0681	57	0.7951	34	0.7958	0.0418
Day:Mutant:Gender	2	16.942	8.471	0.1678	57	0.8459	34	0.8462	0.2062

Annexure-IV Table 4: Comparative time-wise phylum to genus level taxonomic classification for WT and *Atg16l1*^{ΔIEC} (Electronic copy).**Annexure-IV Table 5: Metastats/Indicator analysis for WT and *Atg16l1*^{ΔIEC} (Electronic copy).****Annexure-V Table 1: Showing metagenomics reads passed by quality control from MGRAST**

D00_ <i>Atg16l1</i> ^{WT}			T3_ <i>Atg16l1</i> ^{WT}		
Sequences Count	Post QC: Sequences Count	Individual percentage	Sequences Count	Post QC: Sequences Count	Individual percentage
13,415,537	8,708,430	64.91301839	11,471,838	7,931,560	69.13940033
13,716,708	10,118,060	73.76449218	13,675,085	10,011,611	73.2105943
12,044,797	8,937,015	74.19813717	16,237,464	10,981,574	67.63109067
16,371,216	11,380,188	69.51339473	7,540,271	5,076,189	67.32104191
11,353,936	7,491,783	65.98401647	14,387,217	9,982,695	69.38586524
5,356,238	3,934,266	73.45203854	7,399,512	5,278,177	71.33142023
15,134,167	9,694,542	64.05732142	12,667,193	8,438,441	66.61650296
D00_ <i>Atg16l1</i> ^{ΔIEC}			T3_ <i>Atg16l1</i> ^{ΔIEC}		
Sequences Count	Post QC: Sequences Count	Individual percentage	Sequences Count	Post QC: Sequences Count	Individual percentage
9,964,030	6,653,960	66.77980697	13,472,166	8,848,109	65.67695944
18,204,367	12,248,678	67.28428404	15,058,599	9,871,810	65.55596573

12,915,217	8,525,482	66.01114019		11,624,680	8,924,342	76.77064659
4,205,095	3,303,488	78.55917643		9,773,007	6,626,602	67.80514943
8,437,253	6,206,752	73.56365869		11,925,462	9,105,099	76.35007348
13,583,581	10,604,131	78.0657987		10,148,765	7,543,683	74.33104422
11,594,121	7,934,545	68.43593404		15,802,029	11,143,067	70.51668491

Annexure-V Table 2: Showing alpha diversity for WSG

α -Diversity				
	D00_ <i>Atg16l1</i> ^{fl/fl}	D00_ <i>Atg16l1</i> ^{ΔIEC}	T3_ <i>Atg16l1</i> ^{fl/fl}	T3_ <i>Atg16l1</i> ^{ΔIEC}
Sample1	146.347 species	168.283 species	190.411 species	145.859 species
Sample2	213.424 species	185.923 species	213.722 species	172.888 species
Sample3	222.253 species	130.662 species	165.911 species	212.023 species
Sample4	188.337 species	229.276 species	174.713 species	164.000 species
Sample5	156.307 species	200.421 species	206.856 species	224.428 species
Sample6	178.326 species	226.358 species	161.931 species	225.676 species
Sample7	147.761 species	164.426 species	167.760 species	213.033 species

Annexure-V Table 3 Showing KEGG pathways classification for all metagenomics samples (Electronic attachment).

Annexure-V Table 4 Showing enriched and depleted KEGG ortholog terms showing: 1) D00_ *Atg16l1*^{fl/fl} v/s D00_ *Atg16l1* ^{Δ IEC}, 2) T3_ *Atg16l1*^{fl/fl} v/s T3_ *Atg16l1* ^{Δ IEC}, 3) D00_ *Atg16l1*^{fl/fl} v/s T3_ *Atg16l1*^{fl/fl}, 4) D00_ *Atg16l1* ^{Δ IEC} v/s T3_ *Atg16l1* ^{Δ IEC} (Electronic attachment).

