**PROPOSITIONS**

1. The source of protein used in animal feed shapes small intestinal microbiota composition and diversity.
   (this thesis)

2. Intestinal organoids comprise heterogeneous tissue-specific cells and are more representative of *in vivo* physiology of the gut than intestinal immortalized cell lines.
   (this thesis)

3. Conducting multidisciplinary research requires stepping out of one's comfort zone.

4. Management of wildlife and ecosystems is essential to tackle the global risk of infectious diseases and the spread of antibiotic resistance.

5. The PhD journey is a balancing act, just like life.

6. If success is a recipe, communication is one of the important ingredients.

Propositions belonging to the thesis, entitled

"FeedOmics, an approach to evaluate the functional properties of protein containing feed ingredients"

Soumya Kanti Kar

Wageningen, June 19th, 2017

# FeedOmics, an approach to evaluate the functional properties of protein containing feed ingredients


Soumya Kanti Kar

# Feedomics, an approach to evaluate the functional properties of protein containing feed ingredients

Soumya K. Kar

**Thesis committee**

**Promotors**

Prof. Dr M.A. Smits
Professor of Intestinal Health of Animals
Wageningen University & Research

Prof. Dr J.M. Wells
Professor of Host-Microbe Interactomics
Wageningen University & Research

**Co-promotors**

Dr A.J.M. Jansman
Senior scientist, Animal Nutrition
Wageningen University & Research

Dr D. Schokker
Scientist, Animal Breeding & Genomics
Wageningen University & Research

**Other members**

Prof. Dr H.J. Wichers, Wageningen University & Research
Prof. Dr H. Smidt, Wageningen University & Research
Dr A.K. Kies, DSM N.V., Heerlen, The Netherlands
Dr O. Højberg, Aarhus University, Tjele, Denmark

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS)

# Feedomics, an approach to evaluate the functional properties of protein containing feed ingredients

**Soumya K. Kar**

**Thesis**

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 19 June, 2017
at 4 p.m. in the Aula.

Soumya K. Kar
FeedOmics, an approach to evaluate the functional properties of protein containing feed ingredients, 254 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)
With references, with summary in English

ISBN 978-94-6343-446-1
DOI http://dx.doi.org/10.18174/416108

# TABLE OF CONTENTS

# 1

# CHAPTER 1

## General Introduction

## Increased need of alternative protein sources for livestock

The world population is rapidly increasing and human consumption of animal proteins, is placing enormous pressure on the world plant and animal protein resources (WHO-FAO, 2003; Chadd et al., 2004; UN, 2015). Over the past 50 years, the global livestock of pigs and chickens has increased 2.5-fold and 4.5-fold, respectively, due to the increased demand for meat and poultry (FAOSTAT, 2009). This is expected to continue as livestock farming, particularly the farming of monogastrics, is one of the fastest growing agricultural sectors. This presents opportunities for economic growth, but against a backdrop of increasing societal pressure for more sustainability of farming systems (OECD-FAO, 2011).

Provision of protein in livestock is mainly based on forages, grains, legumes, animal and various industrial by-products. It was reported that in 2008, the world total compound feed production was estimated to be 680 million tonnes equating to the use of approximately 150-170 million tonnes of protein (IFIF, 2009). Poultry consumed 43% of the total compound feed produced worldwide with pigs consuming 25%, beef/sheep 15%, dairy cattle 5% and other species including fish 13% (IFIF, 2009). In 2015, the world total compound feed production was estimated to be 1 billion tonnes (IFIF, 2015). Annual report of the International Feed Industry Federation (IFIF) 2014/2015 predicts an overall 60% increase in demand for animal derived protein for human consumption in 2050 compared to 2010. According to this report, the increase in demand poised for 2050 will be met by poultry that would show an increase by 104%, aquaculture by 90%, beef by 62%, milk and dairy based products by 55% and pork by 38%.

Worldwide, in compounded livestock diet, soybean meal (SBM) dominates as the protein source that accounts for 75% of the total world tonnage of protein, followed by all other major oil meals and fish meal and animal by-products (FAO, 2004; OW, 2015). The protein ingredients used as dietary protein source in animal diet, *per se* are not pure proteins but they are rich in protein. Each year globally, livestock are estimated to consume 77 million tonnes of protein from feed that is potentially suitable for human consumption, while 58 million tonnes of protein are supplied by livestock products (Steinfeld et al., 2006). Increasing human demand for protein may shift protein sources "up the value chain" towards consumption by humans, creating a void in the protein supply intended for livestock products (milk, meat and eggs). One of the main strategies to overcome the expected shortfall of protein for livestock, is the use of alternative protein resources which partly or wholly replace those in current use (FAO, 2004; Boland et al., 2013). This has led to increased interest, particularly among different stakeholders such as animal nutritionist, researchers and feed industries, towards alternatives to conventional protein in livestock diet that can maintain or enhance both animal performance and animal

health. However, before using alternative protein sources for livestock feed, a thorough evaluation of their potential impact on host physiology, in particular metabolism and immunity will be needed. Desirable properties of alternative protein sources would be potential to favourably modulate gut microbiota by supporting the "beneficial microbes" and suppressing pathobionts (Forslund et al., 2014; Martinez et al., 2015; Toutain et al., 2016) and the ability to maintain or even increase livestock growth and performance.

However, there is a lack of comprehensive evaluation system along with a "toolbox" that can measure the impact of the protein ingredients from different sources on host's physiology in particular to metabolism and immunity in target animal. Hence, there has been an increasing investment of research efforts to investigate the effect of the use of alternative protein sources on the growth, performance and health of farm livestock, while at the same time considering the health, safety and acceptability of the resultant animal products for human consumption (Dubeski, 1994).

***Current practice in diet formulation for livestock***

In the current practice of diet formulation using linear programming, nutrient requirements of the animal for which the diet is formulated are matched with the supply of nutrients from available feed ingredients, while at the same time minimizing costs of the diet. Nutrient requirements values are mostly based on a meta-analysis of dose-response studies measuring the performance response of animals in dependence of an increasing dietary supply of a limiting nutrient. Alternatively, nutrient requirements can be based on a factorial approach considering estimated values for the nutrient requirements for maintenance purposes and for production [retention of nutrients in the body of growing animals, or in milk, eggs and progeny in other categories of animals e.g. for pigs (Hauschild et al., 2012; NRC, 2012)]. In practical diet formulation, the nutrient supply of individual feed ingredients to the final diet is supposed to be additive, meaning (digestible) nutrients of different ingredients are of equal value, independent of their origin.

Data on nutrient requirements are used as basis for feeding standards, recommendations and schedules in practice. Development of feeding standards started back in the early 19th century and knowledge had been gained gradually by the means of research and experience over many years. By mid-19$^{th}$ century, feed tables ranked by nitrogen content were available. The turn of the 20$^{th}$ century saw the beginning of research culminating in feed formulation strategies for ruminants based upon, for example, total digestible nutrients and hay, starch or corn equivalents. Table 1 summarises the brief history (from early 19th century to mid-20th century) of the development of methodologies for characterizing nutritional value of feed ingredients, of feeding standards and recommendations for livestock. Although these systems of feed evaluation underpin the present-day concepts of digestible and metabolizable energy, their values faded by the

**Table 1: Summary of the brief history* for development of feeding standards for livestock.**

| Year | Workers/Contributors | Significant contribution |
|---|---|---|
| 1810 | A. Thaer | Devised "hay equivalent". |
| 1827 | W. Prout | Recognised protein, fat and carbohydrate as essential organic nutrients. |
| 1830 | J von Liebig | Developed simple analytical method to estimate protein, fat and carbohydrate. |
| 1859 | H. Grouven | Formulated the first feeding standard with protein, carbohydrate and fat contained in the feed. |
| 1864 | W. Henneberg and F. Stohmann | Bring in the concept of digestible nutrients. |
| 1864 | E. Wolff | Devised *Wolff's Standard*: A feed standard based on digestible protein, digestible fat and digestible carbohydrates derived from the results obtained from the feeding trials. |
| 1880 | E. Wolff | The *Wolff's standards* commenced to be used in the United States of America. |
| 1884 | J. Fjord | Formulated *Scandinavian Feed Unit Standard* based on hay equivalent |
| 1890 | W. Atwater | Proposed *Feeding Standard* based on "*available feed value*" obtained by the used of Rubner's factors applied to digestible nutrients (protein: 4.1 Kcal/g; Fat: 9.3 Kcal/g; Carbohydrate: 4.1 Kcal/g) |
| 1897 | G. Lehmann | Devised *Wolff-Lehmann standards*: standards for various classes of animals. Took into account the quantity of milk produced but not the quality. |
| 1898 | W. Henry | Calculated nutritive ratio as: Digestible crude protein (DCP) = digestible carbohydrates + digestible ether extract x 2.4. Later the factor 2.4 was replaced by 2.25. |
| 1903 | T. Haecker | Published feeding standard for dairy cows showing the nutritive requirement varied not only with the quantity of milk produced but also with the quality, especially the fat content of the milk. |
| 1907 | O. Kellner | Formulated *Kellner's* "starch equivalent system". It was based upon the net energy (NE) and digestible true protein. |
| 1915 | F. Morrison | Formulated *Morrison's Feeding Standards* that indicated the nutrient requirement of animal in a range rather than in one figure. |
| 1917 | H. Armsby | Published *Armsby's Feeding Standards* based on true protein and net energy values |
| 1956 | W.A. Henry & F.B. Morrison | Revised *Morrison's Feeding Standards* by including the allowance of Ca, P, carotene, DCP, total digestible nutrients (TDN) and net energy. |

*Early 19th century to mid -20th century.

latter part of the 20th century because of the introduction of more intensive livestock production systems. In modern times, the practical goal of any feeding system is to optimize the efficiency of feed utilization, animal output and ultimately financial return to the producer. As an example, most recent efforts are being made to develop methods that determine ileal digestible nutrients like amino acids in pigs (Mosenthin et al., 2000; Pan et al., 2016) and broilers (Kidd et al., 2005; Woyengo et al., 2010).The expansion towards the development of methodologies still continues and compiling them is out of scope of this chapter.

Based on these developments in feeding standards and the methods used in evaluation of feeding values of feedstuffs, several nations across the world, including Denmark, Finland, France, Germany, Norway, Poland, Sweden, Switzerland, The Netherlands, UK and USA, assigned technical committees or councils to evaluate and publish the nutrient value of feed ingredients and nutrient requirements of different species and categories used in animal production. The remaining nations adopt or follow one of these feed evaluation systems to build their own feed tables for multiple livestock species. These tables provide reliable information on composition and quantitative nutritional value of feed ingredients either or not in combination with data on nutrient requirements.

The Netherlands publishes its own feed tables known as the CVB tables for different types of livestock animals that include ruminants, pigs and poultry (CVB, 2016). The publication provides information on the nutrient composition and feeding values for livestock animals for different types of feedstuffs *viz*. dry ingredients (frequently used in compound feed), wet ingredients and mineral sources. The following compositional data are given:

- dry matter, ash, crude protein (N*6.25), ether extract (after acid hydrolysis), structural and non-structural carbohydrates, macro- and micro-minerals, fatty acids and amino acids;
- fermentation products (in wet ingredients);
- digestibility values of nutrients;
- calculated or derived feeding values (e.g. for energy) according to the feed evaluation system for the given species of animal.

Currently in intensive livestock farming, the choice of feed ingredients in a diet is merely based on the nutrient values of the ingredient and its cost. These are therefore the "strict-nutritional" properties of feed ingredients which are required by the animal for various physiological functions including maintenance functions (processes necessary for life such as cell metabolism, synthesis and metabolism of enzymes and hormones, transport of substances around the body, maintenance of

body temperature and functioning of organs and muscles and for maintaining the immune system, excluding requirements for physical activity and for metabolic processes related to feeding) and for tissue development and enlargement in growing animals (performance).

In diet and ration formulation, however, other characteristics and potential biological effects designated in this thesis as "non-strict-nutritional" or "functional" effects) of feed ingredients are only considered to a limited extent by using minimum or maximum inclusion constraints for particular ingredients. As the research described in this thesis aimed to investigate the non-strict nutritional or functional properties of protein-containing ingredients, this chapter continues by focussing on protein containing ingredients, although the same principles are applicable to other feed ingredients. There is convincing evidence that non-strict-nutritional components in protein-containing feed ingredients possess important functional properties. Detailed background information on these functional properties is

**Strict Nutritional properties of feed ingredients**

Characteristics of the nutritional value of feed ingredients take into account the species, sex, age, developmental stage, particular production targets and environmental conditions of the animal. The information on the quantitative nutritive value of various ingredients and nutrient requirements of farm animals are available in different tables across the globe (CVB, 2011; NRC, 2012). For protein sources, the nutritional value is in particular characterised by their capacity to provide digestible amino acids. Digestible amino acids (AA) can be used for protein synthesis in farm animals which is related to growth, health, and reproduction.

**Non-Strict-nutritional properties of feed ingredients**

Apart from the nutritional characteristics, feed ingredients possess other functional properties related to their potential effects on health and physiology of the animal. They refer to the "non-strict-nutritional" properties of ingredients. In the context of protein sources for animal feeds, the functional values have been often associated with negative attributes, such as the presence of anti-nutritional factors. Protein-containing feed ingredients, however, have also the ability to release functional biochemical moieties such as bioactive peptides, during the digestion process in gastrointestinal tract (GIT). The released biochemical moieties impart their functional effects in tissue and organ functioning in relation to gut and animal health. In addition, the protein-containing feed ingredients can provide specific AAs such as non-protein $\alpha$-AA (e.g., ornithine, citrulline, and homocysteine) and non-$\alpha$ AA (e.g., taurine and $\beta$-alanine) that play important roles in nutrition and metabolism (Wu, 2009). This strongly suggests that protein sources also have functional properties that can be judged as beneficial attributes.

provided in the section "functional properties of protein-containing feed ingredients", below.

***Effect of processing and digestion on protein sources***

Further, processing of protein-containing ingredients or the resulting diet can cause changes to the organic compounds, in particular to proteins that lead to chemical changes of amino acid residues, that may impact the nutritional value of the protein fraction in the protein source (Gonzalez-Vega et al., 2011; Gerrard et al., 2012; Almeida et al., 2014; Eklund et al., 2015; Hulshof et al., 2016). A significant part of the undigested or unabsorbed protein/amino acids passing the small intestine can be fermented in the large intestine by the residing microbiota (Jha and Berrocoso, 2016). Some of the fermentation breakdown products (e.g. ammonia and amines) can affect the development of the intestinal epithelial lining in the hind gut. In addition, dietary protein reaching the large intestine, can support the growth of harmful bacteria thereby causing enteropathy that may ultimately lead to diarrhoea (Williams et al., 2001; Louis et al., 2007).

***Alternative protein source***

Inclusion of specific protein sources in the diet may support and enhance the development and functioning of the immune system, thereby reducing morbidity and mortality (Liu et al., 2008; Tan et al., 2009; Ewaschuk et al., 2011; Ren et al., 2012). In recent years, however, efforts are being already made to search for and use sustainable food/feed resources with a low environmental footprint such as sea weed (Walsh et al., 2013; Heim et al., 2014), microbial protein (Matassa et al., 2016) and insect proteins (van Huis, 2013; Schanes et al., 2016; Tabassum et al., 2016) as alternative protein sources. In addition, some feed ingredients have shown to positively or negatively impact on the taste and flavour of livestock products (Melton, 1990; Eyng et al., 2013) affecting consumer behaviour and perception about the livestock products (Troy and Kerry, 2010; Font-I-Furnols and Guerrero, 2014). The overview of non-strict-nutritional properties of feed ingredients as presented in Figure 1 does not intend to be complete but provides an illustration of some of the non-strict nutritional effects that are relevant for the nutrition domain.

***Functional properties of protein-containing feed ingredients***

Currently it is the negative anti-nutritional attributes of ingredients that are being considered in relation to their non-strict nutritional properties, rather than any positive aspects. Information of anti-nutritional factors are available in several newly developed feed resources information system, maintained and regularly updated by various agencies (Feedipedia, 2016; FAO, 2017). In current practice, animal nutritionists use this information to put limits on the amounts of anti-nutritional-factor-containing feed ingredients used in livestock diet.

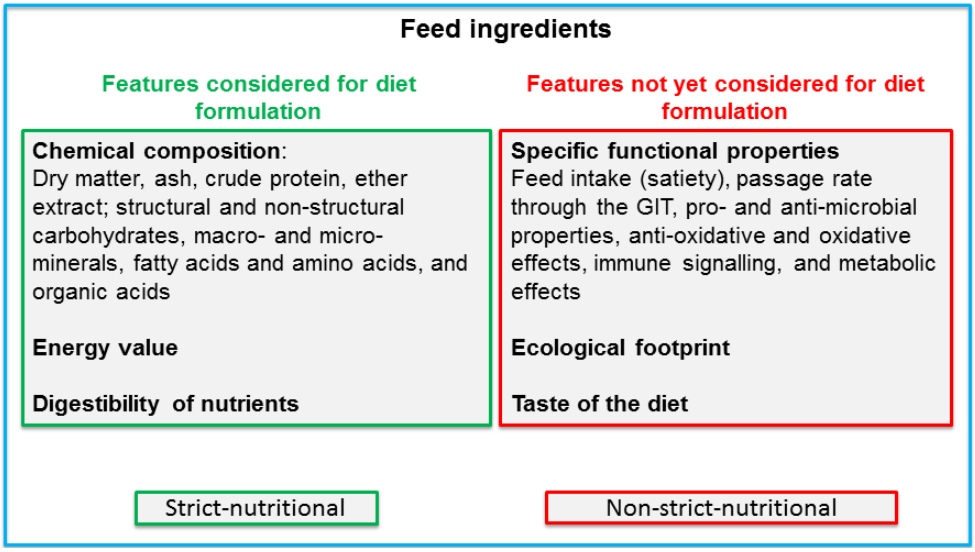



**Figure 1: Important features of feed ingredients.**
The blue outlined box represents feed ingredients and the green outlined box represents the strict nutritional components which are currently being considered for formulation of livestock diet. The non-strict components in the red outlined box are not yet considered for formulation of livestock diet.

The functional properties of a protein-containing feed ingredient is defined in this thesis as the potential of the ingested protein-containing feed ingredient to display bioactivity with impact on the performance and health of the animal. Such functional properties of feed ingredients can be related to feed intake (satiety), passage rate through the gut, pro- and anti-microbial properties, anti-oxidative and oxidative effects, immunity and metabolic effects (Jansman, 2016). These functional properties can arise from an inherent component of the protein-, carbohydrate- (including starch and fibre), fat-, vitamin-, mineral-fraction associated with the ingredient (McIntosh et al., 1998; Borchani et al., 2012; Christaki et al., 2013). The definition of functional properties as explained in this thesis should not be confused with the physico-chemical properties such as water binding capacity, solubility , dispersibility , viscosity or reactivity of feed ingredients that influence processing properties during feed manufacturing (Hermansson, 1979; Martinez, 1979; Howling, 1980; Wood, 1987).

Among functional properties, immunomodulatory effects have received special attention by the scientific community as it is assumed to have direct consequences for immunity and health. Several animal-based protein sources, particularly the dairy-based products (casein) and by-products (whey powder) were evaluated for immunomodulatory proteins and peptides (see Table 2). *In vitro* experiments and feeding trials in laboratory

**Table 2. Immunomodulatory proteins and peptides.**

| Protein/peptide | Effect | Model | References |
|---|---|---|---|
| Caseins (and digests) | T-lymphocyte proliferation ↑ | Cell culture | (Carr et al., 1990; Kayser and Meisel, 1996; Hata et al., 1998) |
| | T-lymphocyte proliferation ↓ | Cell culture | (Otani et al., 1992; Kayser and Meisel, 1996) |
| | Immunoglobulin secretion↑ | Cell culture | (Hata et al., 1998) |
| Whey | Lymphocyte blastogenesis↓ | Cell culture | (Barta et al., 1991) |
| Milk | IgG Antibody secretion ↓ | Cell culture | (Playford et al., 2000) |
| Lactoperoxidase | T-cell mitogenesis ↓ | Cell culture | (Wong et al., 1997) |
| Lactoferrin | Cytokine release↓ | Cell culture | (Crouch et al., 1992) |
| | Mammary gland mononuclear cell proliferation ↑ | Cell culture | (Rejman et al., 1993) |
| Prolin-rich polypeptides (and derivates) | B-lymphocyte growth, differentiation ↑ | Cell culture | (Julius et al., 1988) |
| Tuftsin | Leucocyte chemotaxis ↑ | Animal | (Janusz et al., 1987) |
| Fish protein | IgA, IL-4-, IL-6-, IL-10-positive cells↑ | Fish | (Duarte et al., 2006) |

(↑/↓: Increase/decrease)

animals with purified diet were conducted to understand the mechanisms induced by protein sources that possess immunomodulatory properties (Table 2).