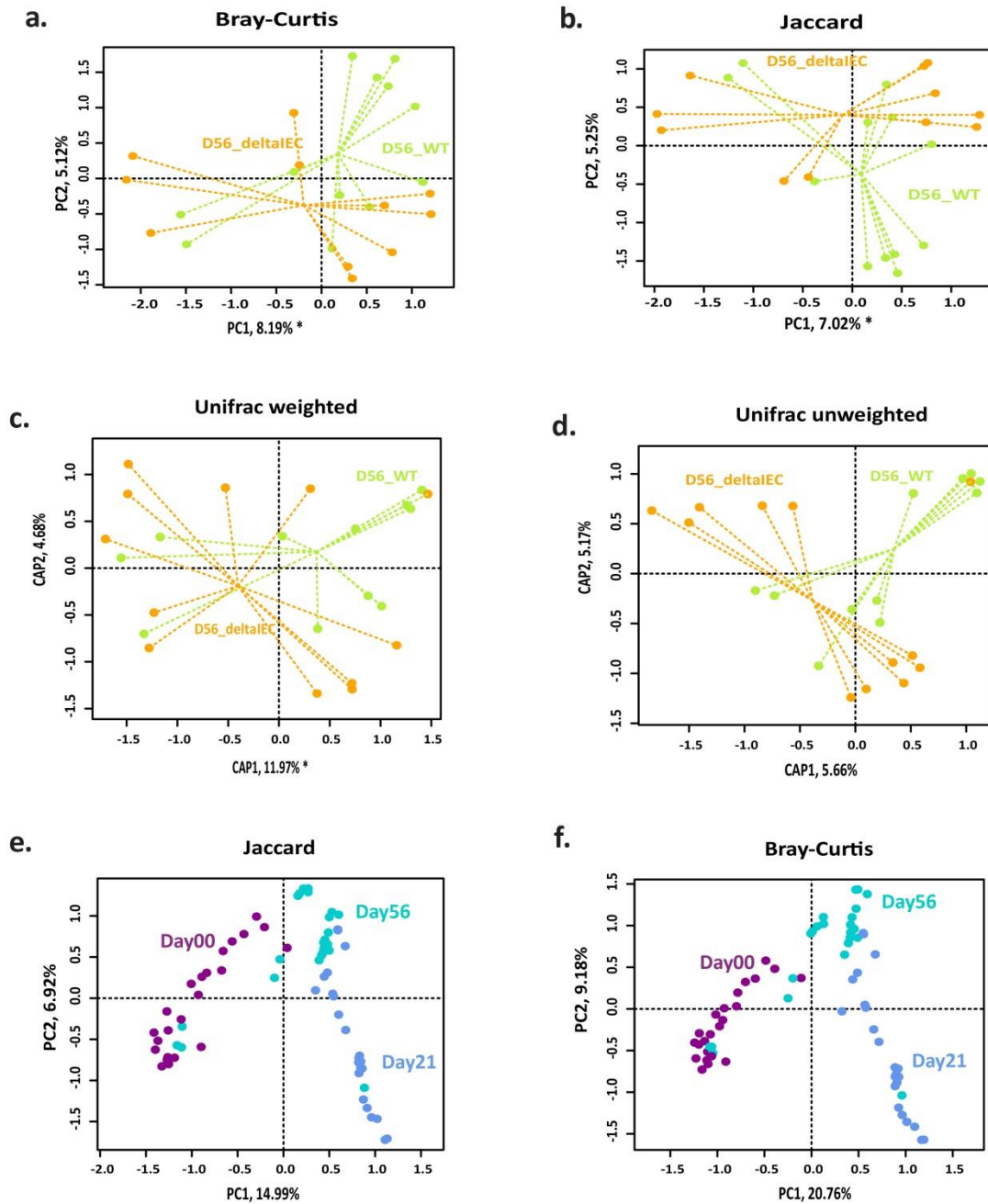


Figure S1 Constrained ordination analysis constrained with Genotype, Gender for Day56 using (a) Bray-Curtis dissimilarity; (b) Jaccard index; (c) unweighted Unifrac distances and (d) weighted Unifrac distances. (d-e) Ordination analysis of the significant constraining factors of Day/AB effect. Each dot represents one mouse and are colored based on Day/AB: Day00-darkmagenta, Day21-blue, Day56-cyan) using (d) Bray-Curtis dissimilarity; (e) Jaccard index (**Also available as Electronic Copy – S1**)

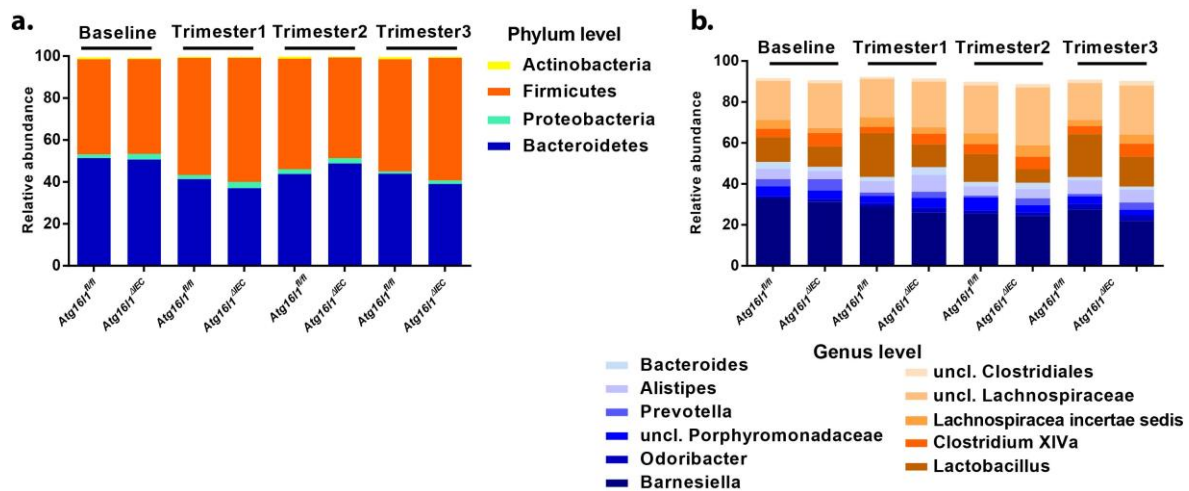


Figure S2 Taxonomic profiling of 16S rRNA data showing (a) phylum and (b) genus level for *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* at Baseline (D00), Trimester1 (T1) and Trimester3 (T3) for WT and *Atg16l1^{ΔIEC}*. (Also available as Electronic Copy – S2)