Besides the immunomodulatory properties of dairy based products, they are also known to have the angiotensin I-converting enzyme (ACE)– inhibitory peptides that have hypotensive effects on the cardio-vascular system (Yamamoto et al., 1994; Maeno et al., 1996; Mullally et al., 1997; Pihlanto-Leppala et al., 2000). In addition dairy based product contains lactoferrin , a peptide that has anti-oxidative properties and antimicrobial activity through its chelation of iron (Lindmark-Mansson and Akesson, 2000; Shinmoto et al., 2014). Spray dried plasma protein (SDPP), a by-product derived from slaughter houses, has been recommended in animal diet as a source of immunological support due to their high levels of globulin proteins, including immunoglobulins (Campbell et al., 2010; Perez-Bosque et al., 2010; Gao et al., 2011) which could bind pathogens. There is also quite some evidence that plant based protein sources, in particular those derived from soybean, have functional properties. Soy β-conglycinin (7S globulin), a protein

found in soybean has shown to ameliorate atherosclerosis in mice (Adams et al., 2004). In addition, both α- and β-subunits of β-conglycinin have proven anti-inflammatory effects (Burris et al., 2014). Soybean protein has also been reported to be hypotriglycerdemic in rats, having cholesterol lowering effects and reducing fatty liver (Ascencio et al., 2004; Lin et al., 2004; Moriyama et al., 2004; Xiao et al., 2004). The hypotriglycerdemic effect of soy protein was shown to be due to suppression of retinoic acid receptor expression in liver of rats fed soy bean protein but not casein protein in their diet (Standeven et al., 1996; Radcliffe et al., 1998; Xiao et al., 2007). Additionally, soy-fermented food digested with proteolytic enzymes contain bioactive peptides with ACE inhibition and antioxidative properties (Gibbs et al., 2004).

The supply of amino acids to the host, results from several interdependent digestive processes: the sequential degradation of dietary proteins during their passage in the gastro-intestinal tract (GIT) and the absorption of their hydrolysis products. Digestive processes are influenced by length of stay and transit of digesta throughout the GIT. Peptides of microbial origin or partly hydrolyzed bacterial protein also contribute in supplying amino acids to the host (Darcy, 1984; Metges, 2000; Morowitz et al., 2011; Neis et al., 2015). Several of these peptides function as bioactive components that may exert biological functions locally (e.g., in the gut) or systemically (i.e., via the bloodstream). The so-called biologically active motifs in polypeptide chains remain inactive as long as they ''reside'' in their precursor forms, however, upon release by proteolytic enzymes, they may interact with receptors and exert bioactivity (Lindmark-Mansson and Akesson, 2000). Numerous bioactivities have been described for peptides released from dietary proteins by enzymatic proteolysis including immunomodulating and ileum contracting properties (Dewey et al., 2006; Moller et al., 2008; Antunes et al., 2011; Walther and Sieber, 2011; Caetano Faria et al., 2013).

According to Shimizu (2004), functional proteins and peptides can be divided into three categories according to the locations where their functions are expressed. Proteins and peptides from category 1 express their functions in the lumen of the GIT, including modulating the secretion of digestive enzymes and the absorption of nutrients and inhibiting the growth of pathogens. For example, phosphopeptides derived from caseins can bind to calcium and form a soluble complex, which increases the absorption of calcium (Sato et al., 1986). Furthermore, the peptides lactoferricin B and lactoferricin H, originating from milk lactoferrin after pepsin digestion, have antimicrobial properties (Bellamy et al., 1992). Category 2, proteins and peptides express their functions in the body, i.e. after their absorption by the GIT. They are capable of, for instance, modulating the activity of the immune system and the nervous system. Peptides from bovine milk hydrolysates can stimulate the proliferation and activity of leukocytes and increase the production of antibodies (Shahidi and Zhong, 2008). In

addition to milk, immunomodulatory peptides are also identified in other food and feed ingredients (Walther and Sieber, 2011). A bioactive peptide, with the ability to stimulate phagocytosis, could be isolated from soybean protein after trypsin digestion (Maruyama et al., 2003). Proteins and peptides from category 3 express their functions directly on epithelial cells of the GIT. Hashimoto et al. (1995) suggested that β-lactoglobulin from bovine milk may stabilize the tight junction of intestinal epithelial cells, thereby inhibiting the paracellular diffusion of compounds with high molecular weight.

Peptides formed during protein hydrolysis may also have negative effects on intestinal functioning an example being the formation of toxic protein metabolites such as ammonia, amines, volatile phenols and indole during bacterial protein fermentation. These potentially toxic metabolites may disturb the delicate balance between beneficial and pathogenic bacteria in the GIT, negatively impacting on symbiosis, gut health and animal performance (Williams et al., 2001). It is also known that peptides formed during hydrolysis are prone to aggregation processes that lead to indigestible peptide aggregates, a phenomenon that was also observed for soy-derived peptides in the intestine of pigs (Fischer et al., 2007). Formation of aggregated peptides is dependent on the nature of the protein source, the degree of processing on the physico-chemical conditions and enzymatic, microbial activity in the digestive tract.

Such anti-nutritional factors receive most attention in formulation of animal feed and potential beneficial functions aside from the nutrition generally receives less attention. This may in part be due to the lack of assessment methods for functional properties of protein sources beyond their capacity to provide (essential) amino acids and other nutrients (Jahan-Mihan et al., 2011).

***Interactions in the gut***

In mammals, the intestinal mucosa and mucosal-associated lymphoid tissue contains many specialised cells including the largest repertoire of immune cells (Figure 2). This includes epithelial, secretory, endocrine and specialized immune cells which are specifically involved in sensing signals from their environment and regulate (activation, repression) local immune mechanisms, mainly active in, but not restricted to, the mucosa of the GIT. Epithelial cells collaborate with resident and infiltrating immune cells to respond to constituents in the lumen of the gut (digestion products of feed/food, host-derived molecules, pathogenic/non-pathogenic micro-organisms and toxins). Along with the epithelial cells that express Toll-like receptors (TLR) and a network of mucosal dendritic cells (DC's) in the lamina propria of the intestine act as sentinels for antigens or microbes which cross the epithelium and participate in innate as well as adaptive defence mechanisms.

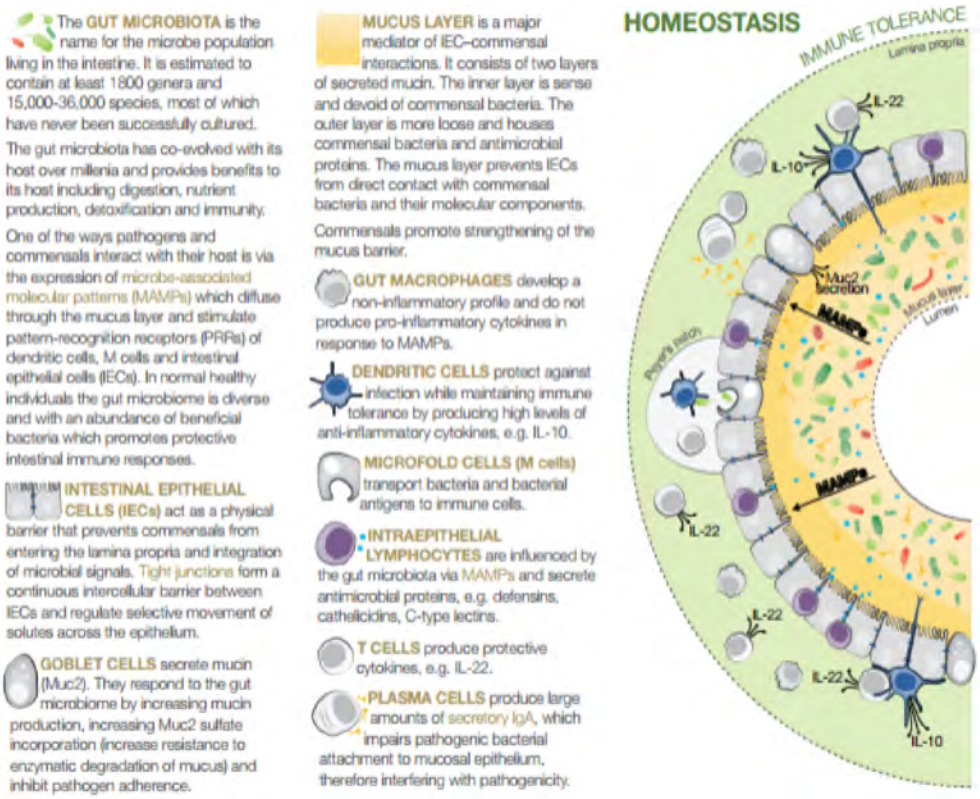



**Figure 2: Components and function of intestinal host-microbe interaction.**
Source: Lee (2016)

The GIT of mammals is colonised by microbes that constitute a vast ecosystem known as the gut microbiota. The intestinal microbiota is a complex ecosystem which increases in numbers throughout the length of GI tract, from $10^1$-$10^4$ cells per ml of intestinal content in the stomach, reaching a density of $10^{10}$-$10^{12}$ cells per gram of intestinal content in the large intestine (Booijink et al., 2007; Hugenholtz, 2015). Although it is well known that gut microbiota is not homogeneously distributed within the GIT, it is still largely unknown how the diversity and function varies in the different niches along the GIT. Current microbiome studies in humans have associated changes in composition and diversity of the gut microbiota with several diseases (Dicksved et al., 2008; Konig et al., 2016), including obesity (Ley, 2010), diabetes (Larsen et al., 2010) and allergies (Round and Mazmanian, 2009). Many studies have suggested a role of gut microbiota and diet in host metabolism, growth and overall health (Heinritz et al., 2013; Xiao et al., 2015).

Beside maintaining immune resilience (Belkaid and Tarbell, 2009; Spencer and Belkaid, 2012), the residing microbiota in the intestine contributes to the digestive process, thereby maximizing the nutrient delivery from the diet to the host. Small-molecule

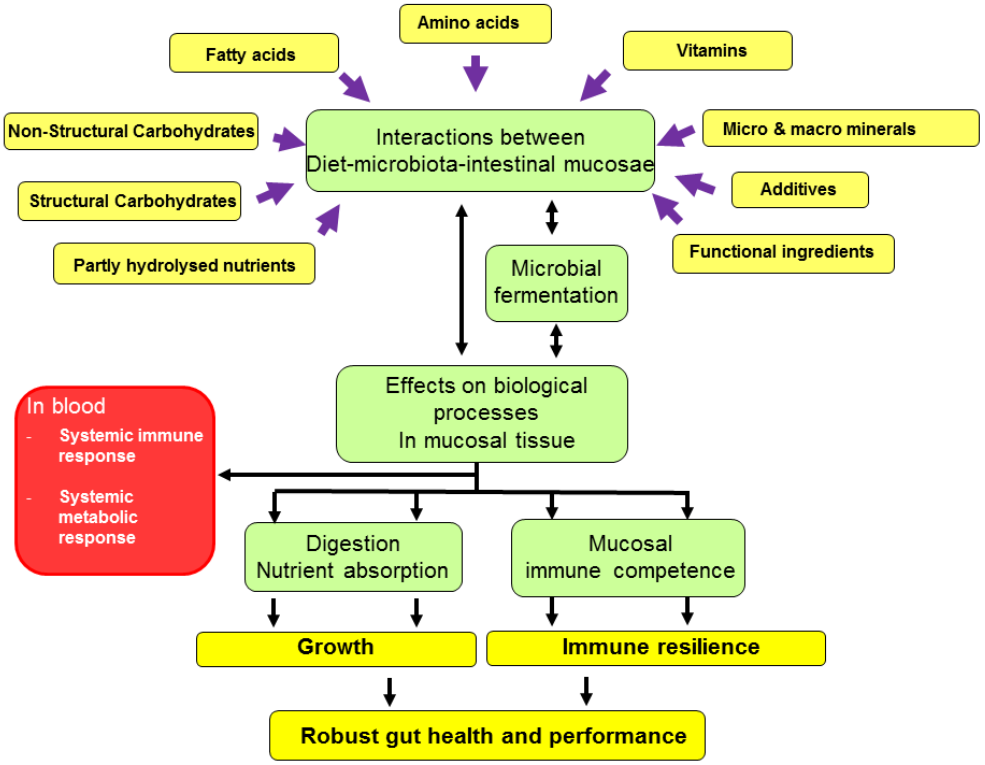



**Figure 3: Schematic overview of the consequence of interaction with nutrients and other constituents in animal feeds, intestinal microbes and mucosae.**

metabolites of dietary, host and microbial origin, including (modified) amino acids and peptides show bioactive effects on intestinal epithelium (Antunes et al., 2011) and are likely to have important functions in maintaining immune homeostasis and digestive efficiency. Nutrients and other constituents in animal feeds, microbes and host epithelial cells interact strongly with each other (Figure 3).

These interactions affect, directly or indirectly, a number of processes related to digestion, fermentation, nutrient absorption, nutrient metabolism, intestinal integrity, satiety, immune recognition, immune regulation and development of immune tolerance and immunity. Disturbance of this system often results in excessive inflammatory reactions that leads to enteropathy. Protracted inflammation can cause serious damage to the mucosal layer resulting in loss of its barrier function. Human intestinal diseases like Crohn's disease and inflammatory bowel disease (IBD) are characterised by a chronic state of inflammation of the small and/or large intestine, causing abdominal pain and diarrhoea and consequently, lead to malabsorption of nutrients. Typically, this is associated with

dysbiosis (microbial imbalance) of the intestinal microbiota (Konig et al., 2016). Parallel research is ongoing to investigate the role of intestinal microbiota in animal (both livestock and companion species) health and disease but also in animal performance and feed efficiency. These studies are focussed on a better understanding of the composition, diversity and function of the residing microbial communities, their mutual interactions and their interactions with the host and the metabolites produced and utilised by microbes. The majority of studies focus on the identification of differences between high performing healthy animals and diseased and/or low performing animals. In livestock, disturbances of immune homeostasis in the gut often results in growth retardation, a lower feed efficiency and an increased susceptibility to infections, the latter with consequences for the use of antibiotics (Looft et al., 2014a; Looft et al., 2014b; Schokker et al., 2015). Therefore, recent animal nutrition research is directed towards the development of nutritional strategies that support immune and digestive system development, stabilise the microbiota, induce immune resilience, maintain intestinal integrity and, at the same time, fulfil the nutrient requirements for maintenance and growth.

Owing to the release of protein digestive enzymes, the small intestine is an important site of protein digestion. The small intestine is considered the first region of GIT where the protein fraction from the diet comes into contact with the residing microbiota. This suggests that the interactions between diet, microbiota and host in the small intestine contribute in maintaining and contribute to a stable diversity and abundance of the microbiota (Lozupone et al., 2012), an efficient digestibility (Krajmalnik-Brown et al., 2012) and an optimal immune resilience (Belkaid and Hand, 2014). However, the contribution of dietary components in the maintenance of small intestinal health is not well understood. Moreover, the inaccessibility of collecting proper small intestinal samples results in limited information in literature with regard to diet-microbiota-host interaction in the small intestine.

***Rationale and approach undertaken in this thesis***

Presently, there is limited knowledge about the mechanisms involved in the interaction of protein source (as diet), microbial communities and host physiology in the complex environment of the GIT. This hinders the knowledge-based inclusion of alternative functionally active natural protein sources to improve intestinal health in animals. To fill the knowledge gap, this thesis studies the functional properties of protein-containing feed ingredients with a focus on comparing traditional protein sources with new alternatives.

A unique approach, combining proteomics and bioinformatics (*in silico*), an intestinal organoid model (*in vitro*) and animal models, including mice and pigs (*in vivo*), was used to evaluate the functional properties of protein sources. Genomics, transcriptomics,

**Genomics**: The branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes.

**Transcriptomics**: The study of transcriptomes (mRNA) and their functions. This allows the examination of whole transcriptome changes across a variety of biological conditions.

**Proteomics**: The study of proteomes and their functions.

**Metabolomics**: The study of metabolite profile present within an organism, cell, or tissue. This can be exploited for the early detection of disease or phenotypic trait.

proteomics and metabolomics techniques are applied to identify changes in a variety of molecular parameters due to the use of different protein sources in animal diet. Subsequently, bioinformatic approaches are used to get insight into the mechanisms imparted by the functional properties of protein sources. Finally, through a systems biology approach, multi-omics data obtained from the experiment with mice are integrated to establish relationships between various biological scales. A schematic representation of the approaches and the Chapters of this thesis is shown in Figure 4.

### Aim and Outline of the thesis

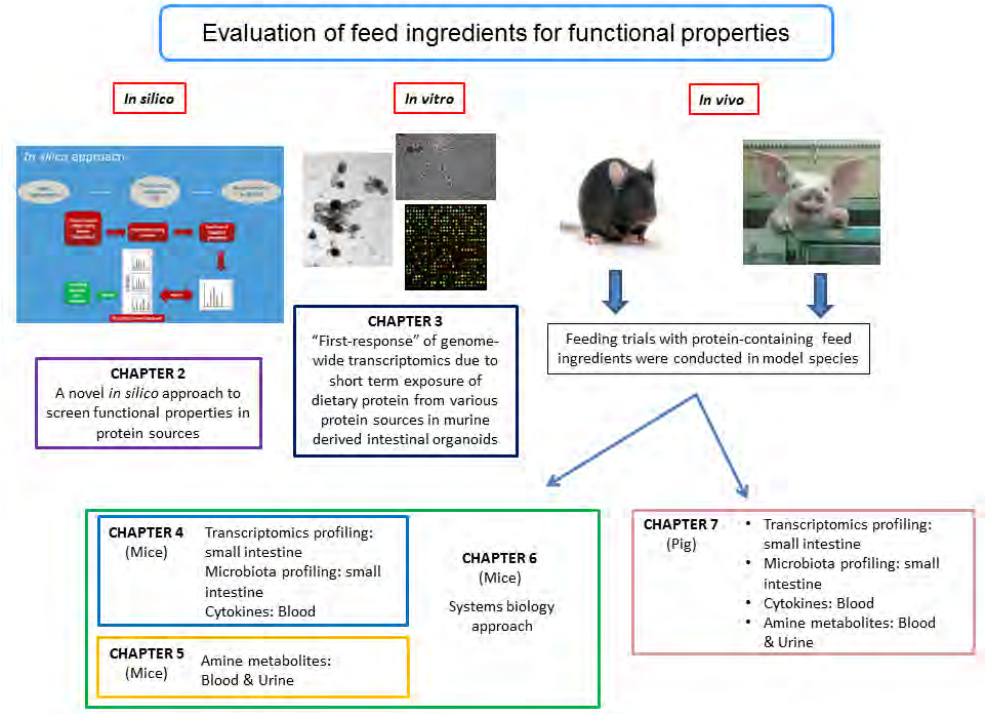



**Figure 4: Schematic overview of the approach and the chapters of this thesis.**

The ultimate goal of this research is to identify protein sources that can replace the protein sources that are currently used for the formulation of animal diet. However, before alternative protein sources can be used, a thorough evaluation of their potential impact on the physiology of the animal is required, in particular with regard to their effects on metabolism and immunity. Thus the research objective of this thesis is to evaluate the functional properties of a variety of different protein containing feed ingredients. For evaluation we used multiple –omics technologies (FeedOmics approach) to improve our understanding of the gut associated changes induced by the dietary protein sources, especially on the level of the intestinal microbiota and the mucosal gene expression. In addition, we also investigated their effects on several systemic immune and metabolic parameters. The first specific aim of the thesis research was to characterise in detail the protein component of existing and potential novel protein sources to be included in feeds for monogastric animals; and in addition to demonstrate the added value of the use of proteomic and bioinformatic procedures to approximate the protein, peptide and amino acid composition as well as to predict the potential bioactive properties of the (novel) protein-containing feed ingredients (**Chapter 2).** This Chapter describes the evaluation of six different protein sources and the bio-functionalities of their *in vivo* digestion products are predicted, using advanced proteomic and bioinformatic approaches.

Given that novel protein sources may contain bioactive components with activities beyond their strictly nutritional properties, in **Chapter 3** we have investigated the use of intestinal organoids as an *in vitro* model to test the functional properties of different protein sources on the intestinal mucosa. We use a genome-wide transcriptome analysis to measure the short-term (relative to *in vivo* experiment performed in this thesis) response of two-dimensional grown murine derived small intestinal organoids upon exposure to different protein sources.

Further, an experiment was conducted with laboratory animal *viz.* mice by feeding them with diet containing protein from various sources (same sources as used in Chapter 2). The third specific aim of the thesis was to understand the long term (relative to *in vitro* experiment performed in Chapter 3) *in vivo* effect of protein sources on intestinal functioning and health in mice. **Chapter 4** describes the effects of dietary protein sources in mice on both local intestinal (immune) gene expression and microbial colonization as well as systemic immune responses. Here, we specifically focus on the expression of cellular signalling systems that connect the availability of nutrients to the growth of cells involved in the renewal of intestinal epithelium and the proliferation of mucosal immune cells. Measuring amine metabolites in serum or urine is considered a useful approach to assess the host and microbiota metabolism of proteins in the diet. On one hand, a profile of blood amine metabolites could explain the efficiency of nutrient utilization from the diet. On the other hand this profile may explain the metabolic and absorptive capacity

of the gut, including its microbiome (Kogut and Arsenault, 2016). **Chapter 5** describes the effect of dietary protein sources on metabolism and metabolic amine profiles in serum and urine of mice The fourth aim of the thesis was to use a set of computational methods generally being utilized in the systems biology for understanding the effect of protein sources with regard to its functional properties. **Chapter 6** describes potential correlations relationships between molecular parameters as measures on various biological scales. This is established by successfully integrating multi-scale quantitative (~omics) data, originating from the mice experiment.

Final evaluation of protein sources was carried out in a feeding trial with pigs. The fifth aim of the thesis was to understand the long term (relative to *in vitro* experiment performed in this thesis) *in vivo* effect of protein sources on intestinal functioning and health in the target animal pigs. **Chapter 7** describes the effects of diet prepared with different protein sources. In this Chapter we investigate whether these diets differ in their ability to modulate host physiology by analysing and measuring three different molecular level parameters as an intestinal functioning and health indicator. The same readout parameters as used in Chapters 4 and 5 are used.

Finally, **Chapter 8** summarises the research undertaken in this thesis and discusses the results in a broader context. This Chapters describes how the research of the thesis contributes to the improvement of scientific knowledge related to the functionalities of the GIT. It also describes methods to evaluate the functional properties of protein-containing feed ingredients by using (but not limited to) modern day (~omics) technologies and novel *in vitro* (intestinal models). Finally, I discuss a new potential application of using this knowledge generated in this thesis in the area of animal nutrition.

## References


1. Adams, M. R., D. L. Golden, A. A. Franke, S. M. Potter, H. S. Smith, and M. S. Anthony. 2004. Dietary soy beta-conglycinin (7S globulin) inhibits atherosclerosis in mice. J. Nutr. 134: 511-516.
2. Almeida, F. N., J. K. Htoo, J. Thomson, and H. H. Stein. 2014. Effects of heat treatment on the apparent and standardized ileal digestibility of amino acids in canola meal fed to growing pigs. Animal Feed Science and Technology 187: 44-52. doi:10.1016/j.anifeedsci.2013.09.009
3. Antunes, L. C. M., J. E. Davies, and B. B. Finlay. 2011. Chemical signaling in the gastrointestinal tract. F1000 biology reports 3: 4-4. doi:10.3410/b3-4
4. Ascencio, C., N. Torres, F. Isoard-Acosta, F. J. Gomez-Perez, R. Hernandez-Pando, and A. R. Tovar. 2004. Soy protein affects serum insulin and hepatic SREBP-1 mRNA and reduces fatty liver in rats. J Nutr 134: 522-529.
5. Barta, O., V. D. Barta, M. V. Crisman, and R. M. Akers. 1991. Inhibition of lymphocyte blastogenesis by whey. Am. J. Vet. Res. 52: 247-253.
6. Belkaid, Y., and T. Hand. 2014. Role of the Microbiota in Immunity and inflammation. Cell 157: 121-141.
7. Belkaid, Y., and K. Tarbell. 2009. Regulatory T cells in the control of host-microorganism interactions (*). Annual review of immunology 27: 551-589. doi:10.1146/annurev.immunol.021908.132723
8. Bellamy, W., M. Takase, K. Yamauchi, H. Wakabayashi, K. Kawase, and M. Tomita. 1992. Identification of the bactericidal domain of lactoferrin. Biochimica et biophysica acta 1121: 130-136. doi:0167-4838(92)90346-F [pii]
9. Boland, M. J., A. N. Rae, J. M. Vereijken, M. P. M. Meuwissen, A. R. H. Fischer, M. van Boekel, S. M. Rutherfurd, H. Gruppen, P. J. Moughan, and W. H. Hendriks. 2013. The future supply of animal-derived protein for human consumption. Trends in Food Science & Technology 29: 62-73. doi:10.1016/j.tifs.2012.07.002
10. Booijink, C., E. G. Zoetendal, M. Kleerebezem, and W. M. de Vos. 2007. Microbial communities in the human small intestine: coupling diversity to metagenomics. Future Microbiology 2: 285-295. doi:10.2217/17460913.2.3.285
11. Borchani, C., S. Besbes, M. Masmoudi, M. A. Bouaziz, C. Blecker, and H. Attia. 2012. Influence of Oven-Drying Temperature on Physicochemical and Functional Properties of Date Fibre Concentrates. Food and Bioprocess Technology 5: 1541-1551. doi:10.1007/s11947-011-0549-z
12. Burris, R. L., H. P. Ng, and S. Nagarajan. 2014. Soy protein inhibits inflammation-induced VCAM-1 and inflammatory cytokine induction by inhibiting the NF-kappaB and AKT signaling pathway in apolipoprotein E-deficient mice. Eur J Nutr 53: 135-148. doi:10.1007/s00394-013-0509-7

13. Caetano Faria, A. M., A. C. Gomes-Santos, J. L. Goncalves, T. G. Moreira, S. R. Medeiros, L. P. Dourado, and D. C. Cara. 2013. Food components and the immune system: from tonic agents to allergens. Frontiers in Immunology 4. doi:10.3389/fimmu.2013.00102
14. Campbell, J. M., J. Polo, L. E. Russell, and J. D. Crenshaw. 2010. Review of spray-dried plasma's impact on intestinal barrier function. Livest Sci 133: 239-241. doi:10.1016/j.livsci.2010.06.075
15. Carr, R. I., D. Webster, D. Sadi, H. Williams, and N. Walsh. 1990. Immunomodulation by Opioids from Dietary Casein (Exorphins). Ann. N. Y. Acad. Sci. 594: 374-376. doi:10.1111/j.1749-6632.1990.tb40500.x
16. Chadd, S. A., W. P. Davies, and J. M. Koivisto. 2004. Practical production of protein for food animals Protein sources for animal feed industries. p 77-124. FAO of UN, Rome, Italy.
17. Christaki, E., E. Bonos, I. Giannenas, and P. Florou-Paneri. 2013. Functional properties of carotenoids originating from algae. Journal of the Science of Food and Agriculture 93: 5-11. doi:10.1002/jsfa.5902
18. Crouch, S. P. M., K. J. Slater, and J. Fletcher. 1992. Regulation of cytokine release from mononuclear-cells by the iron-binding protein lactoferrin. Blood 80: 235-240.
19. CVB. 2011. Chemical compositions and nutritional values of feed ingredients., Centraal Veevoeder Bureau, Lelystad, the Netherlands.
20. Darcy, B. 1984. Availability of amino-acids in monogastric animals - variations of digestive origin. Diabetes & Metabolism 10: 121-133.
21. Dewey, C. E., W. T. Johnston, L. Gould, and T. L. Whiting. 2006. Postweaning mortality in Manitoba swine. Can J Vet Res 70: 161-167.
22. Dicksved, J., J. Halfvarson, M. Rosenquist, G. Jarnerot, C. Tysk, J. Apajalahti, L. Engstrand, and J. K. Jansson. 2008. Molecular analysis of the gut microbiota of identical twins with Crohn's disease. Isme Journal 2: 716-727. doi:10.1038/ismej.2008.37
23. Duarte, J., G. Vinderola, B. Ritz, G. Perdigon, and C. Matar. 2006. Immunomodulating capacity of commercial fish protein hydrolysate for diet supplementation. Immunobiology 211: 341-350. doi:10.1016/j.imbio.2005.12.002
24. Dubeski, P. L. 1994. The living gut: An introduction to micro-organisms in nutrition. Animal Feed Science and Technology 68: 190-192. doi:10.1016/s0377-8401(97)00019-9
25. Eklund, M., N. Sauer, F. Schone, U. Messerschmidt, P. Rosenfelder, J. K. Htoo, and R. Mosenthin. 2015. Effect of processing of rapeseed under defined conditions in a pilot plant on chemical composition and standardized ileal amino acid digestibility in rapeseed meal for pigs. J Anim Sci 93: 2813-2825. doi:10.2527/jas.2014-8210
26. Ewaschuk, J. B., G. K. Murdoch, I. R. Johnson, K. L. Madsen, and C. J. Field. 2011. Glutamine supplementation improves intestinal barrier function in a weaned piglet model of Escherichia coli infection. Br J Nutr 106: 870-877. doi:S0007114511001152 [pii] 10.1017/S0007114511001152 [doi]

27. Eyng, C., R. V. Nunes, P. C. Pozza, A. E. Murakami, C. Scherer, and R. A. Schone. 2013. Carcass yield and sensorial analysis of meat from broiler chicken fed with tilapia byproducts meal. Ciencia E Agrotecnologia 37: 451-456.
28. FAO. 2004. Protein sources for the animal feed industry., Food and Agriculture Organisation, Bangkok.
29. FAO. 2017. Aquaculture Feed and Fertilizer Resources Information System. Food and Agriculture Organization of the United Nations, Rome, Italy.
30. FAOSTAT. 2009. FAO Statistical Databases, Food and Agriculture Organization of the United Nations, Rome, Italy.
31. Feedipedia. 2016. Animal Feed Resources Information System - INRA CIRAD AFZ and FAO.
32. Fischer, M., A. G. J. Voragen, S. R. Piersma, L. V. Kofod, C. I. Joergensen, P. Guggenbuhl, C. S. Nunes, and H. Gruppen. 2007. Presence of indigestible peptide aggregates of soybean meal in pig ileal digesta residue. Journal of the Science of Food and Agriculture 87: 2229-2238. doi:10.1002/jsfa.2948
33. Font-I-Furnols, M., and L. Guerrero. 2014. Consumer preference, behavior and perception about meat and meat products: An overview. Meat Science 98: 361-371. doi:10.1016/j.meatsci.2014.06.025
34. Forslund, K., S. Sunagawa, L. P. Coelho, and P. Bork. 2014. Metagenomic insights into the human gut resistome and the forces that shape it. Bioessays 36: 316-329. doi:10.1002/bies.201300143 [doi]
35. Gao, Y. Y., Z. Y. Jiang, Y. C. Lin, C. T. Zheng, G. L. Zhou, and F. Chen. 2011. Effects of spray-dried animal plasma on serous and intestinal redox status and cytokines of neonatal piglets. J Anim Sci 89: 150-157. doi:10.2527/jas.2010-2967
36. Gerrard, J. A., M. Lasse, J. Cottam, J. P. Healy, S. E. Fayle, I. Rasiah, P. K. Brown, S. M. BinYasir, K. H. Sutton, and N. G. Larsen. 2012. Aspects of physical and chemical alterations to proteins during food processing - some implications for nutrition. British Journal of Nutrition 108: S288-S297. doi:10.1017/s000711451200236x
37. Gibbs, B. F., A. Zougman, R. Masse, and C. Mulligan. 2004. Production and characterization of bioactive peptides from soy hydrolysate and soy-fermented food. Food Res. Int. 37: 123-131. doi:10.1016/j.foodres.2003.09.010
38. Gonzalez-Vega, J. C., B. G. Kim, J. K. Htoo, A. Lemme, and H. H. Stein. 2011. Amino acid digestibility in heated soybean meal fed to growing pigs. J Anim Sci 89: 3617-3625. doi:10.2527/jas.2010-3465
39. Hashimoto, K., K. Takeda, T. Nakayama, and M. Shimizu. 1995. Stabilization of the tight junction of the intestinal Caco-2 cell monolayer by milk whey proteins. Bioscience, biotechnology, and biochemistry 59: 1951-1952.
40. Hata, I., S. Higashiyama, and H. Otani. 1998. Identification of a phosphopeptide in bovine alpha(s1)-casein digest as a factor influencing proliferation and immunoglobulin production in lymphocyte cultures. Journal of Dairy Research 65: 569-578. doi:10.1017/s0022029998003136

41. Hauschild, L., P. A. Lovatto, J. Pomar, and C. Pomar. 2012. Development of sustainable precision farming systems for swine: estimating real-time individual amino acid requirements in growing-finishing pigs. J Anim Sci 90: 2255-2263. doi:jas.2011-4252 [pii] 10.2527/jas.2011-4252 [doi]
42. Heim, G., T. Sweeney, C. J. O'Shea, D. N. Doyle, and J. V. O'Doherty. 2014. Effect of maternal supplementation with seaweed extracts on growth performance and aspects of gastrointestinal health of newly weaned piglets after challenge with enterotoxigenic Escherichia coli K88. British Journal of Nutrition 112: 1955-1965. doi:10.1017/s0007114514003171
43. Heinritz, S. N., R. Mosenthin, and E. Weiss. 2013. Use of pigs as a potential model for research into dietary modulation of the human gut microbiota. Nutrition Research Reviews 26: 191-209. doi:10.1017/s0954422413000152
44. Hermansson, A. M. 1979. Methods of studying functional-characteristics of vegetable proteins. Journal of the American Oil Chemists Society 56: 272-279. doi:10.1007/bf02671471
45. Howling, D. 1980. The influence of the structure of starch on its rheological properties. Food Chemistry 6: 51-61. doi:10.1016/0308-8146(80)90006-0
46. Hugenholtz, F. 2015. Mouse gut microbiomics of short chain fatty acid metabolism and mucosal responses, Wageningen University and Research, Wageningen, NL.
47. Hulshof, T. G., P. Bikker, A. F. B. van der Poel, and W. H. Hendriks. 2016. Assessment of protein quality of soybean meal and 00-rapeseed meal toasted in the presence of lignosulfonate by amino acid digestibility in growing pigs and Maillard reaction products. J Anim Sci 94: 1020-1030. doi:10.2527/jas2015-9700
48. IFIF. 2009. International Feed Industry Federation annual report 2008/09, International Feed Industry Federation - IFIF, Luxembourg.
49. IFIF. 2015. International Feed Industry Federation annual report 2014/2015, International Feed Industry Federation - IFIF, Luxembourg
50. Jahan-Mihan, A., B. L. Luhovyy, D. E. Khoury, and G. H. Anderson. 2011. Dietary Proteins as Determinants of Metabolic and Physiologic Functions of the Gastrointestinal Tract. Nutrients 3: 574-603.
51. Jansman, A. J. M. 2016. Health and functions of the gastrointestinal tract in pigs: Effects of functional ingredients and feed and ingredient processing. J Anim Sci 94: 12-21. doi:10.2527/jas2015-9886
52. Janusz, M., Z. Wieczorek, K. Spiegel, A. Kubik, Z. Szewczuk, I. Siemion, and J. Lisowski. 1987. Immunoregulatory properties of synthetic peptides, fragments of a proline-rich polypeptide (prp) from ovine colostrum. Mol Immunol 24: 1029-1031. doi:10.1016/0161-5890(87)90069-1
53. Jha, R., and J. F. D. Berrocoso. 2016. Dietary fiber and protein fermentation in the intestine of swine and their interactive effects on gut health and on the environment: A review. Animal Feed Science and Technology 212: 18-26. doi:10.1016/j.anifeedsci.2015.12.002
54. Julius, M. H., M. Janusz, and J. Lisowski. 1988. A colostral protein that induces the growth and differentiation of resting b-lymphocytes. Journal of Immunology 140: 1366-1371.

55. Kayser, H., and H. Meisel. 1996. Stimulation of human peripheral blood lymphocytes by bioactive peptides derived from bovine milk proteins. Febs Letters 383: 18-20. doi:10.1016/0014-5793(96)00207-4
56. Kidd, M. T., A. Corzo, D. Hoehler, E. R. Miller, and W. A. Dozier. 2005. Broiler responsiveness (Ross x 708) to diet varying in amino acid density. Poultry Science 84: 1389-1396.
57. Kogut, M. H., and R. J. Arsenault. 2016. Editorial: Gut Health: The New Paradigm in Food Animal Production. Frontiers in Veterinary Science 3: 71. doi:10.3389/fvets.2016.00071
58. Konig, J., J. Wells, P. D. Cani, C. L. Garcia-Rodenas, T. MacDonald, A. Mercenier, J. Whyte, F. Troost, and R.-J. Brummer. 2016. Human Intestinal Barrier Function in Health and Disease. Clinical and translational gastroenterology 7: e196. doi:ctg201654 [pii] 10.1038/ctg.2016.54 [doi]
59. Krajmalnik-Brown, R., Z.-E. Ilhan, D.-W. Kang, and J. K. DiBaise. 2012. Effects of Gut Microbes on Nutrient Absorption and Energy Regulation. Nutr Clin Pract 27: 201-214.
60. Larsen, N., F. K. Vogensen, F. W. J. van den Berg, D. S. Nielsen, A. S. Andreasen, B. K. Pedersen, W. Abu Al-Soud, S. J. Sorensen, L. H. Hansen, and M. Jakobsen. 2010. Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults. PLoS One 5. doi:10.1371/journal.pone.0009085
61. Lee, M. 2016. The human gut microbiota: stability and diversity. Fx Medicine.
62. Ley, R. E. 2010. Obesity and the human microbiome. Current Opinion in Gastroenterology 26: 5-11. doi:10.1097/MOG.0b013e328333d751
63. Lin, Y., G. W. Meijer, M. A. Vermeer, and E. A. Trautwein. 2004. Soy protein enhances the cholesterol-lowering effect of plant sterol esters in cholesterol-fed hamsters. J Nutr 134: 143-148.
64. Lindmark-Mansson, H., and B. Akesson. 2000. Antioxidative factors in milk. Br. J. Nutr. 84: S103-S110.
65. Liu, Y., J. Huang, Y. Hou, H. Zhu, S. Zhao, B. Ding, Y. Yin, G. Yi, J. Shi, and W. Fan. 2008. Dietary arginine supplementation alleviates intestinal mucosal disruption induced by Escherichia coli lipopolysaccharide in weaned pigs. Br J Nutr 100: 552-560. doi:S0007114508911612 [pii] 10.1017/S0007114508911612 [doi]
66. Looft, T., H. K. Allen, B. L. Cantarel, U. Y. Levine, D. O. Bayles, D. P. Alt, B. Henrissat, and T. B. Stanton. 2014a. Bacteria, phages and pigs: the effects of in-feed antibiotics on the microbiome at different gut locations. Isme Journal 8: 1566-1576. doi:10.1038/ismej.2014.12
67. Looft, T., H. K. Allen, T. A. Casey, D. P. Alt, and T. B. Stanton. 2014b. Carbadox has both temporary and lasting effects on the swine gut microbiota. Frontiers in Microbiology 5. doi:10.3389/fmicb.2014.00276
68. Louis, P., K. P. Scott, S. H. Duncan, and H. J. Flint. 2007. Understanding the effects of diet on bacterial metabolism in the large intestine. Journal of applied microbiology 102: 1197-1208. doi:JAM3322 [pii] 10.1111/j.1365-2672.2007.03322.x [doi]
69. Lozupone, C. A., J. I. Stombaugh, J. I. Gordon, J. K. Jansson, and R. Knight. 2012. Diversity, stability and resilience of the human gut microbiota. Nature 489: 220-230.

70. Maeno, M., N. Yamamoto, and T. Takano. 1996. Identification of an antihypertensive peptide from casein hydrolysate produced by a proteinase from Lactobacillus helveticus CP790. J. Dairy Sci. 79: 1316-1321.
71. Martinez, J. L., T. M. Coque, and F. Baquero. 2015. What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol 13: 116-123. doi:nrmicro3399 [pii] 10.1038/nrmicro3399 [doi]
72. Martinez, W. H. 1979. Functionality of vegetable proteins other than soy. Journal of the American Oil Chemists' Society 56: 280-284. doi:10.1007/bf02671472
73. Maruyama, N., Y. Maruyama, T. Tsuruki, E. Okuda, M. Yoshikawa, and S. Utsumi. 2003. Creation of soybean beta-conglycinin beta with strong phagocytosis-stimulating activity. Biochimica et biophysica acta 1648: 99-104. doi:S1570963903001134 [pii]
74. Matassa, S., N. Boon, I. Pikaar, and W. Verstraete. 2016. Microbial protein: future sustainable food supply route with low environmental footprint. Microb Biotechnol 9: 568-575. doi:10.1111/1751-7915.12369 [doi]
75. McIntosh, G. H., P. J. Royle, R. K. Le Leu, G. O. Regester, M. A. Johnson, R. L. Grinsted, R. S. Kenward, and G. W. Smithers. 1998. Whey proteins as functional food ingredients? International Dairy Journal 8: 425-434. doi:10.1016/s0958-6946(98)00065-x
76. Melton, S. L. 1990. Effects of feeds on flavor of red meat - a review. J Anim Sci 68: 4421-4435.
77. Metges, C. C. 2000. Contribution of microbial amino acids to amino acid homeostasis of the host. J Nutr 130: 1857S-1864S.
78. Moller, N. P., K. E. Scholz-Ahrens, N. Roos, and J. Schrezenmeir. 2008. Bioactive peptides and proteins from foods: indication for health effects. European journal of nutrition 47: 171-182. doi:10.1007/s00394-008-0710-2 [doi]
79. Moriyama, T., K. Kishimoto, K. Nagai, R. Urade, T. Ogawa, S. Utsumi, N. Maruyama, and M. Maebuchi. 2004. Soybean beta-conglycinin diet suppresses serum triglyceride levels in normal and genetically obese mice by induction of beta-oxidation, downregulation of fatty acid synthase, and inhibition of triglyceride absorption. Bioscience, biotechnology, and biochemistry 68: 352-359. doi:10.1271/bbb.68.352 [doi]
80. Morowitz, M. J., E. Carlisle, and J. C. Alverdy. 2011. Contributions of Intestinal Bacteria to Nutrition and Metabolism in the Critically Ill. Surg Clin North Am 91: 771-785.
81. Mosenthin, R., W. C. Sauer, R. Blank, J. Huisman, and M. Z. Fan. 2000. The concept of digestible amino acids in diet formulation for pigs. Livestock Production Science 64: 265-280. doi:10.1016/s0301-6226(99)00139-6
82. Mullally, M. M., H. Meisel, and R. J. FitzGerald. 1997. Identification of a novel angiotensin-I-converting enzyme inhibitory peptide corresponding to a tryptic fragment of bovine β-lactoglobulin. FEBS Lett. 402: 99-101. doi:10.1016/s0014-5793(96)01503-7
83. Neis, E. P. J. G., C. H. C. Dejong, and S. S. Rensen. 2015. The Role of Microbial Amino Acid Metabolism in Host Metabolism. Nutrients 7: 2930-2946.
84. NRC. 2012. Nutrient Requirements of Swine. National Academic Press, Washington, D.C. 20418 USA.