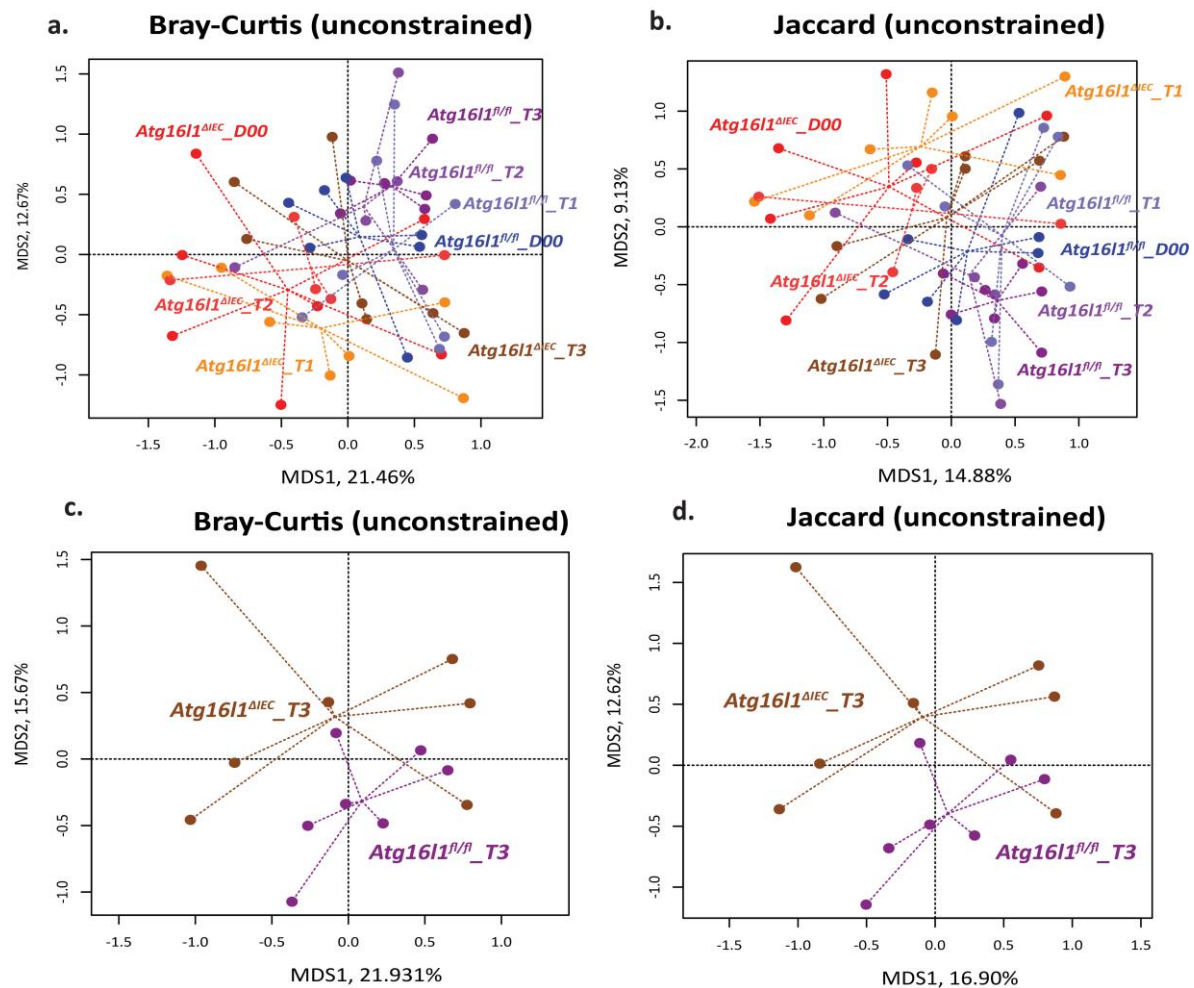


Figure S3 Unconstrained ordination plots for WT and *Atg16l1^{ΔIEC}* showing (a-b) from day 0(D00), Trimester1 (T1), Trimester2 (T2) and Trimester3 (T3) using Bray-Curtis and Jaccard distances; (c-d) from Trimester3 (T3) of WT and *Atg16l1^{ΔIEC}* using Bray-Curtis and Jaccard distances. (Also available as Electronic Copy – S3)

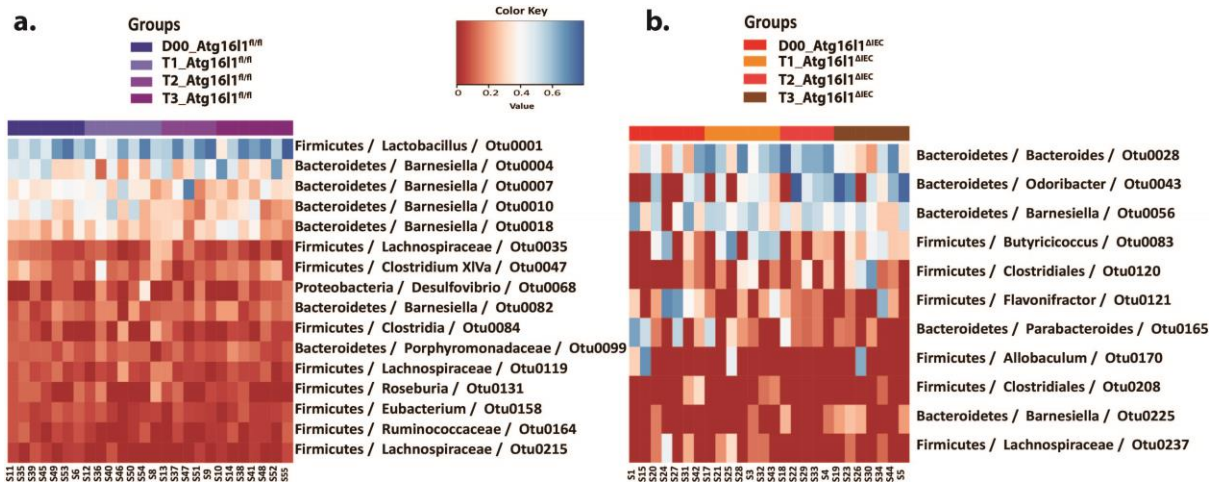
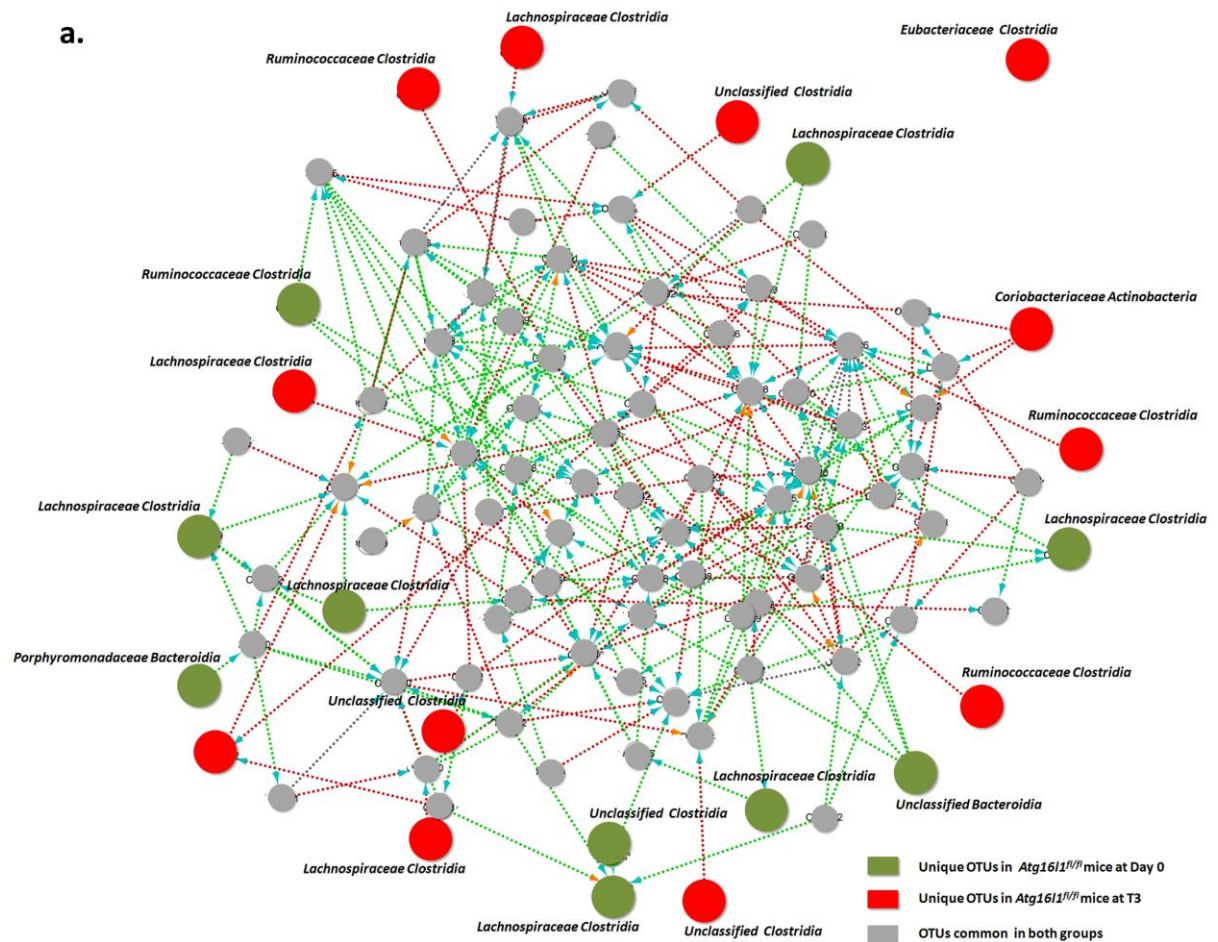