85. OECD-FAO, 2011. OECD-FAO agricultural outlook 2011-2020., OECD-FAO of the United Nations., Rome, Italy.
86. Otani, H., M. Monnai, and A. Hosono. 1992. Bovine kappa-casein as inhibitor of the proliferation of mouse splenocytes induced by lipopolysaccharide stimulation. Milchwissenschaft 47: 512-515.
87. OW. 2015. Oil World Annual 2015. http://www.oilworld.de Accessed Date Accessed.| doi:DOI|
88. Pan, L., P. F. Zhao, Z. Y. Yang, S. F. Long, H. L. Wang, Q. Y. Tian, Y. T. Xu, X. Xu, Z. H. Zhang, and X. S. Piao. 2016. Effects of Coated Compound Proteases on Apparent Total Tract Digestibility of Nutrients and Apparent Ileal Digestibility of Amino Acids for Pigs. Asian-Australasian Journal of Animal Sciences 29: 1761-1767. doi:10.5713/ajas.16.0041
89. Perez-Bosque, A., L. Miro, J. Polo, L. Russell, J. Campbell, E. Weaver, J. Crenshaw, and M. Moreto. 2010. Dietary Plasma Protein Supplements Prevent the Release of Mucosal Proinflammatory Mediators in Intestinal Inflammation in Rats. J. Nutr. 140: 25-30. doi:10.3945/jn.109.112466
90. Pihlanto-Leppala, A., P. Koskinen, K. Piilola, T. Tupasela, and H. Korhonen. 2000. Angiotensin I-converting enzyme inhibitory properties of whey protein digests: concentration and characterization of active peptides. J. Dairy Sci. 67: 53-64. doi:10.1017/s0022029999003982
91. Playford, R. J., C. E. Macdonald, and W. S. Johnson. 2000. Colostrum and milk-derived peptide growth factors for the treatment of gastrointestinal disorders. Am J Clin Nutr 72: 5-14.
92. Radcliffe, J. D., V. L. Imrhan, and A. M. Hsueh. 1998. The use of soy protein isolate to reduce the severity of 13-cis retinoic acid-induced hypertriglyceridemia. Cancer Detect Prev 22: 526-532.
93. Rejman, J. J., M. J. Lewis, and S. P. Oliver. 1993. Enhancement of mammary gland mononuclear cell proliferation by interleukin-2 in the presence of lactoferrin. Food Agric Immunol 5: 51-56. doi:10.1080/09540109309354782
94. Ren, W., Y. Yin, G. Liu, X. Yu, Y. Li, G. Yang, T. Li, and G. Wu. 2012. Effect of dietary arginine supplementation on reproductive performance of mice with porcine circovirus type 2 infection. Amino Acids 42: 2089-2094.
95. Round, J. L., and S. K. Mazmanian. 2009. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol 9: 313-323. doi:10.1038/nri2515
96. Sato, R., T. Noguchi, and H. Naito. 1986. Casein phosphopeptide (CPP) enhances calcium absorption from the ligated segment of rat small intestine. J Nutr Sci Vitaminol (Tokyo) 32: 67-76.
97. Schanes, K., S. Giljum, and E. Hertwich. 2016. Low carbon lifestyles: A framework to structure consumption strategies and options to reduce carbon footprints. Journal of Cleaner Production 139: 1033-1043. doi:10.1016/j.jclepro.2016.08.154
98. Schokker, D., J. Zhang, S. A. Vastenhouw, H. G. H. J. Heilig, H. Smidt, J. M. J. Rebel, and M. A. Smits. 2015. Long-Lasting Effects of Early-Life Antibiotic Treatment and Routine Animal Handling on Gut Microbiota Composition and Immune System in Pigs. PLoS One 10: e0116523.
99. Shahidi, F., and Y. Zhong. 2008. Bioactive peptides. Journal of AOAC International 91: 914-931.

100. Shimizu, M. 2004. Food-derived peptides and intestinal functions. Biofactors 21: 43-47.
101. Shinmoto, H., S. i. Dosako, and I. Nakajima. 2014. Anti-oxidant Activity of Bovine Lactoferrin on Iron/Ascorbate Induced Lipid Peroxidation. Biosci. Biotechnol. Biochem. 56: 2079-2080. doi:10.1271/bbb.56.2079
102. Spencer, S. P., and Y. Belkaid. 2012. Dietary and commensal derived nutrients: shaping mucosal and systemic immunity. Current opinion in immunology 24: 379-384. doi:10.1016/j.coi.2012.07.006
103. Standeven, A. M., R. L. Beard, A. T. Johnson, M. F. Boehm, M. Escobar, R. A. Heyman, and R. A. Chandraratna. 1996. Retinoid-induced hypertriglyceridemia in rats is mediated by retinoic acid receptors. Fundam Appl Toxicol 33: 264-271. doi:S0272059096901645 [pii]
104. Steinfeld, H., H. A. Mooney, F. S. &, and L. E. Neville. 2006. Livestock's Long Shadow: Environmental issues and options, Food and Agriculture Organisation, Rome.
105. Tabassum, A., T. Abbasi, and S. A. Abbasi. 2016. Reducing the global environmental impact of livestock production: the minilivestock option. Journal of Cleaner Production 112: 1754-1766. doi:10.1016/j.jclepro.2015.02.094
106. Tan, B., X. G. Li, X. Kong, R. Huang, Z. Ruan, K. Yao, Z. Deng, M. Xie, I. Shinzato, Y. Yin, and G. Wu. 2009. Dietary L-arginine supplementation enhances the immune status in early-weaned piglets. Amino acids 37: 323-331. doi:10.1007/s00726-008-0155-1 [doi]
107. Toutain, P.-L., A. A. Ferran, A. Bousquet-Melou, L. Pelligand, and P. Lees. 2016. Veterinary Medicine Needs New Green Antimicrobial Drugs. Frontiers in microbiology 7: 1196. doi:10.3389/fmicb.2016.01196 [doi]
108. Troy, D. J., and J. P. Kerry. 2010. Consumer perception and the role of science in the meat industry. Meat Science 86: 214-226. doi:10.1016/j.meatsci.2010.05.009
109. UN. 2015. World population prospects: The 2015 revisions, key findings and advance tables, United Nations, Department of Economic and Social Affairs, Population Divisions.
110. van Huis, A. 2013. Potential of Insects as Food and Feed in Assuring Food Security. Annu. Rev. Entomol. 58: 563-583. doi:10.1146/annurev-ento-120811-153704
111. Walsh, A. M., T. Sweeney, C. J. O'Shea, D. N. Doyle, and J. V. O'Doherty. 2013. Effect of dietary laminarin and fucoidan on selected microbiota, intestinal morphology and immune status of the newly weaned pig. British Journal of Nutrition 110: 1630-1638. doi:10.1017/s0007114513000834
112. Walther, B., and R. Sieber. 2011. Bioactive Proteins and Peptides in Foods. International Journal for Vitamin and Nutrition Research 81: 181-191. doi:10.1024/0300-9831/a000054
113. WHO-FAO. 2003. Global and regional food consumption patterns and trends No. Technical Report Series, No 916. p 13-29. World Health Organization, Geneva (Switzerland).
114. Williams, B. A., M. W. Verstegen, and S. Tamminga. 2001. Fermentation in the large intestine of single-stomached animals and its relationship to animal health. Nutrition research reviews 14: 207-228. doi:S0954422401000117 [pii] 10.1079/NRR200127 [doi]

115. Wong, C. W., H. F. Seow, A. J. Husband, G. O. Regester, and D. L. Watson. 1997. Effects of purified bovine whey factors on cellular immune functions in ruminants. Vet. Immunol. Immunopathol. 56: 85-96. doi:10.1016/s0165-2427(96)05732-7
116. Wood, J. F. 1987. The functional-properties of feed raw-materials and their effect on the production and quality of feed pellets. Animal Feed Science and Technology 18: 1-17. doi:10.1016/0377-8401(87)90025-3
117. Woyengo, T. A., E. Kiarie, and C. M. Nyachoti. 2010. Metabolizable energy and standardized ileal digestible amino acid contents of expeller-extracted canola meal fed to broiler chicks. Poultry Science 89: 1182-1189. doi:10.3382/ps.2009-00595
118. Wu, G. 2009. Amino acids: metabolism, functions, and nutrition. Amino acids 37: 1-17. doi:10.1007/s00726-009-0269-0 [doi]
119. Xiao, C. W., M. R. L'Abbe, G. S. Gilani, G. M. Cooke, I. H. Curran, and S. A. Papademetriou. 2004. Dietary soy protein isolate and isoflavones modulate hepatic thyroid hormone receptors in rats. J Nutr 134: 743-749.
120. Xiao, C. W., J. Mei, W. Huang, C. Wood, M. R. L'Abbe, G. S. Gilani, G. M. Cooke, and I. H. Curran. 2007. Dietary soy protein isolate modifies hepatic retinoic acid receptor-beta proteins and inhibits their DNA binding activity in rats. J Nutr 137: 1-6. doi:137/1/1 [pii]
121. Xiao, L., Q. Feng, S. Liang, S. B. Sonne, Z. Xia, X. Qiu, X. Li, H. Long, J. Zhang, D. Zhang, C. Liu, Z. Fang, J. Chou, J. Glanville, Q. Hao, D. Kotowska, C. Colding, T. R. Licht, D. Wu, J. Yu, J. J. Y. Sung, Q. Liang, J. Li, H. Jia, Z. Lan, V. Tremaroli, P. Dworzynski, H. B. Nielsen, F. Backhed, J. Dore, E. Le Chatelier, S. D. Ehrlich, J. C. Lin, M. Arumugam, J. Wang, L. Madsen, and K. Kristiansen. 2015. A catalog of the mouse gut metagenome. 33: 1103-1108.
122. Yamamoto, N., A. Akino, and T. Takano. 1994. Antihypertensive Effects of Different Kinds of Fermented Milk in Spontaneously Hypertensive Rats. Biosci. Biotechnol. Biochem. 58: 776-778.

# CHAPTER 2

## Protein, peptide, amino acid composition, and potential functional properties of existing and novel dietary protein sources for monogastrics

*Soumya K. Kar [1], Alfons J. M. Jansman[2], Sjef Boeren[3], Leo Kruijt[2,4], Mari A. Smits[1,4,5]*

[1]Host Microbe Interactomics Group, Wageningen University & Research, Wageningen, The Netherlands
[2]Wageningen Livestock Research, Animal Nutrition, Wageningen University & Research, Wageningen, The Netherlands
[3]Laboratory of Biochemistry, Wageningen University & Research, Wageningen, The Netherlands
[4]Wageningen University & Research Animal Breeding and Genomics, Wageningen, The Netherlands
[5]Wageningen Bioveterinary Research, Wageningen University & Research, Wageningen, The Netherlands

*Journal of Animal Science 2016, 94:30–39*

## Abstract

Replacement of current protein resources for novel or alternative sources may be one of the solutions to abolish the expected scarcity of dietary protein for animal feed. However, little is known about the nutritional, protein composition and potential functional value of such novel or alternative protein sources. In the present study, we used advanced proteomic and bioinformatic approaches to characterise the protein component of six different protein sources and predicted the bio-functionalities of their *in vivo* digestion products. We used casein (CAS), partially delactosed whey powder (DWP), spray dried porcine plasma (SDPP), soybean meal (SBM), wheat gluten meal (WGM) and yellow meal worm (YMW). We characterised and semi-quantified the individual proteins present in these resources by NanoLC-LTQ-Orbitrap-Mass Spectrometry (nLCMS). Based on the data obtained, we calculated the amino acid composition of the proteins that constitute 90% of the total calculated protein content and compared this with the chemically determined amino acid compositions of the corresponding protein sources. By the use of bioinformatic procedures, we predicted the bioactive properties of these protein sources after *in silico* digestion with monogastric proteolytic enzymes. We detected and semi-quantified 37, 58, 85, 188, 113 and 33 different individual proteins in CAS, DWP, SDPP, SBM, WGM and YMW, respectively. The calculated amino acid composition of the various protein sources was almost identical to the chemically determined composition, with correlation values (r) ranging from 0.85 to 0.94. Furthermore, we revealed that the selected protein sources are potentially rich in bioactive peptides, in particular of angiotensin-converting enzyme inhibitors and peptides with antioxidative properties. We discuss the results in terms of the benefit of the applied nLCMS-based approach for analysing protein feed ingredients and the use of these alternative sources of protein in animal feed for monogastrics. Furthermore, we discuss new potential applications of this method in the area of (animal) nutrition.

**Key words: amino acid, bioactive peptide, bioinformatics, monogastric, novel proteins, peptides.**

## Introduction

Nutritionists, food and feed manufacturers become increasingly conscious of the growing pressure upon the world's existing protein resources, mainly due to global shortage and increasing prices of available resources. One of the main strategies for the development of more sustainable livestock feeding systems, is the use of alternative protein resources which partly or wholly replace the current protein sources for monogastric animals. Therefore, it is urgently required to initiate evaluation studies on (novel) protein containing feed ingredients of different origin considering both their nutritional and functional value in terms of their capacity to support or modify nutrient supply, the animal's physiology, tissue development and functioning.

Currently, a range of protein chemistry techniques are used by feed analysis laboratories to evaluate quantitative and quality parameters of protein containing feed ingredients. Surprisingly, the impressive capabilities of modern-day proteomic techniques to characterise and quantify individual protein components within complex mixtures, are not fully exploited yet. The more so as the information resulting from such measurements can be enriched by information present in comprehensive, high-quality and freely accessible databases with information on their genome and protein composition. Such an approach allows to obtain information on the collective group of individual proteins in the resource instead of the protein component as a whole.

The objective of the present study was to demonstrate the added value of the use of proteomic and bioinformatic procedures to approximate the protein, peptide and amino acid composition as well as to predict the potential bioactive properties of the (novel) protein containing feed ingredients. Another objective was to characterise in detail the protein component of existing and potential novel protein sources to be included in feed for monogastric animals.

## Materials and Methods

### *Nutrient and Amino Acid Composition: Chemical Analysis*

Representative samples of dried and grounded casein (CAS), partially delactosed whey powder (DWP), spray dried porcine plasma (SDPP), soybean meal (SBM), wheat gluten meal (WGM) and yellow meal worm (YMW) were chemically analysed for dry matter (DM; NEN-ISO 6496 by 4 hours drying at 104°C), nitrogen (N; NEN-ISO 5983-2 by Kjeldahl method and CP calculated as Nx6.25), ash (NEN-ISO 5984 after 3 hours ashing at 550°C), ether extract (EE, NEN-ISO 6492 by extraction with petroleum ether), gross energy (GE; NEN- EN-ISO 9831 by bomb calorimetry) and amino acid composition (NEN-ISO 13903 by acid hydrolysis at 110°C for 23 hours and ion-exchange

chromatography with post column derivatization with ninhydrin; no analysis concerning tryptophan was carried out). Values for tryptophan (Trp) and non-protein nitrogen (NPN) were not measured directly but estimated according to the formula described below:

$$1.\ \text{Trp (g/kg)} = \left\{ \frac{\text{Trp}\ \left(\frac{\text{g}}{\text{kg}}\right)}{\text{Crude Protein}\ \left(\frac{\text{g}}{\text{kg}}\right)} \right\}^{*} \text{x Crude protein analysed (as fed basis)}$$

Where,
* For CAS, DWP, SBM and WGM, reference values were obtained from CVB (CVB, 2007). For SDPP and YMW reference values were obtained from Angulo and Cubilo (1998) and Finke (2002).

2. NPN (g/kg)=∑ Proportion of Nitrogen from individual amino acid residue (g/kg) -Nitrogen analysed (g/kg)

It was assumed that Asx and Glx consist of 50:50 aspartic acid (Asp) and asparagine (Asn) for the former and 50:50 for glutamic acid (Glu) and glutamine (Gln) in the latter.

***Mass Spectrometry and in silico Determination of Amino Acid Composition***
Isolation of the protein fraction from protein sources were based on the method of Filter-Aided Sample Preparation (FASP) where the samples were reduced, alkylated and digested with trypsin (Wisniewski et al., 2009). Protein digests were analysed on a NanoLC-LTQ-Orbitrap-MS (nLCMS) to identify and quantify (label free) the peptides/proteins in the protein sources evaluated (Lu et al., 2011). For identification of the proteins, MS/MS spectra were analysed with MaxQuant 1.3.0.5 (Cox and Mann, 2008) using default settings for the andromeda search engine (Cox et al., 2011) except that extra variable modifications were set for de-amidation of asparagine (N) and glutamine (Q) as described by Smaczniak et al. (2012). NanoLC and LTQ-Orbitrap-MS quality checks were carried out according to Lu et al. (2011). Bovine (*Bos taurus*), pig (*Sus scrofa*), soybean (*Glycine max*), wheat (*Triticum aestivum*) and yellow meal worm (*Tenebrio molitor*) with 23,868, 26,082, 64,675, 94,854 and 384 protein sequences respectively, were downloaded from Uniprot databases (Leinonen et al., 2004; Bairoch et al., 2005; Jain et al., 2009) available at http://www.uniprot.org. These were then used together with a database that contained 36 protein sequences of potential contaminants like BSA (P02769, bovine serum albumin precursor), Trypsin (P00760, bovine), Trypsin (P00761, porcine), Keratin K22E (P35908, human), Keratin K1C9 (P35527, human), Keratin K2C1 (P04264, human) and Keratin K1CI (P35527, human) usually contributed by the biologicals used in the experiment and/or from the environment. Peptides and proteins with a false discovery

rate (FDR) of < 1% and a minimum of two peptides per protein (one or more of which unique for each protein) were accepted for reliable identification of proteins in the feed ingredients. The "label-free quantification" as well as the "match between runs" (set to 2 minutes) options were enabled in MaxQuant. De-amidated peptides were allowed to be used for protein quantification and all other quantification settings were kept default (Hubner et al., 2010). iBAQ intensities (that is the total protein intensity divided by the number of tryptic peptides that have 7 to 25 amino acids) were used to rank the proteins in accordance to their abundance. Based on the total iBAQ of all identified proteins a list of top-ranking proteins was made that together constituted 90% of the total protein content of the protein source. Subsequently, amino acid sequences without post-translational modifications of these proteins (along with their isoforms) were manually retrieved from the UniProt database, and a data file was created for each of the evaluated protein sources. To allow single step analysis, the protein sequences (belonging to each protein source) were connected in a consecutive order in their respective data files. The data with aggregated amino acid sequences of the abundant proteins in each protein containing feed ingredient were termed as "representative protein from feed ingredients" (RPFI). For the in *silico* assessment of the amino acid composition of RPFI, the sequences of RPFI were used as input file for the ProtParam program tool (Gasteiger et al., 2005) available at http://www.expasy.org/tools/protparam.html. Pearson's correlation was used to calculate the r value between amino acid composition as determined by chemical analysis and as calculated through the in *silico* approach.

All feed ingredients samples were analysed in duplicate for nutrients, amino acid composition and mass spectrometry and results are shown as the average of duplicate measurements.

***In silico Digestion and Bio-functionality Determination***

To mimic *in vivo* protein digestion in monogastrics and to determine the bio-functionalities of the peptides potentially derived from various parent protein sources, an in *silico* enzymatic digestion on the RPFI sequence was performed with pepsin (E.C. 3.4.23.1), trypsin (E.C. 3.4.21.4) and chymotrypsin (E.C. 3.4.21.1). The bio-functional annotation of the formed peptides is built-in into the BIOPEP database (Minkiewicz et al., 2008; Iwaniak and Dziuba, 2011) available at http://uwm.edu.pl/biochemia. The proportion of bio-functional activities from individual protein sources were calculated relative to the total amount of bio-functional activities observed for all peptides for representation into a doughnut chart.