Figure S4 Metastats analysis reporting significant OTUs from pairwise comparison from baseline to T3 for both *Atg16l1*^{fl/fl} and *Atg16l1*^{ΔIEC}. (Also available as Electronic Copy – S4)



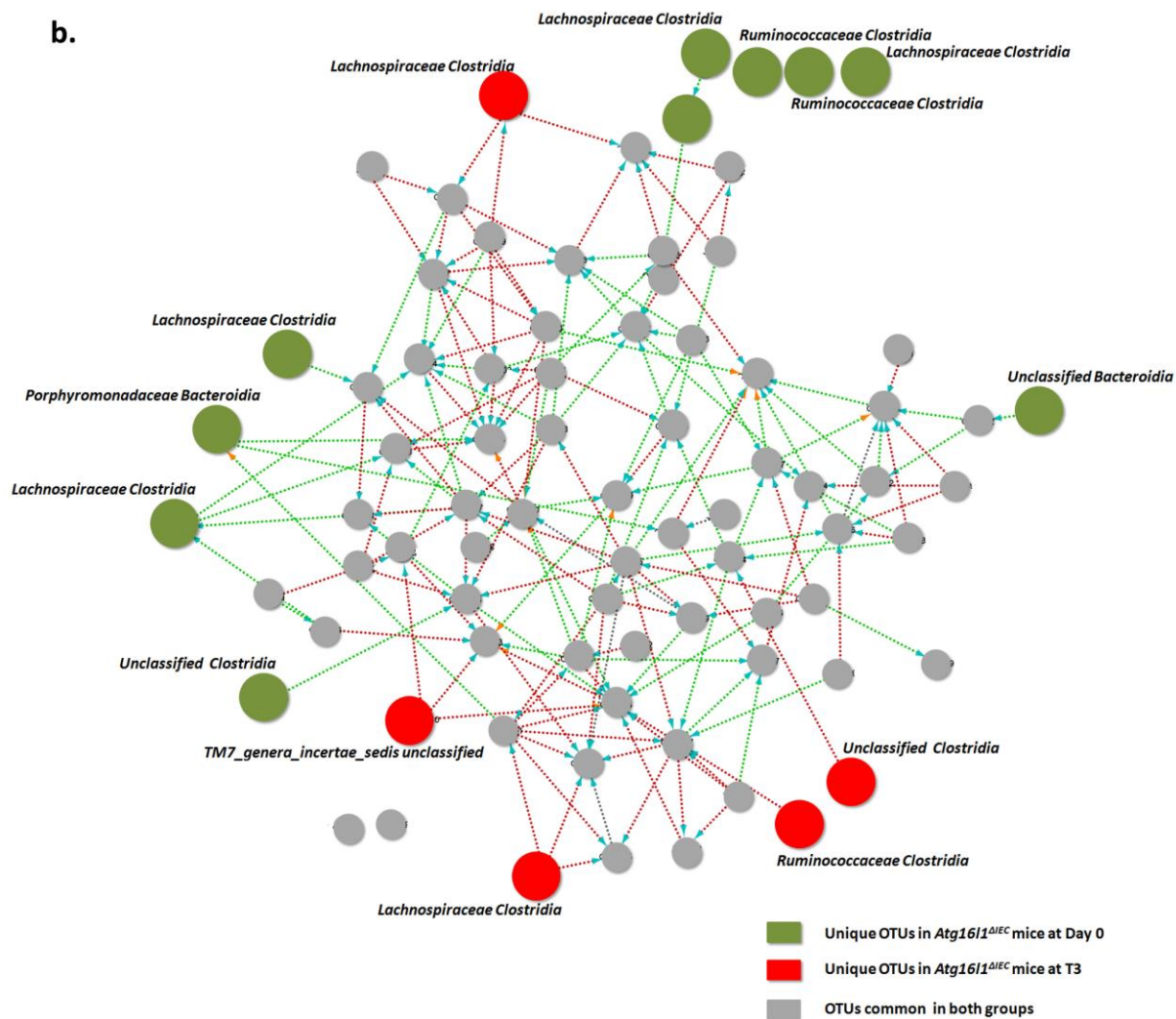


Figure S5 Network analysis showing the unique OTUs at D00 (green) and T3 (red) associated with each group are (a) *Atg16l1*^{f/f} and (b) *Atg16l1*^{ΔIEC}. The arrows represent correlation between the OTUs (Blue; positive, Red; negative). A *P*-value cutoff of 0.01 and a correlation coefficient of ± 0.6 were kept for identifying unique OTUs and generating the network respectively. **(Also available as Electronic Copy – S5)**

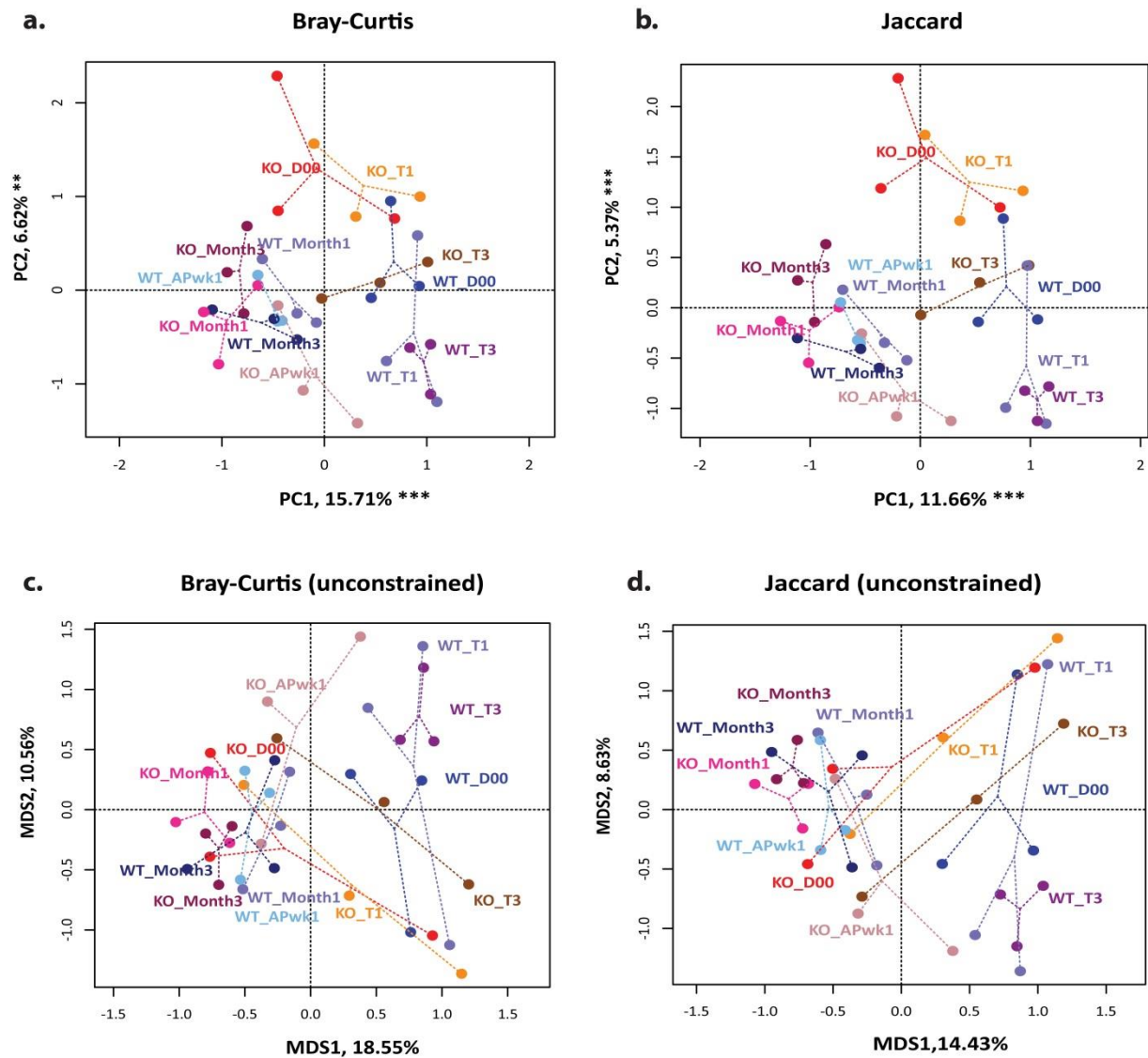


Figure S6 Temporal distance based ordination plots for and *Atg16l1^{fl/fl}/WT* and *Atg16l1^{AIEC}/KO* groups from day 0 (D00) to month 3 (M3) showing: constrained (a) Bray-Curtis (b) Jaccard and Unconstrained (c) Bray-Curtis (d) Jaccard analysis. (Also available as Electronic Copy – S6)