## Results

The nutrient composition of each protein feed ingredient, as determined by classical chemical analysis, is shown in Table 1. The amino acid composition of the feed ingredients as determined by chemical analysis is summarised in Table 2.

**Table 1. Nutrient composition of the evaluated protein sources.**

| | Feed ingredients[1*] | | | | | |
|---|---|---|---|---|---|---|
| | Animal (Vertebrate) | | | Plant | | Insects (Invertebrate) |
| Items | CAS[3] | DWP[4] | SDPP[5] | SBM[4] | WGM[6] | YMW[7] |
| DM[2], g/kg (as in) | 906 | 945 | 915 | 890 | 928 | 959 |
| Protein, g/kg (DM) | 968 | 263 | 875 | 548 | 827 | 476 |
| Ash, g/kg (DM) | 23 | 184 | 78 | 66 | 10 | 34 |
| Total Fat, g/kg (DM) | 6 | 22 | 5 | 35 | 63 | 270 |
| Gross energy, kJ/g (DM) | 24 | 16 | 22 | 20 | 24 | 28 |

[1]Feed ingredients: CAS is casein, DWP is partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal, WGM is wheat gluten meal and YMW is yellow meal worm.
[2]DM: Dry matter
[*]Source of procurement for feed ingredients
[3]Fronterra, Auckland, New Zealand
[4]Reserach Diet Services BV, The Netherlands
[5]Darling Ingredients Internationals, The Netherlands
[6]Cargill BV, The Netherlands
[7]Kreca, The Netherlands

**Table 2. Amino acid composition of the evaluated protein sources as determined chemically**

| | Feed ingredients[1] | | | | | |
|---|---|---|---|---|---|---|
| | | Animal (Vertebrate) | | Plant | | Insects (Invertebrate) |
| Items | CAS, g/kg | DWP, g/kg | SDPP, g/kg | SBM, g/kg | WGM, g/kg | YMW, g/kg |
| Ala | 31.2 | 12.2 | 48.4 | 24.0 | 21.1 | 35.1 |
| Arg | 36.6 | 5.5 | 53.1 | 39.2 | 27.3 | 25.5 |
| Asx[2] | 72.4 | 27.4 | 87.9 | 63.3 | 27.0 | 39.5 |
| Cys | 3.8 | 6.1 | 31.7 | 8.1 | 17.1 | 5.0 |
| Glx[3] | 231.1 | 43.4 | 126.2 | 101.6 | 310.0 | 56.6 |
| Gly | 19.0 | 5.2 | 32.2 | 23.5 | 27.6 | 25.7 |
| His | 30.7 | 5.1 | 30.5 | 15.5 | 17.6 | 15.5 |
| Ile | 65.8 | 18.8 | 41.1 | 31.5 | 36.7 | 26.0 |
| Leu | 97.0 | 26.0 | 88.0 | 42.9 | 57.9 | 36.0 |
| Lys | 86.1 | 19.9 | 83.1 | 35.1 | 13.8 | 28.4 |
| Met | 26.9 | 3.5 | 7.3 | 6.9 | 11.3 | 5.9 |
| Phe | 53.6 | 7.7 | 52.2 | 29.0 | 44.4 | 15.3 |
| Pro | 114.9 | 18.0 | 59.1 | 30.0 | 123.1 | 42.0 |
| Ser | 59.5 | 12.8 | 51.6 | 28.3 | 40.4 | 22.3 |
| Thr | 44.5 | 17.6 | 53.0 | 22.1 | 21.3 | 19.8 |
| Tyr | 59.4 | 8.8 | 50.6 | 21.1 | 29.5 | 47.1 |
| Val | 69.9 | 16.1 | 62.2 | 28.4 | 37.1 | 31.9 |
| Trp | 12.6 | 3.7 | 15.7 | 7.2 | 7.4 | 3.8 |
| NPN[4] | 16.8 | 8.9 | 11.8 | 13.4 | 28.4 | 12.1 |

[1]Feed ingredients: CAS is casein, DWP is partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal , WGM is wheat gluten meal and YMW is yellow meal worm.

[2]Asx = 50:50 Asp:Asn

[3]Glx = 50:50 Glu:Gln

[4]NPN = Non Protein Nitrogen.

By using the proteomic / bioinformatic approach we detected and semi-quantified 37, 58, 85, 188, 113 and 33 different proteins in CAS, DWP, SDPP, SBM, WGM and YMW, respectively (Table 3 and Supplemental Table S1). The total number of proteins which contribute to at least 90% of the calculated protein content are shown in Table 3. Based on the identified proteins, which together make up 90% of the total calculated protein content, we determined the amino acid composition of each feed ingredient and compared the results with those of the chemically determined amino acid composition.

**Table 3. Number of identified proteins and number of proteins together constituting 90% of the total protein content from different protein sources. The identified proteins along with their Uniprot ID, names of identified proteins, molecular weight, sequence length and relative intensity (iBAQ) are given in supplemental Table S1.**

| Source | Feed ingredients[1] | Species | Identified proteins from feed ingredients by nLC-MS | Number of proteins (including isoform) forming 90% of total protein fraction |
|---|---|---|---|---|
| Animal (Vertebrates) | CAS | *Bos taurus* | 37 | 3 |
| | DWP | *Bos taurus* | 58 | 3 |
| | SDPP | *Sus scrofa* | 85 | 25 |
| Plant | SBM | *Glycine max* | 188 | 68 |
| | WGM | *Triticum aestivum* | 113 | 74 |
| Insect (Invertebrates) | YMW | *Tenebrio molitor* | 33 | 13 |

[1]Feed ingredients: CAS is casein, DWP is partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal , WGM is wheat gluten meal and YMW is yellow meal worm.

As shown in Fig 1, the amino acid composition of the different protein sources based on both approaches was about the same with correlation values (r) ranging from 0.85 to 0.94 for the different protein sources (Supplemental Table S2). Further, Fig 2 displays the predicted proportion of the various bio-functional activities exerted by the selected protein sources as determined by the in *silico* procedure described in Materials and Methods.

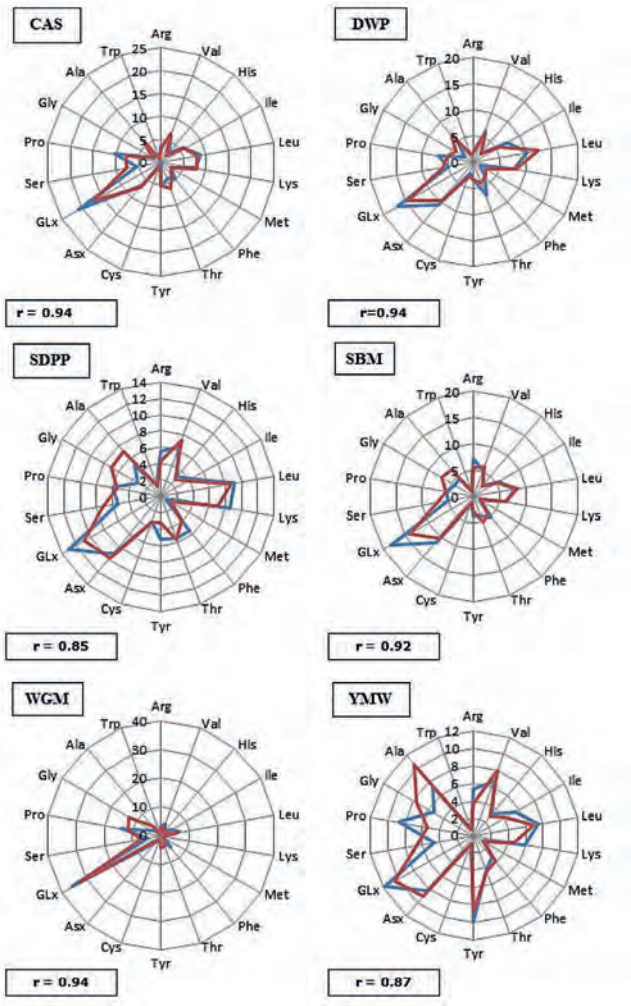



**Figure 1. Radial representation of amino acid composition of protein sources as analysed chemically and predicted in *silico*.**

Here, CAS is casein, DWP is partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal , WGM is wheat gluten meal and YMW is yellow meal worm.

Number on circles refer to amino acid values in percentage.

'r' means correlation values between in *silico* predicted and chemically defined amino acid composition.

Blue line: Amino acid composition as determined by chemical analysis

Red line: Amino acid composition as determined by the proteomic in *silico* approach.

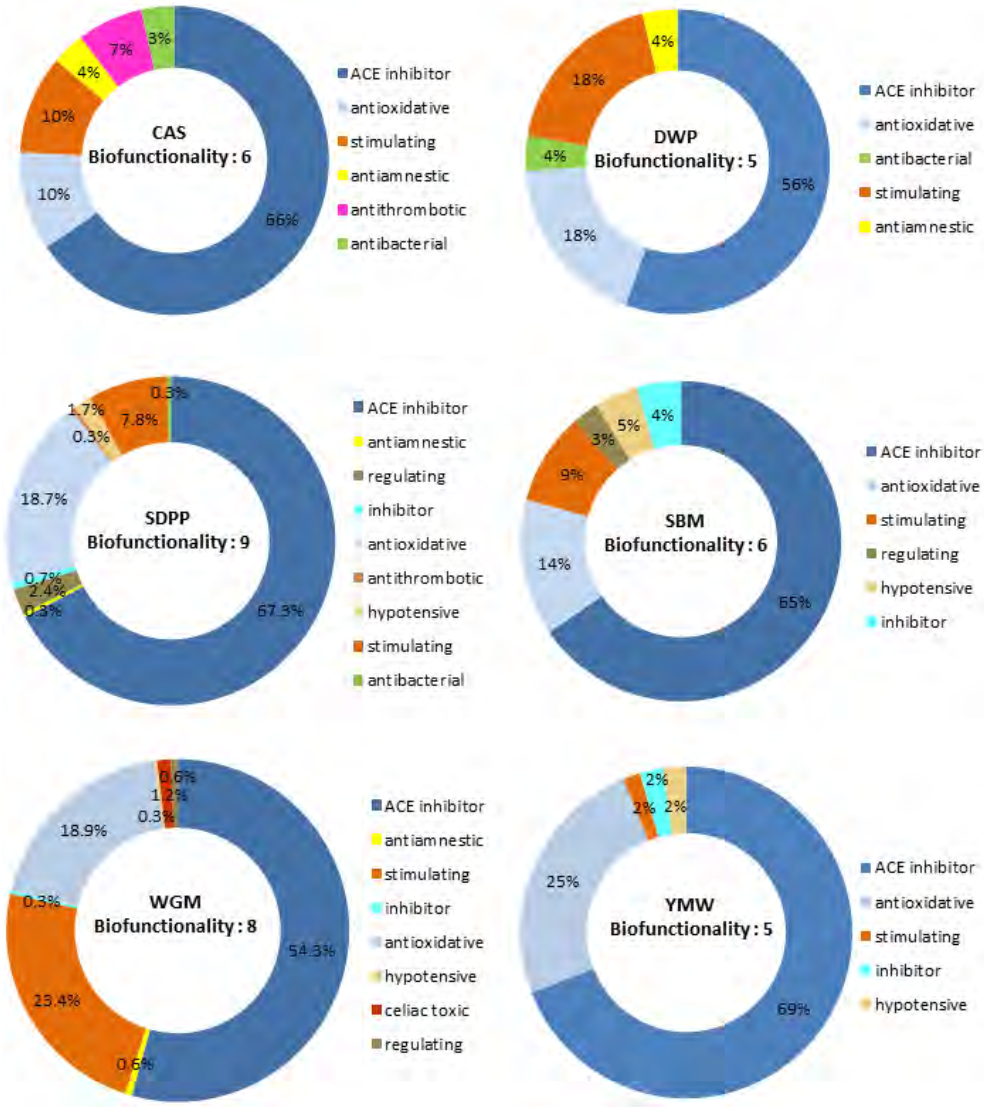



**Figure 2. Proportion1 (%) of determined bio-functional activities in various protein sources.**

[1]Proportion of bio functional activities exhibited after *in silico* digestion by pepsin, trypsin and chymotrypsin of the proteins that constitute 90% of the protein content in the different protein sources. The number within the circles exhibits bio-functional activities in various protein sources.

Here, CAS is casein, DWP is partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal , WGM is wheat gluten meal and YMW is yellow meal worm.

Different colours display various bio-functional activities.

According to BIOPEP database, stimulating include glucose uptake stimulating peptide and peptide stimulating vasoactive substance release; inhibitors include dipeptidyl-aminopeptidase IV inhibitor, CaMPDE inhibitor and Renin inhibitor and regulators include ion flow regulating peptide.

## Discussion

Traditionally, sources of dietary protein for animal feed are seen as either being of animal or vegetable origin (Hoffman and Falvo, 2004). Insects are increasingly considered as an alternative source of protein which can be efficiently grown on organic side streams (van Huis, 2013). Therefore, we included both animal, vegetable as well as insect derived protein sources in our study.

In general, we observed differences in the nutrient composition of the tested feed ingredients, in line with other references with data on feed ingredient composition (CVB, 2007; NRC, 2012). The nutrient composition of CAS, DWP, SBM, WGM as analysed are comparable to those given in previous reports (CVB, 2007). We recorded, however, a 1.1 to 1.3 times higher crude protein value for SDPP compared to previous reports (Angulo and Cubilo, 1998; Jamroz et al., 2011). The nutrient composition of YMW was comparable to that reported by Yi et al. (2013), although the total fat content for YMW was 2.2 times higher, which is also reflected in its relatively high gross energy value. The former may be explained by differences in the nutritional composition of the medium on which worms were grown and the stage of development of the worms at harvest (Ghaly and Alkoaik, 2009).

Our mass spectrometry (MS) analyses have resulted in the detection of high as well as low-abundant individual proteins in a variety of protein sources for animal feed. When compared to other protein sources, we identified relatively low number of proteins in YMW, which is almost certainly due to incompleteness of the YMW (*Tenebrio molitor*) UniProt database. This result is suggestive of the fact that a higher enrichment of the proteomic database for a particular source will improve identification and quantification of its constituent proteins. An advantage of this non-targeted MS approach is that it provides qualitative and quantitative information on the protein molecules present in the feed ingredients. Some important proteins which were identified by MS in the different sources and their functional biological characteristics are briefly discussed below.

*CAS and DWP*. We have detected the four main subunits of casein: s1 α casein, s2 α casein, β-casein and κ-casein (Swaisgood, 1993; Audic et al., 2003). Results also showed that CAS contained proteins or peptides such as α-lactalbumin, β-lactoglobulin, lactoferrin, lactoglobulin or immunoglobulins. Several milk proteins have been reported to modulate lymphocyte proliferation *in vitro* such as whole casein (Otani et al., 1992), α-, β-, κ-casein (Carr et al., 1990; Wong et al., 1996), whole whey protein (Barta et al., 1991), lactoferrin (Rejman et al., 1993), lactoperoxidase (Wong et al., 1997) and milk immunoglobulin G (Kulczycki et al., 1987). Technically, whey proteins are those that remain in milk serum after coagulating caseins at pH 4.6 and 20°C (Otani and Hata,

1995). Therefore, DWP contains many proteins which were also detected in CAS e.g. β-lactoglobulin, α-lactalbumin and immunoglobins (Papenburg et al., 1990; Otani et al., 1992). This resulted in a similar profiling of predictive bio-functionality for both protein sources. Moreover, it is well known that caseins function as precursors of various bioactive peptides and that whey proteins exhibit activities related to immune modulation and host defence (Madureira et al., 2007). During digestion process, s1 α-casein, β-casein, β-lactoglobulin and α-lactalbumin release several angiotensin I-converting enzyme (ACE)-inhibitory peptides (Yamamoto et al., 1994; Maeno et al., 1996; Mullally et al., 1997; Pihlanto-Leppala et al., 2000). Lactoferrin peptide obtained from casein yielded antioxidative activity (Lindmark-Mansson and Akesson, 2000; Shinmoto et al., 2014). Most of these published knowledge is in agreement with the predictions made by our in *silico* approach.

*SDPP.* Spray-dried plasma protein (SDPP) as feed ingredient is composed of a diverse mixture of functional proteins and other biologically important components. SDPP is largely composed of proteins, most importantly albumins and globulins. Within the globulins, the gamma globulins have an immune function, of which IgG's are the most important functional fraction of blood plasma. The other types, IgM, IgA, IgD and IgE are minor in blood plasma. As expected, the abundance of these proteins was rather high in our analysis. Furthermore, our results have shown that still other proteins are present in SDPP, especially proteins that belong to transferrin, glycoproteins, apolipoproteins, enzyme inhibitors (trypsin, chymotrypsin, carbonic anhydrase) and proteins involved in the blood coagulation mechanism. It has been suggested, however, that the health beneficial effects of spray dried plasma are related to its immunoglobulin content (Gatnau and Zimmerman, 1992).

*SBM.* Our results have shown that soy protein consist of a mixture of globular proteins - conglycinin and glycinin (Isanga and Zhang, 2008). The latter protein was found to be highly abundant. Other low abundant (glyco) proteins included lipoxygenases, lectins, trypsin inhibitors (Kunitz and Bowman-Birk) and amylases inhibitor (Stauffer, 1990; Liener, 1994; Cho et al., 1995). Protease inhibitors, physiologically active components obtained from soy were shown to have beneficial effects on lowering blood pressure in human (Teixeira et al., 2000). Administration of soy β-conglycinin (7S globulin) has been shown to ameliorate atherosclerosis in mice (Adams et al., 2004). Both α- and β-subunit of β-conglycinin have proven anti-inflammatory effects (Burris et al., 2014). Other identified bioactive peptides in soy-fermented food digested with proteolytic enzymes were ACE-inhibition and antioxidative properties (Pena-Ramos and Xiong, 2002; Gibbs et al., 2004). The predictions made by the in *silico* approach are in line with most of these previous observations.

*WGM*. We revealed 586 individual WGM proteins which mainly correspond to globulins (soluble in dilute salt solutions), the storage protein prolamin, gliadins (soluble in 70-90% ethanol) and glutenin (insoluble under all of the previously mentioned conditions). Proteomic analysis of mature wheat grain has revealed the presence of about 1125 individual proteins (Skylas et al., 2000). Based on our results, WGM proves to be an excellent source of glutamine and proline. The high concentration of these amino acids may be due to the presence of the glutamine and proline rich gliadin, glutenin and prolamin proteins in WGM, which are considered unique to wheat endosperm (Wieser, 2007).The observed relative high levels of glutamine and proline in WGM is in agreement with the finding of Apper-Bossard et al. (2013). Its high glutamine proportion acts as a major substrate for all rapidly proliferating cells and is a preferred energy source for intestinal tissue playing an important role in gut physiology and immunity (Souba, 1993; Wu et al., 1996; Stoll et al., 1999).



*YMW*. Our results suggest that feed ingredient YMW is mainly composed of proteins derived from exoskeleton chiefly cuticular protein. Other proteins within YMW belong to various haemolymph proteins, endurance proteins (like hexamerin, heat shock proteins -70 and -90) and different enzymatic proteins. Unfortunately, information on individual proteins in YMW and their bio-functional properties is scarce.

Knowledge on the abundance and identity of the various proteins in the feed ingredients allowed a rapid and reliable prediction of their amino acid compositions. These predicted amino acid composition profiles were almost completely identical to the amino acid profiles as determined by the standard chemical procedures. Pearson's correlation values were found to be high, indicating that the MS-based proteomic approach provides almost the same information regarding the amino acid composition of the feed ingredients compared to the classical analytical technique using HPLC after acid digestion of the sample.

Apart from being a source of amino acids to the host for growth and maintenance, proteins and peptides derived from the dietary sources during digestion process provide various protein motifs with defined biological functions to the host. Such peptide motifs may have opiate, antithrombotic, antihypertensive, immunomodulating, antilipemic, antithrombotic, osteoprotective, antioxidative, antimicrobial, ileum contracting, anticarcinogenic and growth promoting properties effects (Moller et al., 2008). The complete in silico digestion of the identified proteins in the various protein sources with pepsin, trypsin and chymotrypsin resulted mainly in the release of di- or tri-peptides. It appeared that most of the di- or tri-peptides that encompass beneficial bioactive peptides encode for ACE inhibitors and antioxidative activity. These were derived from all protein sources evaluated. In addition, we have observed different bio-functionalities

in varying proportions in the different protein sources. It should be noted, however, that we predicted bioactivities only from the whole protein sequence in all protein sources as digested with pepsin, trypsin and chymotrypsin and did not take into account the kinetics of protein hydrolysis which may result in additional sets of bioactive peptides. Furthermore, it should be stressed that our *in silico* analysis only points towards potentially available bioactivities. There are numerous reports on the bioactivity of proteins and peptides *in vitro*. Such data, however, are insufficient in claiming an effect on host health since the active compound may be degraded during *in vivo* intestinal digestion, may not be absorbed or not attain the appropriate concentrations in blood and target tissues that are required for acting significantly. In addition, the bioactivity may be reduced through molecular alterations during feed/food processing or interactions with constituents in other food ingredients. Furthermore, to exert the predicted bioactivities, di- and tripeptides should be absorbed by the intestinal mucosa, transferred to the blood circulation and reach the target cells in the body in substantial concentrations (Adibi, 1971; Hara et al., 1984; Masuda et al., 1996; van der Pijl et al., 2008). Further research needs to be carried out to gather more knowledge on the actual *in vivo* release, availability and abundance of biologically active peptides in protein sources for feed and food.