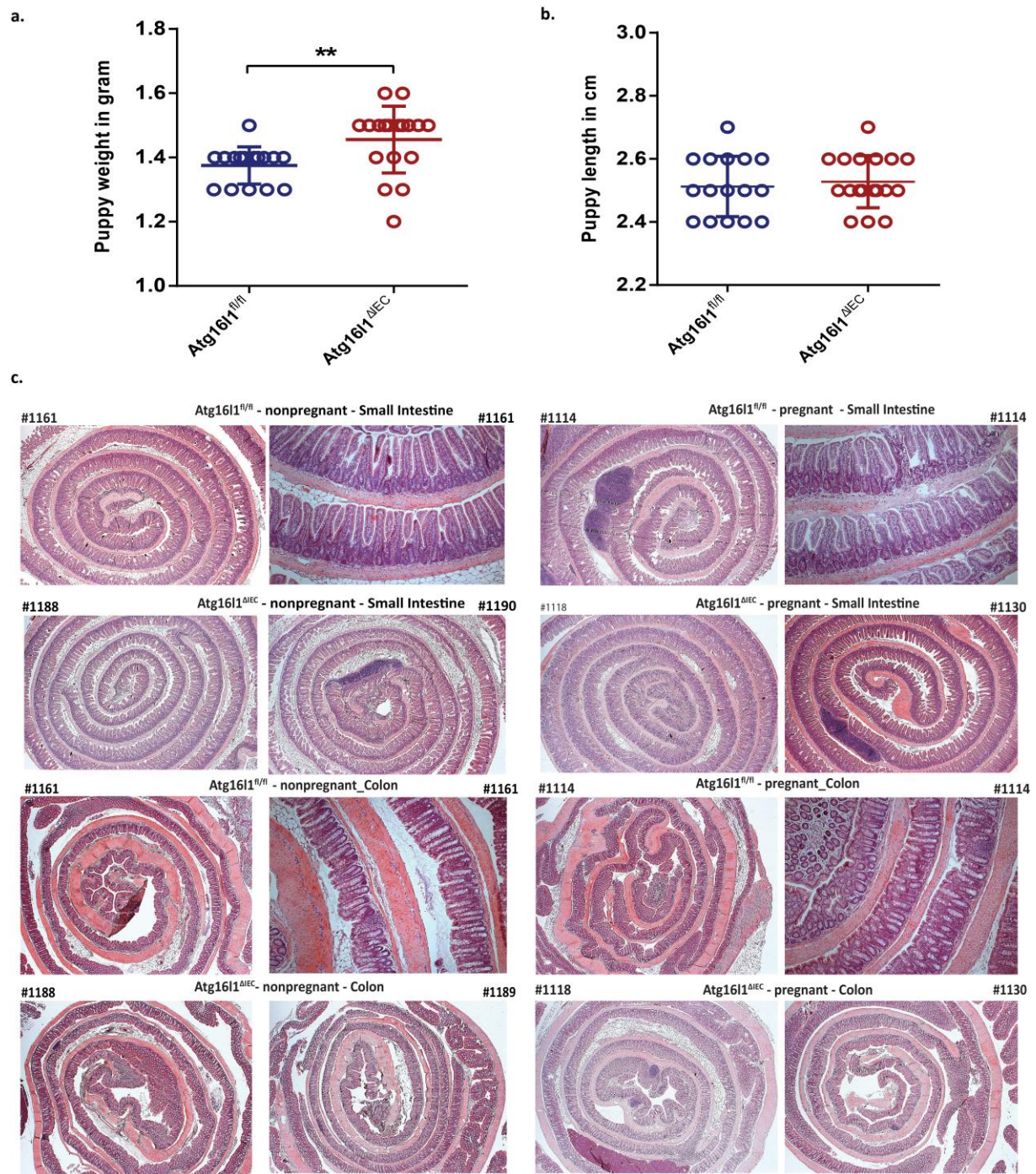


Figure S7 Histological phenotype parameters for nulliparous and pregnant *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}* mice showing (a) Puppy weight (b) Embryo length (c) histology using H&E staining for small intestine and colon. **(Also available as Electronic Copy - S7)**

Figure S8. All KEGG pathways mentioned in section 5.2.8 showing changes at D00 and T3 in *Atg16l1^{fl/fl}* and *Atg16l1^{ΔIEC}*. Color represents the number of genes mapped on pathway for each group where blue represents *Atg16l1^{fl/fl}* and red represents *Atg16l1^{ΔIEC}*. **(Also available as Electronic Copy - S8)**

11 Curriculum vitae

First and Last Name: Richa Bharti

Address: Fraunhoferstraße 4, 97076, Würzburg

Date of birth: 7th April 1986

Nationality: Indian

04/2004 – 04/2007 D.A.V College, Amritsar, India
(Bachelor in Life Sciences)
Major subjects: Botany, Zoology, Inorganic-Organic Chemistry,
English

06/2007 – 05/2009 Gurunanak Dev University (G.N.D.U), Amritsar, India
(Masters in Bioinformatics)

7/2009-5/2012 Jawaharlal Nehru University (JNU), New Delhi, India
(Research fellow)

06/2012 – 08/2016 PhD in Bioinformatics at Institute of Clinical Molecular Biology
Christian-Albrechts-University, Kiel

12 Declaration

I hereby declare that the present work - apart from the consultations through my academic teachers - in form and content is my own work. The work has not been submitted elsewhere. It has been carried out in strict accordance with the rules of good scientific practice of the *Deutsche Forschungsgesellschaft*

Erklärung

Ich erkläre hiermit, dass die vorliegende Arbeit - abgesehen von den Beratungen über meine akademische Lehrer - in Form und Inhalt meiner eigenen Arbeit ist. Die Arbeit wurde nicht an anderer Stelle eingereicht. Es wurde in strikter Übereinstimmung mit den Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgesellschaft durchgeführt.

Kiel, _____

(Richa Bharti)

13 Publications

1. Arlt A, **Bharti R** *et al.* Characteristic changes in microbial community composition and expression of innate immune genes in acute appendicitis. **Innate Immun. 2015 Jan;21(1):30-41.** (Parts of the dissertation are contained in this manuscript).
2. Tschurtschenthaler M⁺, Adolph TE*, Jonathan W. Ashcroft*, Niederreiter L*, **Bharti R*** *et al.* Defective ATG16L1-mediated removal of IRE1 α drives Crohn's disease-like ileitis **J Exp Med., 2017.** *Second author
3. Ott SJ*, Waetzig GH*, Rehman A*, Anderson JM, **Bharti R** *et al.* Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients with Clostridium difficile Infection. **Gastroenterology, 2016.**
4. Aden K...(o)...**Bharti R.** *et al.* Epithelial IL-23R Signaling Licenses Protective IL-22 Responses in Intestinal Inflammation. **Cell Rep., 2016.**
5. Adolph TE, Tomczak MF...(o)...**Bharti R** *et al.* Paneth cells as a site of origin for intestinal inflammation. **Nature. 2:10, 2013.**
6. Sommer F*, Anderson JM, **Bharti R**, Jeroen Raes and Rosenstiel P*. The resilience of the intestinal microbiota influences health and disease. **Nature microbiology review, 2017.**

Manuscripts in preparation/under revision

7. **Bharti R**, Billmann S *et al.*, ***Atg16l1* contributes to resilience of intestinal microbiota after antibiotic perturbation.** (Parts of the dissertation are contained in this manuscript and is under revision)
8. **Bharti R** *et al.*, Role of *Atg16l1* in microbial variations during pregnancy and postpartum recovery (Parts of the dissertation are contained in this manuscript in preparation)
9. Keil M, Paulsen MK, Sonner JK, **Bharti R**, Oezen I, Wanke F, Tang Y, Brando I, Reinhardt C *et al.* Dietary tryptophan is required for CNS infiltration of encephalitogenic T cells, Short Title: Dietary tryptophan deprivation blunts EAE. (In preparation).

14 Acknowledgements

The work presented in the following thesis was performed by me as a doctoral researcher at the Institute for clinical and molecular biology (ICMB) with cell biology group at Kiel University. I would like to first express my sincere thanks to **Prof. Dr. Philip Rosenstiel** for his trust in me and providing me with the opportunity to work in his group. His inputs have always led to a substantial progress at different stages of my work. I thank him for being available and pointing to the right directions when I needed it the most. He created an open environment in which our individual ideas thrived, at the same time contributing to the support system of the group. He has personally been very kind and has advised me many times in different aspects of my PhD life.

I would like thank **Prof. Alok Bhattacharya** for being my mentor for my master thesis. His suggestions and point of view for different topics have had a long lasting impression on my approach to research.

Besides my advisor, I would like to thank the rest of my thesis committee: **Prof Dr. Thomas Roeder and Prof. Dr. Hinrich Schulenburg.**

I would also thank Dr. Felix Sommer for proof reading my thesis. I further extend my thanks to Dr. Ateequr Rehman, Dr. Susanne Billman, Dr. Maren Paulsen and Dr. Anne Luzius who provided me feedback during my analysis and for performing the experiments.

A special mention to Bruhn's foundation for awarding me one year fellowship and SysmedIBD for providing me funding for my PhD and to let me pursue my research project. Also I would like to thank the research training group (RTG) for integrating me as an associate student. I also thank all my colleagues for providing a very warm and friendly working environment.

All of this has vastly helped me shape my work and thesis. I express my sincere gratitude to the whole institute of ICMB, sequencing team and all the technicians.

Last but not the least I would like to extend my heartfelt thanks to my family and friends for providing me with love and support throughout my previous years. I thank them for being there with me during my lowest phases and believing in my abilities. This would not have been possible without them.