Relative to the amino acid requirements of monogastrics, SDPP has a very appropriate amino acid composition (Bureau et al., 1999), except for a relatively low concentration of methionine and isoleucine. SDPP also recorded the highest number of bioactivities compared to the other protein sources of animal origin evaluated (CAS and DWP). Along with the milk source (CAS and DWP), SDPP showed antibacterial properties in our study. This bio-functionality was confirmed *in vivo* by other workers where supplementation of SDPP reduced incidence of scours in piglets (Owusu-Asiedu et al., 2003). SDPP is recognized as a safe, high-quality feed ingredient for farmed animals, including pigs (Ferreira et al., 2009; Fruge et al., 2009). The high nutritional value of SDPP is related to the spray drying process applied during its manufacturing in which the proteins are subjected to less heat damage and denaturing conditions compared to traditional rendering processes to which other protein sources can be subjected (Luzier et al., 1995). SDPP has an good amino acid profile and apparent digestibility coefficients were recorded close to 99% (Bureau et al., 1999) and, when included in pig diet, resulted in an improvement of feed intake, body weight gain and feed efficiency (Ferreira et al., 2009). In addition, SDPP has also been recommended in animal diet as a source of immunological support due to their high levels of globulin proteins, including immunoglobulins (Campbell et al., 2010; Perez-Bosque et al., 2010; Gao et al., 2011). It should be noted that the use of animal protein in feed for livestock is strictly regulated by national and international bodies due to its potential to transmit infectious diseases (Sapkota et al., 2007).

By using the described proteomic approach, we were able to detect individual proteins in protein rich feed ingredients, thereby providing more detailed information on the composition of complex protein sources compared to conventional (nutritional) analytical approaches. This proteomic approach may be an attractive alternative to the conventional chemical methods for determining the amino acid composition of food/feed stuffs, thereby making it time and cost effective. The bioinformatic pipeline outlined in this study was not only effective in predicting the amino acid composition but it also allowed the prediction of potential bio-functional properties that may be displayed by the proteins from the various sources during or after in *vivo digestion*.

Our combined proteomic and bioinformatic approach may be exploited for other nutrition-related purposes such as (1) development of improved method to estimate protein quality of feed ingredients, (2) more precise monitoring of the effects of processing of feed and food ingredients and, (3) proteomic analysis of intestinal digesta for studying *in vivo* protein digestion kinetics.



## Acknowledgement

The authors acknowledge the financial support from the Wageningen UR "IPOP Customized Nutrition" programme financed by Wageningen UR, the Dutch Ministry of Economic Affairs, Agriculture & Innovation, the graduate school (WIAS) and industrial partners Nutreco, The Netherlands and Darling Ingredients Internationals, The Netherlands.

## References


1. Adams, M. R., D. L. Golden, A. A. Franke, S. M. Potter, H. S. Smith, and M. S. Anthony. 2004. Dietary soy beta-conglycinin (7S globulin) inhibits atherosclerosis in mice. J. Nutr. 134: 511-516.
2. Adibi, S. A. 1971. Intestinal transport of dipeptides in man: relative importance of hydrolysis and intact absorption. J. Clin. Invest. 50: 2266-2275 doi:10.1172/jci106724.
3. Angulo, E., and D. Cubilo. 1998. Effect of different dietary concentrations of spray-dried porcine plasma and a modified soyprotein product on the growth performance of piglets weaned at 6 kg body weight. Anim. Feed Sci. Technol. 72: 71-79 doi:10.1016/s0377-8401(97)00166-1.
4. Apper-Bossard, E., A. Feneuil, A. Wagner, and F. Respondek. 2013. Use of vital wheat gluten in aquaculture feeds. Aquat Biosyst 9: 21-21 doi:10.1186/2046-9063-9-21.
5. Audic, J.-L., B. Chaufer, and G. Daufin. 2003. Non-food applications of milk components and dairy co-products: A review. Lait 83: 417-438 doi:10.1051/lait:2003027.
6. Bairoch, A., R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Z. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L. S. L. Yeh. 2005. The universal protein resource (UniProt). Nucleic Acids Res. 33: D154-D159 doi:10.1093/nar/gki070.
7. Barta, O., V. D. Barta, M. V. Crisman, and R. M. Akers. 1991. Inhibition of lymphocyte blastogenesis by whey. Am. J. Vet. Res. 52: 247-253
8. Bureau, D. P., A. M. Harris, and C. Y. Cho. 1999. Apparent digestibility of rendered animal protein ingredients for rainbow trout (Oncorhynchus mykiss). Aquaculture 180: 345-358 doi:10.1016/s0044-8486(99)00210-0.
9. Burris, R. L., H. P. Ng, and S. Nagarajan. 2014. Soy protein inhibits inflammation-induced VCAM-1 and inflammatory cytokine induction by inhibiting the NF-kappaB and AKT signaling pathway in apolipoprotein E-deficient mice. Eur J Nutr 53: 135-148 doi:10.1007/s00394-013-0509-7.
10. Campbell, J. M., J. Polo, L. E. Russell, and J. D. Crenshaw. 2010. Review of spray-dried plasma's impact on intestinal barrier function. Livest Sci 133: 239-241 doi:10.1016/j.livsci.2010.06.075.
11. Carr, R. I., D. Webster, D. Sadi, H. Williams, and N. Walsh. 1990. Immunomodulation by Opioids from Dietary Casein (Exorphins). Ann. N. Y. Acad. Sci. 594: 374-376 doi:10.1111/j.1749-6632.1990.tb40500.x.
12. Cho, M. J., J. M. Widholm, and L. O. Vodkin. 1995. Cassettes for seed-specific expression tested in transformed embryogenic cultures of soybean. Plant Mol. Biol. Rep. 13: 255-269 doi:10.1007/bf02670904.
13. Cox, J., and M. Mann. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology 26: 1367-1372 doi:10.1038/nbt.1511.
14. Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann. 2011. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. J. Proteome Res. 10: 1794-1805 doi:10.1021/pr101065j.

15. CVB. 2007. CVB table pigs 2007: Chemical composition and Nutritional values of feedstuffs and feeding standards. CVB, Product Board Animal Feed, The Hague, The Netherlands.
16. Ferreira, A. S., F. F. Barbosa, M. D. Tokach, and M. Santos. 2009. Spray dried plasma for pigs weaned at different ages. Recent Pat Food Nutr Agric 1: 231-235
17. Finke, M. D. 2002. Complete nutrient composition of commercially raised invertebrates used as food for insectivores. Zoo Biol. 21: 269-285 doi:10.1002/zoo.10031.
18. Fruge, E. D., M. L. Roux, R. D. Lirette, T. D. Bidner, L. L. Southern, and J. D. Crenshaw. 2009. Effects of dietary spray-dried plasma protein on sow productivity during lactation. J. Anim. Sci. 87: 960-964 doi:10.2527/jas.2008-1353.
19. Gao, Y. Y., Z. Y. Jiang, Y. C. Lin, C. T. Zheng, G. L. Zhou, and F. Chen. 2011. Effects of spray-dried animal plasma on serous and intestinal redox status and cytokines of neonatal piglets. J. Anim. Sci. 89: 150-157 doi:10.2527/jas.2010-2967.
20. Gasteiger, Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., and B. A.;. 2005. Protein Identification and Analysis Tools on the ExPASy Server. Humana Press Inc., Totowa, NJ.
21. Gatnau, R., and D. R. Zimmerman. 1992. Determination of optimum levels of inclusion of spray-dried porcine plasma (SDPP) in diet for weanling pigs fed in practical conditions. J. Anim. Sci. 70: 60
22. Ghaly, A. E., and F. Alkoaik. 2009. The yellow mealworm as a novel source of protein. Am J Agric Biol Sci 4: 319-331
23. Gibbs, B. F., A. Zougman, R. Masse, and C. Mulligan. 2004. Production and characterization of bioactive peptides from soy hydrolysate and soy-fermented food. Food Res. Int. 37: 123-131 doi:10.1016/j.foodres.2003.09.010.
24. Hara, H., R. Funabiki, M. Iwata, and K. I. Yamazaki. 1984. Portal absorption of small peptides in rats under unrestrained conditions. J. Nutr. 114: 1122-1129
25. Hoffman, J. R., and M. J. Falvo. 2004. Protein - Which is best? J Sports Sci Med 3: 118-130
26. Hubner, N. C., A. W. Bird, J. Cox, B. Splettstoesser, P. Bandilla, I. Poser, A. Hyman, and M. Mann. 2010. Quantitative proteomics combined with BAC TransgeneOmics reveals *in vivo* protein interactions. J. Cell Biol. 189: 739-754 doi:10.1083/jcb.200911091.
27. Isanga, J., and G.-N. Zhang. 2008. Soybean Bioactive Components and their Implications to Health—A Review. Food Rev Int 24: 252-276 doi:10.1080/87559120801926351.
28. Iwaniak, A., and J. Dziuba. 2011. BIOPEP-PBIL Tool for the Analysis of the Structure of Biologically Active Motifs Derived from Food Proteins. Food Technol. Biotechnol. 49: 118-127
29. Jain, E., A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B. E. Suzek, M. J. Martin, P. McGarvey, and E. Gasteiger. 2009. Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics 10: 19 doi:10.1186/1471-2105-10-136.
30. Jamroz, D., A. Wiliczkiewicz, J. Orda, J. Skorupinska, M. Slupczynska, and J. Kuryszko. 2011. Chemical composition and biological value of spray dried porcine blood by-products and bone protein hydrolysate for young chickens. Br. Poult. Sci. 52: 589-605 doi:10.1080/00071668.2011.610298.

31. Kulczycki, A., G. S. Nash, M. J. Bertovich, H. D. Burack, and R. P. Macdermott. 1987. Bovine milk IgG, but not serum IgG, inhibits pokeweed mitogen-induced antibody secretion by human peripheral blood mononuclear cells. J. Clin. Immunol. 7: 37-45 doi:10.1007/bf00915423.
32. Leinonen, R., F. G. Diez, D. Binns, W. Fleischmann, R. Lopez, and R. Apweiler. 2004. UniProt archive. Bioinformatics 20: 3236-3237 doi:10.1093/bioinformatics/bth191.
33. Liener, I. E. 1994. Implications of antinutritional components in soybean foods. Crit Rev Food Sci Nutr 34: 31-67 doi:10.1080/10408399409527649.
34. Lindmark-Mansson, H., and B. Akesson. 2000. Antioxidative factors in milk. Br. J. Nutr. 84: S103-S110
35. Lu, J., S. Boeren, S. C. de Vries, H. J. F. van Valenberg, J. Vervoort, and K. Hettinga. 2011. Filter-aided sample preparation with dimethyl labeling to identify and quantify milk fat globule membrane proteins. J. Proteomics 75: 34-43 doi:10.1016/j.jprot.2011.07.031.
36. Luzier, J. M., R. C. Summerfelt, and H. G. Ketola. 1995. Partial replacement of fish meal with spray-dried blood powder to reduce phosphorus concentrations in diet for juvenile rainbow trout, Oncorhynchus mykiss (Walbaum). Aqacul. Res. 26: 587
37. Madureira, A. R., C. I. Pereira, A. M. P. Gomes, M. E. Pintado, and F. Xavier Malcata. 2007. Bovine whey proteins – Overview on their main biological properties. Food Res. Int. 40: 1197-1211 doi:10.1016/j.foodres.2007.07.005.
38. Maeno, M., N. Yamamoto, and T. Takano. 1996. Identification of an antihypertensive peptide from casein hydrolysate produced by a proteinase from Lactobacillus helveticus CP790. J. Dairy Sci. 79: 1316-1321
39. Masuda, O., Y. Nakamura, and T. Takano. 1996. Antihypertensive peptides are present in aorta after oral administration of sour milk containing these peptides to spontaneously hypertensive rats. J Nutr 126: 3063-3068
40. Minkiewicz, P., J. Dziuba, A. Iwaniak, M. Dziuba, and M. Darewicz. 2008. BIOPEP database and other programs for processing bioactive peptide sequences. J. AOAC Int. 91: 965-980
41. Moller, N. P., K. E. Scholz-Ahrens, N. Roos, and J. Schrezenmeir. 2008. Bioactive peptides and proteins from foods: indication for health effects. Eur J Nutr 47: 171-182 doi:10.1007/s00394-008-0710-2.
42. Mullally, M. M., H. Meisel, and R. J. FitzGerald. 1997. Identification of a novel angiotensin-I-converting enzyme inhibitory peptide corresponding to a tryptic fragment of bovine β-lactoglobulin. FEBS Lett. 402: 99-101 doi:10.1016/s0014-5793(96)01503-7.
43. NRC. 2012. Nutrient requirement of swine. 11th ed. Natl. Acad. Press, Washington, DC.
44. Otani, H., and I. Hata. 1995. Inhibition of proliferative responses of mouse spleen lymphocytes and rabbit Peyer's patch cells by bovine milk caseins and their digests. Journal of Dairy Research 62: 339-348
45. Otani, H., M. Monnai, and A. Hosono. 1992. Bovine kappa-casein as inhibitor of the proliferation of mouse splenocytes induced by lipopolysaccharide stimulation. Milchwissenschaft 47: 512-515

46. Owusu-Asiedu, A., C. M. Nyachoti, S. K. Baidoo, R. R. Marquardt, and X. Yang. 2003. Response of early-weaned pigs to an enterotoxigenic Escherichia coli (K88) challenge when fed diet containing spray-dried porcine plasma or pea protein isolate plus egg yolk antibody. J. Anim. Sci. 81: 1781-1789
47. Papenburg, R., G. Bounous, D. Fleiszer, and P. Gold. 1990. Dietary milk proteins inhibit the development of dimethylhydrazine-induced malignancy. Tumour Biol. 11: 129-136
48. Pena-Ramos, E. A., and Y. L. Xiong. 2002. Antioxidant activity of soy protein hydrolysates in a liposomal system. J. Food Sci. 67: 2952-2956
49. Perez-Bosque, A., L. Miro, J. Polo, L. Russell, J. Campbell, E. Weaver, J. Crenshaw, and M. Moreto. 2010. Dietary Plasma Protein Supplements Prevent the Release of Mucosal Proinflammatory Mediators in Intestinal Inflammation in Rats. J. Nutr. 140: 25-30 doi:10.3945/jn.109.112466.
50. Pihlanto-Leppala, A., P. Koskinen, K. Piilola, T. Tupasela, and H. Korhonen. 2000. Angiotensin I-converting enzyme inhibitory properties of whey protein digests: concentration and characterization of active peptides. J. Dairy Sci. 67: 53-64 doi:10.1017/s0022029999003982.
51. Rejman, J. J., M. J. Lewis, and S. P. Oliver. 1993. Enhancement of mammary gland mononuclear cell proliferation by interleukin□2 in the presence of lactoferrin. Food Agric Immunol 5: 51-56 doi:10.1080/09540109309354782.
52. Sapkota, A. R., L. Y. Lefferts, S. McKenzie, and P. Walker. 2007. What do we feed to food-production animals? A review of animal feed ingredients and their potential impacts on human health. Environmental Health Perspectives 115: 663-670 doi:10.1289/ehp.9760.
53. Shinmoto, H., S. i. Dosako, and I. Nakajima. 2014. Anti-oxidant Activity of Bovine Lactoferrin on Iron/Ascorbate Induced Lipid Peroxidation. Biosci. Biotechnol. Biochem. 56: 2079-2080 doi:10.1271/bbb.56.2079.
54. Skylas, D. J., J. A. Mackintosh, S. J. Cordwell, D. J. Basseal, B. J. Walsh, J. Harry, C. Blumenthal, L. Copeland, C. W. Wrigley, and W. Rathmell. 2000. Proteome Approach to the Characterisation of Protein Composition in the Developing and Mature Wheat-grain Endosperm. J. Cereal Sci. 32: 169-188 doi:10.1006/jcrs.2000.0321.
55. Smaczniak, C., N. Li, S. Boeren, T. America, W. van Dongen, S. S. Goerdayal, S. de Vries, G. C. Angenent, and K. Kaufmann. 2012. Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues. Nat. Protoc. 7: 2144-2158 doi:10.1038/nprot.2012.129.
56. Souba, W. W. 1993. Intestinal glutamine metabolism and nutrition. J. Nutr. Biochem. 4: 2-9 doi:10.1016/0955-2863(93)90013-m.
57. Stauffer, C. E. 1990. Measuring Trypsin Inhibitor in Soy Meal: Suggested Improvements in the Standard Method. Cereal Chem 67: 296-302
58. Stoll, B., D. G. Burrin, J. Henry, H. Yu, F. Jahoor, and P. J. Reeds. 1999. Substrate oxidation by the portal drained viscera of fed piglets. Am. J. Physiol. Endocrinol. Metab. 277: E168-E175
59. Swaisgood, H. E. 1993. Review and update of casein chemistry. J. Dairy Sci. 76: 3054-3061

60. Teixeira, S. R., S. M. Potter, R. Weigel, S. Hannum, J. W. Erdman, and C. M. Hasler. 2000. Effects of feeding 4 levels of soy protein for 3 and 6 wk on blood lipids and apolipoproteins in moderately hypercholesterolemic men. Am. J. Clin. Nutr. 71: 1077-1084
61. van der Pijl, P. C., A. K. Kies, G. A. M. Ten Have, G. S. M. J. E. Duchateau, and N. E. P. Deutz. 2008. Pharmacokinetics of proline-rich tripeptides in the pig. Peptides 29: 2196-2202. doi:S0196-9781(08)00361-6.
62. van Huis, A. 2013. Potential of Insects as Food and Feed in Assuring Food Security. Annu. Rev. Entomol. 58: 563-583 doi:10.1146/annurev-ento-120811-153704.
63. Wieser, H. 2007. Chemistry of gluten proteins. Food Microbiol. 24: 115-119 doi:10.1016/j.fm.2006.07.004.
64. Wisniewski, J. R., A. Zougman, N. Nagaraj, and M. Mann. 2009. Universal sample preparation method for proteome analysis. Nat Methods 6: 359-362 doi:10.1038/nmeth.1322.
65. Wong, C. W., H. F. Seow, A. J. Husband, G. O. Regester, and D. L. Watson. 1997. Effects of purified bovine whey factors on cellular immune functions in ruminants. Vet. Immunol. Immunopathol. 56: 85-96 doi:10.1016/s0165-2427(96)05732-7.
66. Wong, C. W., H. F. Seow, A. H. Liu, A. J. Husband, G. W. Smithers, and D. L. Watson. 1996. Modulation of immune responses by bovine beta-casein. Immunol. Cell Biol. 74: 323-329 doi:10.1038/icb.1996.58.
67. Wu, G. Y., S. A. Meier, and D. A. Knabe. 1996. Dietary glutamine supplementation prevents jejunal atrophy in weaned pigs. J. Nutr. 126: 2578-2584
68. Yamamoto, N., A. Akino, and T. Takano. 1994. Antihypertensive Effects of Different Kinds of Fermented Milk in Spontaneously Hypertensive Rats. Biosci. Biotechnol. Biochem. 58: 776-778
69. Yi, L. Y., C. M. M. Lakemond, L. M. C. Sagis, V. Eisner-Schadler, A. van Huis, and M. van Boekel. 2013. Extraction and characterisation of protein fractions from five insect species. Food Chem 141: 3341-3348 doi:10.1016/j.foodchem.2013.05.115.

## Supplemental Information

Available online : Journal of Animal Science, J. Anim. Sci. 2016.94:30–39. doi:10.2527/jas2015-9677

Supplemental Table S1: The identified proteins along with their Uniprot ID, names of identified proteins, molecular weight, sequence length and relative intensity (iBAQ).

Supplemental Table S2: The amino acid composition of the different protein sources as determined chemically and predicted in silico along with their correlation values.

# CHAPTER 3

## Evaluation of functional properties of current and novel protein sources using enteroids

*Soumya K. Kar[1,2], Bart van der Hee[1], Linda M. P. Loonen[1], Nico Taverne[1], Alfons J. M. Jansman[3], Dirkjan Schokker[2], Johanna J. Taverne-Thiele1, Mari A. Smits[1,2,4], Jerry M. Wells[1].*

[1]Host Microbe Interactomics Group, Wageningen University & Research, Wageningen, the Netherlands
[2]Wageningen University & Research Animal Breeding and Genomics, Wageningen, The Netherlands
[3]Wageningen Livestock Research, Animal Nutrition, Wageningen University & Research, Wageningen, The Netherlands
[4]Wageningen Bioveterinary Research, Wageningen University & Research, Wageningen, the Netherlands

*Manuscript submitted for publication*

## Abstract

Protein sources may contain bioactive components with activities beyond their strictly nutritional properties. We have investigated the use of enteroids as a model to test the effects of different protein sources on the intestinal epithelium. Confluent 2D monolayers generated from mouse 3D enteroids from the duodenum were shown to form polarized monolayers with distinct basolateral and apical membrane surfaces and epithelial cell linages found in the tissue of origin. Mouse enteroids were exposed to different undigested protein sources (4% w/v, viz. soybean meal, SBM; casein, spray dried plasma protein and yellow meal worm) or DMEM as a control and RNA was isolated for genome-wide transcriptomics. All protein sources induced unique biological-processes and the only biological-process altered by each protein sources compared to control was cell proliferation. Results of SBM suggests that a component of SBM negatively regulates cholesterol and lipid biosynthetic pathways through down regulation of retinoic acid receptors in the intestinal epithelium. These results were consistent with *in vivo* studies on the hypo-triglycerdemic effect of soy protein. Taken together these results highlight enteroids as a promising new model to evaluate complex interaction between feed/ food ingredients and the intestinal epithelium.

**Key words: dietary protein source, enteroids, intestinal organoids, mice, two dimensional.**

## Introduction

Currently, the approach adopted by the animal nutrition field to formulate animal diet are based on the delivery of "strict-nutritional" values of ingredients in the diet that relates to ileal or faecal digestible nutrients and derived net energy during the digestion process. The nutrients relate to proteins/amino acids (AA), starch and sugars, fat, fermentable non-starch-polysaccharides, minerals and vitamins. In this context, multiple feed tables are available that provide reliable information on composition and nutritional (quantitative) values of feed ingredients that forms the base for feed formulation (CVB, 2007; NRC, 2012). However, apart from the "strict-nutritional" value, diet and their constituents may have other "non-strict-nutritional", properties e.g. anti-oxidative and oxidative effects, immune signalling, which may arise from digestion of proteins (Kar et al., 2016), carbohydrate structures recognised by innate receptors (Rosch et al., 2017; Wells et al., 2017) or other components (Biesalski et al., 2009; Jansman, 2016).



Currently, there is much interest in alternative, economical sources of protein for animal feed, driven by the expected growth in the human population and increasing demand for animal protein. Efficiencies are being looked for in animal production, including alternative sources of protein for feed. Proposed alternative sources of protein include insect larvae and blood plasma for which there is little knowledge about constituents that might have non-strict-nutritional effects on the host. Moreover, the current practise of formulating animal diet ignores the potential non-strict-nutritional properties of various feed ingredients (Jahan-Mihan et al., 2011).

Commonly used models to study intestinal responses to luminal factors include intestinal cancer cell lines, such as Caco-2, which can be grown as polarised cell monolayers with a distinct apical and basolateral surface. However, cancer cell lines display aneuploidy, and have undergone genetic rearrangements and deletions making them less physiologically relevant than primary tissue. Cancer cell lines also show greatly differing gene and protein expression, leading to pathway specific differences compared to primary tissue cells (Ertel et al., 2006; Kosti et al., 2016). Intestinal tissue explants are an alternative model for studying gut physiology (Randall et al., 2011), but rapidly undergo necrosis after excision, complicating the interpretation of results and leading to a high variability between tissue samples.

Recently a solution to the limitations of *in vitro* models has arisen though knowledge of how to maintain and proliferate stem cells and programme tissue specific stem cell differentiation using 'niche' factors and Matrigel as a growth support matrix (Sato et al., 2011b; Date and Sato, 2015). These advances have made it possible to generate three-dimensional (3D) mini-guts from isolated crypts of the small intestine (enteroids)

or colon (colonoids), collectively known as organoids, because of their resemblance to the organ from which they were derived (Ootani et al., 2009; Sato et al., 2009; Sato et al., 2011a; Spence et al., 2011; Mustata et al., 2013; Wang et al., 2013; Cao et al., 2015). Three-dimensional enteroids can be generated from isolated crypts within about 2 weeks, can be maintained for > 2 years in culture or cryopreserved and contain all the different cell linages found at the location of origin. Recently we and others (Moon et al., 2014; VanDussen et al., 2015) have used the primary cells from 3D cultures to grow monolayers in wells or semipermeable supports, enabling apical exposure to luminal factors. The two-dimensional (2D) gut models also contain different cell types found in the intestine and produce a secreted mucus layer composed of MUC2, unlike any of the cancer cell line models (Moon et al., 2014; VanDussen et al., 2015).

The aim of this study was to investigate the use of 2D enteroid monolayers as a model system to test the effects of current and alternative protein sources on the intestinal epithelium. Building on the pioneering work of Sato and Clevers (Sato 2011), we generated murine enteroids from isolated crypts of the small intestine. To facilitate apical exposure to different protein sources we mechanically-dissociated mature 3D enteroids, seeded them into 96 well plates and grew them until they reached a confluent 2D monolayer. The presence of different intestinal cell types and a secreted mucus layer in the 2D enteroid model was verified by histological methods. Soybean meal (SBM) protein (4% w/v) and other sources of dietary protein such as casein from milk (CAS), spray dried plasma protein from porcine blood (SDPP) and grounded freeze dried yellow meal worm larvae (YMW) were used without prior *in vitro* digestion to the apical side of the confluent 2D monolayers for 6 hours, after which RNA was isolated for genome-wide transcriptomics.

## Materials and Methods

### *Animals*

All protocols were approved by the animal ethics board of Wageningen University and Research, The Netherlands. Three nine-week-old wild type C57BL/6J female mice were purchased from Harlan Sprague Dawley Inc. (Horst, the Netherlands) and kept in a light and temperature-controlled animal facility of Wageningen University & Research (12:12 h reversed light/dark cycle, 20 ± 2°C). The mice were housed together in a specific pathogen-free environment with *ad libitum* access to standard diet (AIN93M) and water. Mice were euthanized and dissected to remove the duodenum which was used to generate the enteroids. A schematic representation of the study design is represented in Figure 1.

### *Crypt isolation and culture of 3D enteroids*

Murine duodenal enteroids were generated and maintained per methods described by Sato et al., (2009), Sato et al., (2011a); Dekkers et al., (2013) with slight modifications. Briefly, a 2 cm part of the duodenum was isolated and opened longitudinally. Duodenal segments were washed in ice-cold PBS until supernatant was clear. Subsequently, the tissue was incubated in phosphate-buffered saline solution (PBS) containing 2mM

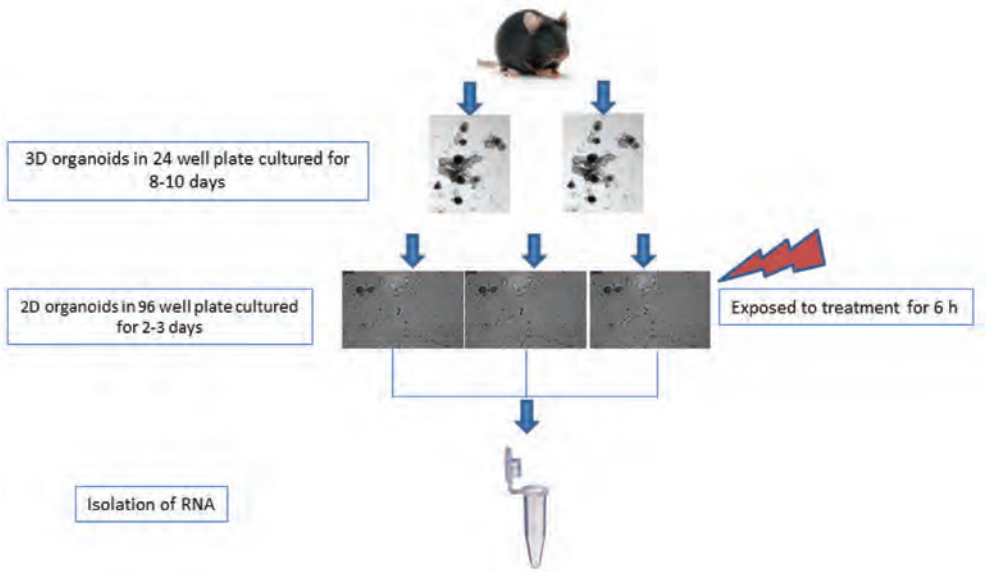



**Figure 1: Schematic representation of design of the experiment.**

Briefly, enteroids were grown in separate batches and each batch includes all treatments along with non-treated control (blank). Three dimensional (3D) murine enteroids were generated from a wild type C57BL/6J female mice by maintaining them under culturing condition for 8-10 days, for each batch. Subsequently, confluent 2D monolayers (n=3) were generated from the primary 3D enteroids cultures and incubated with 4% (w/v) CAS, SBM, SDPP and YMW for 6 hours at 37°C with 5% $CO_2$ prior to RNA isolation. DMEM which was used to suspend the protein sources was added to the 2D monolayers (n=3) as a control (blank).

EDTA for 30 minutes on ice. Intestinal villi were gently removed using a glass slide, the remaining tissue was sectioned in smaller pieces and washed with ice-cold PBS. After precipitation of the tissue fragments, PBS containing EDTA was removed and resulting fragments were thoroughly suspended in advanced Dulbecco's modified Eagle medium (DMEM/F12, ThermoFisher scientific, the Netherlands) containing 1% *v/v* penicillin/streptomycin (PenStrep, Sigma-Aldrich, the Netherlands). Supernatant containing crypts was filtered through a 70µm cell strainer and centrifuged at 200 x *g* for 3 minutes at 4 °C. The pellet was suspended in Matrigel matrix (growth factor reduced, phenol red free, BD biosciences, the Netherlands) and plated at a density of 40-100 crypts per 50µl in a 24-well culture plate (Corning, the Netherlands) for 3D growth of the enteroids. After inverted polymerization of the matrix at 37 $^{0}$C with 5% $CO_2$ for 20 minutes, 600 µl/well basal culture medium (DMEM/F12 medium); enriched with mouse EGF, Hepes 1M (Invitrogen, the Netherlands), N-acetyl cysteine (Sigma, the Netherlands), B-27 (ThermoFisher scientific, The Netherlands), Noggin (Hubrecht, Utrecht, Netherlands) and R-Spondin (Stanford university, Palo Alto, USA) was added. Recombinant mouse Noggin and R-Spondin were obtained from conditioned media of transfected HEK293 cell lines. After seeding, the culture medium was initially replaced after 24h and subsequently every 72h. Enteroids were sub-cultured and passed 1:5 every 8-10 days by mechanical disruption and seeded in fresh Matrigel matrix (BD Biosciences, the Netherlands).

***Culture of 2D enteroids***

Three dimensional (3D) grown enteroids were extracted from the plate by adding 1 ml/well ice-cold DMEM F12 containing 1% *v/v* penicillin/streptomycin. The enteroids were disrupted using a 200 µl pipette for 40 consecutive passages and centrifuged at 250 x *g* for 5 minutes at 4 °C. The resulting pellet was suspended in culture medium at room temperature and plated in a 96-well culture plate, coated with 0.5% Matrigel matrix.

***Histology***

Mouse organoids were grown 2D on transparent transwell filters (0.4 µm pore size, Corning). The filters were cut out of the cups and fixed in Carnoy's fixative. The filters were next embedded in paraffin through a graded series of washing with ethanol (50-100%) and xylene.

***PAS/Alcian blue staining***

Paraffin sections were cut at 4 µm and attached to slides. They were stained with a Pas/Alcian Blue staining in the same way as in Loonen et al. (2014) with some slight modifications: sections were stained in Alcian blue for 20 minutes, in per-iodic acid 0.5% for 5 minutes and in Schiff's reagents for 40 minutes. After the Schiff's reagent the slides were washed in fresh $SO_2$ diluted in water for 3 times 2 minutes, followed by a washing step in tap water for 5 minutes. After this washing step the slides were incubated in

hematoxylin (Mayer) for 45 seconds, followed by a wash in tap water for 15 minutes. The slides were dehydrated and embedded in depex (BDH Chemicals, England).

***CD138 staining***

Paraffin sections were cut at 4 µm and attached to slides. After overnight incubation at 37 ˚C, slides were deparaffinised in xylene and rehydrated in ethanol and distilled water. An antigen retrieval step was performed by heating the sections for 15 min in 0.01 M citrate buffer (pH 6.0) at 95˚C. Sections were washed for 20 minutes in PBS at room temperature and for 10 minutes in PBS-T at 4 ˚C. A blocking step to reduce non-specific binding was included using 5% normal goat serum (Invitrogen) in PBS with 0.1% Triton X-100 for 30 min at room temperature. CD138 expression was detected by incubating the sections with Rat anti-Mouse CD138 antibody (BD Biosciences) PE labelled, 1:100 diluted in Tris-buffered saline, 0.1% Tween 20 (TBS-T) and incubated overnight at 4 ˚C. After the incubation sections were washed for 2 times 10 minutes in PBS-T. After the washing steps the sections were incubated with Draq5 (ThermoFisher Scientific, The Netherlands) (1:1000, Invitrogen) in TBS-T for 30 minutes. The slides were washed 2 times in TBS-T and embedded in fluoromount G (SouthernBiotec, Alabama, USA).

***Stimulation of 2D enteroids with preparation of dietary proteins***

Finely powdered SBM, CAS, SDPP and YMW were added to DMEM F12 media and vortexed for 1 min. Subsequently, we allowed this homogenate to rest for 15 mins at room temperature and vortexed again for 1 min. Finally, this homogenate was centrifuged at 10000g for 1 min to remove insoluble precipitates and the supernatant collected for use in the assay. The 2D enteroids culture was apically exposed to 4% (w/v) of each protein source and incubated for 6 hours.

***RNA isolation, transcriptome and biological pathway analysis***

After incubation, the wells were washed with 200 µl PBS at room temperature. Total RNA was isolated using the RNeasy Mini kit (Qiagen, The Netherlands), with a 15 minute DNase treatment (RNAse free DNAse kit, Qiagen). RNA purity and integrity was verified using spectrophotometry (NanoDrop Technologies, USA) and Bioanalyzer (Agilent, USA). The RNA was only used to generate cDNA and perform microarray hybridization when there was no evidence of RNA degradation (RNA Integrity Number > 8). The labelling, hybridization of individual samples on Affymetrix GeneChip mouse gene 1.1 ST arrays (Affymetrix, USA), scanning, quality control and normalization of the resulting datasets was performed as described previously (Sovran et al., 2015). Briefly, 100 ng of total RNA was labeled using the Ambion WT Expression kit (Life Technologies Ltd, UK) together with the Affymetrix GeneChip WT Terminal Labeling kit (Affymetrix, CA). Labelled samples were hybridized to Affymetrix GeneChip Mouse Gene 1.1 ST arrays. Hybridization, washing and scanning of the array plates was performed on an

Affymetrix GeneTitan Instrument, according to the manufacturer's recommendations. Quality control of the data sets obtained from the scanned Affymetrix arrays was performed using Bioconductor (Gentleman et al., 2004) packages integrated in an online "MADMAX" pipeline (Lin et al., 2011). Probe sets were redefined according to Dai et al. (Dai et al., 2005) using current genome information. In this study, probes were reorganized based on the Entrez Gene database. Normalized expression estimates were obtained from the raw intensity values using the Robust Multiarray Analysis (RMA) preprocessing algorithm available in the Bioconductor library affyPLM using default settings (Bolstad et al., 2004).

Differentially expressed probe sets were identified using linear models, applying moderated T-statistics that implemented empirical Bayes regularization of standard errors (Storey and Tibshirani, 2003). A Bayesian hierarchical model was used to define an intensity based moderated T-statistic (IMBT)(Sartor et al., 2006). Only genes with a fold-change (FC) of at least 1.5 (up or down) and p-value below 0.05 were considered to be significantly different. Biological interaction networks among regulated genes activated in response to protein ingredients from different sources were identified using "GeneAnalytics" (LifeMap Sciences, Inc. a subsidiary of BioTime, Inc., USA). GeneAnalytics (Ben-Ari Fuchs et al., 2016) automatically integrates gene-centric data from ~125 web sources, including genomic, transcriptomic, proteomic, genetic, clinical and functional information to identify Gene Ontology (GO) terms related to their gene sets, providing information about the molecular functions and biological roles of the genes of interest. The annotation of mice genes was performed for the subsequent functional analysis. Our GeneAnalytics analyses compared differentially regulated genes in the 2D enteroids exposed to treatments compared to non-treated control (blank). The input was all differentially regulated genes (p value < 0.05 and FC > 1.5) in the 2D enteroids with/without exposure to the treatments. Here, the GO biological processes were retrieved from GeneAnalytics analysis with a high or medium score (p-value <0.05).

***Validation of microarray by RT-qPCR***

Several genes that according to the microarray data analysis were significantly differentially expressed in protein-treated enteroids compared to non-treated control (blank) were selected for quantitative reverse transcription PCR (RT-qPCR) assays (Supplementary table 1). One microgram of total RNA was reverse transcribed into cDNA using a qScript cDNA synthesis kit (Quantabio, USA) according to manufacturer's instructions and diluted 1:20. Primers were designed using Primer3 software (Koressaar and Remm, 2007; Untergasser et al., 2007) and purchased from Eurogentec (Oligo center, Belgium). RT-qPCR was performed using the Rotor-gene SYBR green PCR kit (Qiagen, the Netherlands) with primers specified in Supplementary table 1. Expression levels were measured in triplicate assays per sample using the Rotor-gene Q2plex real-time cycler

(Qiagen). *18S* and *beta-Actin* acted as endogenous control genes and relative expression was calculated using individual amplification values, following methods described in (Schmittgen and Livak, 2008). RT-qPCR data are presented as mean ± standard error of the mean (SEM). Statistical analysis was performed by one-way analysis of variance (ANOVA) followed by Dunnet's multiple comparisons test (Treatments vs Blank) using GraphPad prism version 5.03 (GraphPad Software, San Diego, California, USA). P values < 0.05 were considered statistically significant.

## Results and Discussion

Confluent 2D monolayers generated from murine 3D enteroids were shown to form polarized monolayers with distinct basolateral and apical membrane surfaces by staining for CD138, a laminin binding protein expressed only on the basolateral membrane of the tight-junction epithelium (Figure 2) (Moon et al., 2014). These organoid-based models also produced a secreted mucus layer from goblet cells present in the monolayer (Figure 2) which is consistent with expression values for goblet cell-specific Muc2 transcripts in the microarray data. Similarly, the microarray data revealed expression of other genes expressed in specific cell linages of intestinal epithelium including *Lgr5, CD24* (stem cells)



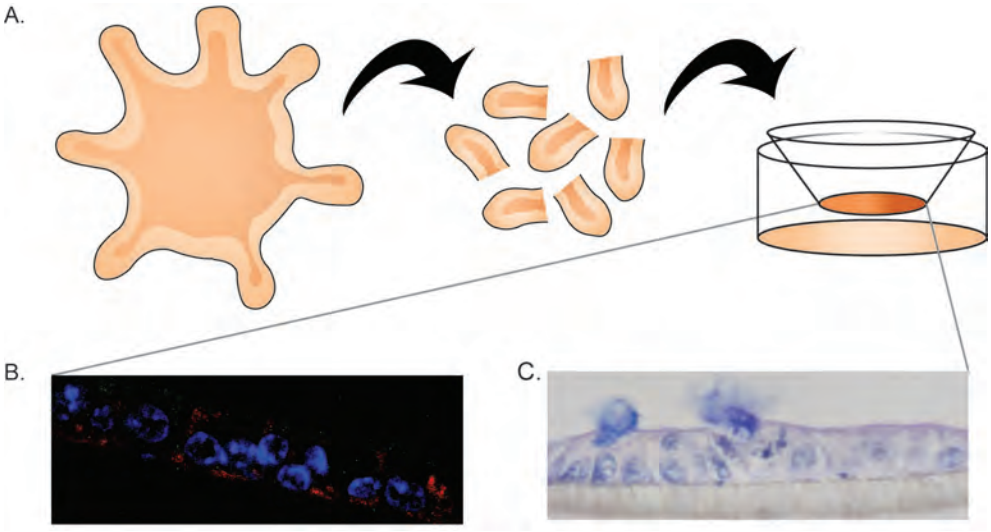



**Figure 2: Mouse enteroids culture systems.**
(A) 3D enteroid in Matrigel with an intestinal region-specific differentiation programme can be dissociated and seeded onto Matrigel-coated Transwells or plates to form 2D epithelial monolayers. (B) 2D enteroid monolayers form polarised monolayers (red= CD138 basolateral membrane marker; blue = nuclei). (C) 2D monolayer stained with PAS/Alcian blue revealing the presence of goblet cells secreting mucus.

(Sato et al., 2009), *Lyz1* (Paneth cells) (Sato et al., 2009) and *Cck* (enteroendocrine cells) (Sato et al., 2009).

The enteroids were exposed to different undigested protein sources (4% w/v, viz. SBM, CAS, SDPP, WGM and YMW) or DMEM as a control and RNA isolated for genome-wide transcriptomics profiling. The data is available in the Gene Expression Omnibus from NCBI with the accession number GSE98051. Principal Component Analysis (PCA) was performed to get more insight into the variability in the data. The variance explained by the first two axes was 77.2%. In the PCA plots the blank group showed a tendency to form a separate cluster, however, an overlap was observed at the 95% confidence regions with the other groups due to one of the samples (Figure 3). The clusters formed by the protein treatment groups showed a clear overlap of the 95% confidence regions.

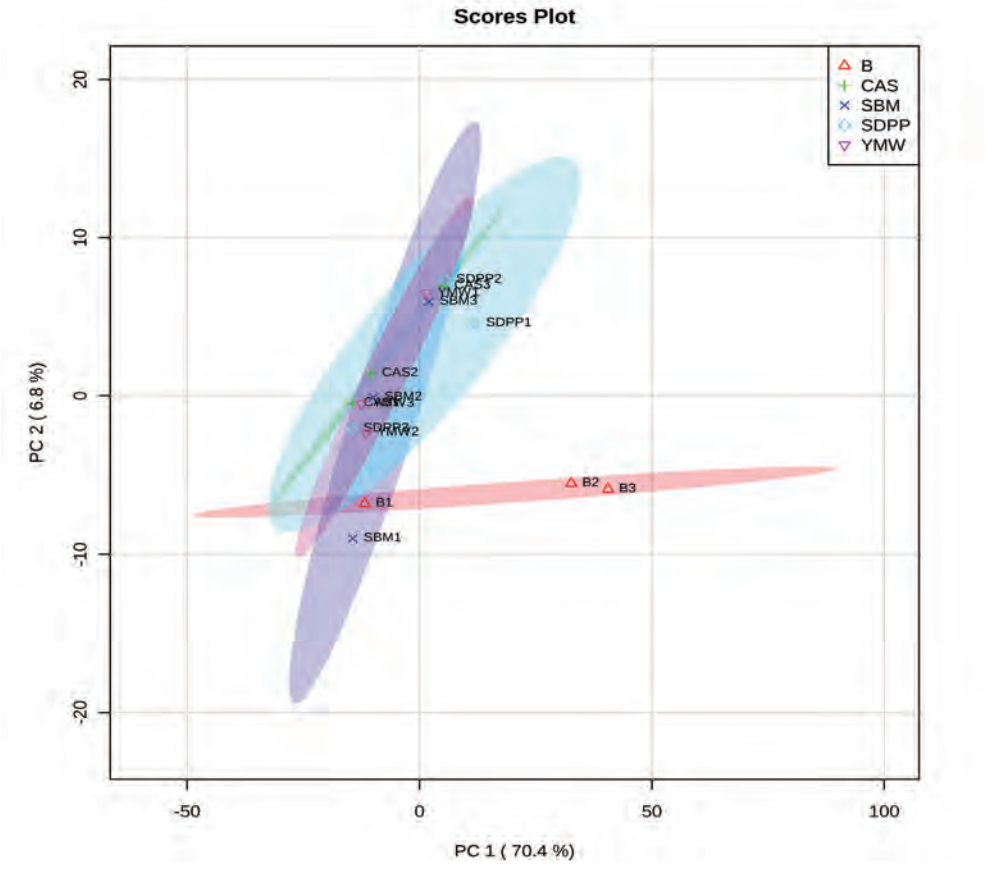



**Figure 3: PCA score plot based on genome-wide transcriptomic response measured by microarray at 2D enteroids stimulated with proteins ingredients from different source.**

Colored spherical area displays 95% confidence region of respective experimental diet. Each dot represent an batch culture of organoids. B: Blank; CAS: Casein; SBM: Soybean meal; SDPP: Spray dried plasma protein; YMW: Yellow meal worm.

Analysis of the microarray transcriptomics data revealed that hundreds of genes were significantly (p value < 0.05 and fold change > 1.5) differentially expressed in 2D enteroids following exposure to different protein sources for 6 hours (Table 1). Several of these differentially regulated genes were verified to have altered expression using RT-qPCR with a strong correlation (r = 0.8) with the microarray expression (Supplementary table 2). Using GeneAnalytics the significantly different genes were matched to gene ontology (GO) biological processes (Table 2). Strikingly, several unique biological processes were influenced by each of the protein sources, while only one biological process was commonly regulated (Figure 4, listed in Table 2). For CAS, SBM, SDPP and YMW, the number of unique GO-biological processes modulated compared to the control were 5, 8, 24 and 11, respectively (Figure 4). The biological process upregulated by all protein ingredients compared to the blank (medium control) was cell proliferation, indicating altered rates of cell death and renewal.



CAS exposure to 2D enteroids significantly downregulate expression of genes related to glutathione- metabolic and glutathione- derivative biosynthetic process. Glutathione is a tripeptide, ubiquitously distributed in living cells and plays an important role in the intracellular defence mechanism against oxidative stress (Diaz-Vivancos et al., 2015; Couto et al., 2016). It is known that glutathione metabolism is important for

**Table 1: Overview of significantly differentially expressed genes along with the functional analysis at gene ontology (GO)-biological process in 2D enteroids exposed to undigested protein from different sources.**

| Comparisons | Regulation | Differentially expressed genes [a] | | Matched GO biological processes [b] |
|---|---|---|---|---|
| CAS vs Blank | Up | 417 | 823 | 26 |
| | Down | 406 | | |
| SBM vs Blank | Up | 340 | 796 | 27 |
| | Down | 456 | | |
| SDPP vs Blank | Up | 262 | 379 | 40 |
| | Down | 117 | | |
| YMW vs Blank | Up | 391 | 824 | 20 |
| | Down | 433 | | |

[a] P < 0.05 and log Fold Change > |1.5|

[b] Analysed by GeneAnalytics; pathway analysis significance at corrected p value < 0.05

Abbreviations used: CAS is Casein, SBM is soybean meal, SDPP is spray dried plasma protein; YMW is yellow meal worm.

the antioxidant and detoxifying action of the intestine (Iantomasi et al., 1997). This suggests that CAS has detoxifying or antioxidative (functional) properties reducing the requirement for glutathione.

Exposure of enteroids to SDPP upregulate several biological pathways associated with processes of cell migration, movement, including for example 'wound healing', 'cytokinesis' 'mitotic spindle midzone assembly'. This is consistent with observed upregulation of 'DNA-templated transcription, initiation' and 'nucleosome assembly'. Additionally, upregulation of angiogenesis could be related to cell apoptosis and the altered expression of membrane metalo-proteases and connective tissue growth factors involved in epithelial repair processes. Overall these findings suggest increased cell turnover reflected in the upregulation of apoptosis and replication processes. The reason for this is unclear but warrants further investigation.

**Table 2: List of significantly unique biological processes [a] as analysed in GeneAnalytics with the significant differential genes [b] at each comparisons of 2D enteroids exposed to treatment vs blank.**

| Sl No. | Biological processes modulated in treatments compared to blank | Putative regulation of the pathways |
|---|---|---|
| | **CAS** | |
| 1 | Glutathione Metabolic Process | Down |
| 2 | Notochord Development | Up |
| 3 | Microtubule-based Movement | Up |
| 4 | Glutathione Derivative Biosynthetic Process | Down |
| 5 | Cellular Detoxification of Nitrogen Compound | Down |
| | **SBM** | |
| 1 | Phosphatidylcholine Biosynthetic Process | Down |
| 2 | Triglyceride Homeostasis | Down |
| 3 | Drug Metabolic Process | Down |
| 4 | Type I Interferon Signaling Pathway | Up |
| 5 | Retinol Metabolic Process | Down |
| 6 | Hexose Transport | Down |
| 7 | Lipoprotein Metabolic Process | Down |
| 8 | Retinoid Metabolic Process | Down |
| | **SDPP** | |
| 1 | Nucleosome Assembly | Up |
| 2 | Positive Regulation of Apoptotic Process | Up |
| 3 | Positive Regulation of Cell Migration | Up |

| Sl No. | Biological processes modulated in treatments compared to blank | Putative regulation of the pathways |
|---|---|---|
| 4 | Response to Wounding | Up |
| 5 | Wound Healing, Spreading of Cells | Up |
| 6 | Negative Regulation of Cell Cycle | Up |
| 7 | DNA Replication-independent Nucleosome Assembly | Up |
| 8 | Cellular Response to Organic Cyclic Compound | Up |
| 9 | Angiogenesis | Up |
| 10 | Telomere Organization | Up |
| 11 | Response to Virus | Up |
| 12 | Cytokinesis | Up |
| 13 | DNA Replication-dependent Nucleosome Assembly | Up |
| 14 | DNA-templated Transcription, Initiation | Up |
| 15 | Response to Hypoxia | Up |
| 16 | Positive Regulation of Fever Generation | Up |
| 17 | Mitotic Spindle Midzone Assembly | Up |
| 18 | Platelet Degranulation | Up |
| 19 | Positive Regulation of Cytokinesis | Up |
| 20 | Chromatin Silencing at RDNA | Up |
| 21 | Protein Stabilization | Up |
| 22 | Protein Localization to Kinetochore | Up |
| 23 | Omega-hydroxylase P450 Pathway | Up |
| 24 | Positive Regulation of Protein Localization to Nucleus | Up |
| | **YMW** | |
| 1 | Cholesterol Biosynthetic Process | Down |
| 2 | Steroid Metabolic Process | Down |
| 3 | Response to Drug | Down |
| 4 | Oxidation-reduction Process | Up |
| 5 | Transmembrane Transport | Down |
| 6 | Isoprenoid Biosynthetic Process | Down |
| 7 | Very Long-chain Fatty Acid Metabolic Process | Down |
| 8 | CDP-choline Pathway | Down |
| 9 | Long-chain Fatty Acid Metabolic Process | Down |
| 10 | Steroid Hormone Mediated Signaling Pathway | Down |
| 11 | Response to Gamma Radiation | Up |

[a] Analysed by GeneAnalytics; pathway analysis significance at corrected p value < 0.05

[b] $P < 0.05$ and log Fold Change > $|1.5|$

Abbreviations used: CAS is Casein, SBM is soybean meal, SDPP is spray dried plasma protein; YMW is yellow meal worm.

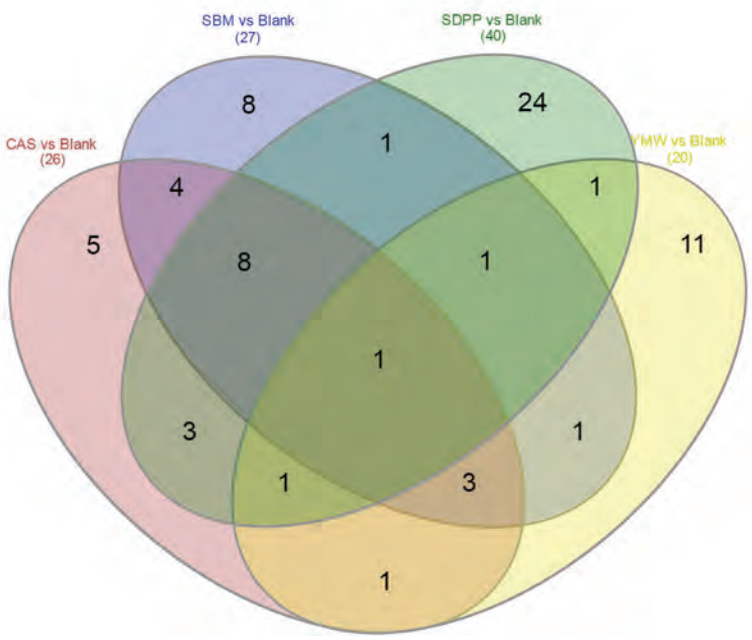



**Figure 4: Overview of non-overlapped and overlapped significant gene ontology (GO)-biological processes modulated by protein ingredients from different sources compared to blank, based on functional analysis result with GeneAnalytics.**

CAS: Casein; SBM: Soybean meal; SDPP: Spray dried plasma protein; YMW: Yellow meal worm

Exposure of 2D enteroids to YMW down-regulate biological processes involved in lipid metabolism processes such as 'steroid metabolic process', 'cholesterol biosynthesis process' very-long chain fatty acid metabolic process' and isoprenoid biosynthesis. Isoprenoid is need for biosynthesis of sterols such as cholesterol which is vital for membrane structure. These effects of YMW seem likely to be linked to the relatively high fat content of this protein source (Kar et al., 2016).

SBM is a common protein ingredient of animal feed and down-regulated biological processes involved in triglyceride and phosphatidylcholine biosynthesis as well as pathways associated with retinoid and retinol metabolism (Table 2). This is consistent with *in vivo* studies showing soy protein intake was shown to be hypotriglyceridemic in rats, having cholesterol lowering effects and reducing fatty liver (Ascencio et al., 2004; Lin et al., 2004; Moriyama et al., 2004; Xiao et al., 2004). The hypo-triglyceridemic effect of soy protein was recently shown to be due to suppression of retinoic acid receptor expression in liver of rats fed soy bean protein but not casein protein in their diet. This is consistent with earlier reports that administration of retinoids in cancer treatments are associated with hypertriglyceridemia via their interaction with retinoid receptors (Standeven et al., 1996; Radcliffe et al., 1998; Xiao et al., 2007). Taken together these

intriguing observations suggest that a component of SBM protein also negatively regulates retinoid and retinol biosynthesis as well as cholesterol and lipid biosynthetic pathways through down regulation of retinoic acid receptors in the intestinal epithelium.

Collectively the transcriptomics results indicate a direct diet-host interaction of the ' undigested-soluble fraction' of different protein sources and suggests that 2D enteroids are a promising new model to evaluate complex interaction between feed/ food ingredients and the intestinal epithelium. One of the main advantages of the enteroid models lies with the use of untransformed cells that can be cryopreserved or propagated in culture for at least 2 years. As such they represent a near physiological model of the intestinal epithelium, containing all intestinal cell types and having a global transcriptome more closely resembling tissue of origin than immortal cancer-derived cell lines. As demonstrated in this study, the possibility to derive polarised 2D monolayers from enteroids while maintaining the presence of different cell types, greatly facilitates the apical application of nutrients and ingredients.

Future studies will be aimed at elucidating the components associated with these altered biological processes through purification of different molecular fractions of the protein sources and further functional studies. Additionally, advances made in human and mouse intestinal organoid research could be extrapolated to other host-species. This could be relevant for tailor-made in-depth molecular studies on feed effects on the species of interest.



## Acknowledgement

We thank Dr. M Boekschoten and H J Jansen of Human Nutrition, Wageningen University and Research for technical expertise and help with microarray scanning. The authors acknowledge the financial support from the Wageningen UR "IPOP Customized Nutrition" programme financed by Wageningen UR, the Dutch Ministry of Economic Affairs, Agriculture & Innovation, the graduate school (WIAS) and industrial partners Nutreco, The Netherlands and Darling Ingredients Internationals, The Netherlands.

## References


1. Ascencio, C., N. Torres, F. Isoard-Acosta, F. J. Gomez-Perez, R. Hernandez-Pando, and A. R. Tovar. 2004. Soy protein affects serum insulin and hepatic SREBP-1 mRNA and reduces fatty liver in rats. J Nutr 134: 522-529.
2. Ben-Ari Fuchs, S., I. Lieder, G. Stelzer, Y. Mazor, E. Buzhor, S. Kaplan, Y. Bogoch, I. Plaschkes, A. Shitrit, N. Rappaport, A. Kohn, R. Edgar, L. Shenhav, M. Safran, D. Lancet, Y. Guan-Golan, D. Warshawsky, and R. Shtrichman. 2016. GeneAnalytics: An Integrative Gene Set Analysis Tool for Next Generation Sequencing, RNAseq and Microarray Data. Omics-a Journal of Integrative Biology 20: 139-151. doi:10.1089/omi.2015.0168
3. Biesalski, H.-K., L. O. Dragsted, I. Elmadfa, R. Grossklaus, M. Muller, D. Schrenk, P. Walter, and P. Weber. 2009. Bioactive compounds: definition and assessment of activity. Nutrition (Burbank, Los Angeles County, Calif ) 25: 1202-1205. doi:S0899-9007(09)00225-1 [pii] 10.1016/j.nut.2009.04.023 [doi]
4. Bolstad, B. M., F. Collin, K. M. Simpson, R. A. Irizarry, and T. P. Speed. 2004. Experimental design and low-level analysis of microarray data. In: M. F. Miles (ed.) DNA Arrays in Neurobiology, Vol 60. International Review of Neurobiology No. 60. p 25-+.
5. Cao, L., A. Kuratnik, W. Xu, J. D. Gibson, F. t. Kolling, E. R. Falcone, M. Ammar, M. D. Van Heyst, D. L. Wright, C. E. Nelson, and C. Giardina. 2015. Development of intestinal organoids as tissue surrogates: cell composition and the epigenetic control of differentiation. Mol Carcinog 54: 189-202. doi:10.1002/mc.22089 [doi]
6. Couto, N., J. Wood, and J. Barber. 2016. The role of glutathione reductase and related enzymes on cellular redox homoeostasis network. Free Radic Biol Med 95: 27-42. doi:S0891-5849(16)00087-3 [pii] 10.1016/j.freeradbiomed.2016.02.028 [doi]
7. CVB. 2007. Chemical composition and nutritional values of feedstuffs and feeding standards No. CVB series nr. 36. Product Board Animal Feed, The Hague, The Netherlands.
8. Dai, M. H., P. L. Wang, A. D. Boyd, G. Kostov, B. Athey, E. G. Jones, W. E. Bunney, R. M. Myers, T. P. Speed, H. Akil, S. J. Watson, and F. Meng. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Research 33. doi:10.1093/nar/gni179
9. Date, S., and T. Sato. 2015. Mini-Gut Organoids: Reconstitution of Stem Cell Niche. In: R. Schekman (ed.) Annual Review of Cell and Developmental Biology, Vol 31. Annual Review of Cell and Developmental Biology No. 31. p 269-289.
10. Dekkers, J. F., C. L. Wiegerinck, H. R. de Jonge, I. Bronsveld, H. M. Janssens, K. M. de Winter-de Groot, A. M. Brandsma, N. W. M. de Jong, M. J. C. Bijvelds, B. J. Scholte, E. E. S. Nieuwenhuis, S. van den Brink, H. Clevers, C. K. van der Ent, S. Middendorp, and J. M. Beekman. 2013. A functional CFTR assay using primary cystic fibrosis intestinal organoids. Nature Medicine 19: 939-+. doi:10.1038/nm.3201

11. Diaz-Vivancos, P., A. de Simone, G. Kiddle, and C. H. Foyer. 2015. Glutathione--linking cell proliferation to oxidative stress. Free Radic Biol Med 89: 1154-1164. doi:S0891-5849(15)00608-5 [pii] 10.1016/j.freeradbiomed.2015.09.023 [doi]
12. Ertel, A., A. Verghese, S. W. Byers, M. Ochs, and A. Tozeren. 2006. Pathway-specific differences between tumor cell lines and normal and tumor tissue cells. Mol Cancer 5: 55.
13. Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome biology 5: R80. doi:gb-2004-5-10-r80 [pii] 10.1186/gb-2004-5-10-r80 [doi]
14. Iantomasi, T., F. Favilli, P. Marraccini, T. Magaldi, P. Bruni, and M. T. Vincenzini. 1997. Glutathione transport system in human small intestine epithelial cells. Biochimica et biophysica acta 1330: 274-283.
15. Jahan-Mihan, A., B. L. Luhovyy, D. E. Khoury, and G. H. Anderson. 2011. Dietary Proteins as Determinants of Metabolic and Physiologic Functions of the Gastrointestinal Tract. Nutrients 3: 574-603.
16. Jansman, A. J. M. 2016. Health and functions of the gastrointestinal tract in pigs: Effects of functional ingredients and feed and ingredient processing. J Anim Sci 94: 12-21. doi:10.2527/jas2015-9886
17. Kar, S. K., A. J. M. Jansman, S. Boeren, L. Kruijt, and M. A. Smits. 2016. Approximation of the amino acid composition and bio-functional properties of current and novel protein sources for pigs. J Anim Sci 94: 30–39. doi:10.2527/jas2015-9677
18. Koressaar, T., and M. Remm. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23: 1289-1291. doi:10.1093/bioinformatics/btm091
19. Kosti, I., N. Jain, D. Aran, A. J. Butte, and M. Sirota. 2016. Cross-tissue Analysis of Gene and Protein Expression in Normal and Cancer Tissues. 6: 24799.
20. Lin, K., H. Kools, P. J. de Groot, A. K. Gavai, R. K. Basnet, F. Cheng, J. Wu, X. Wang, A. Lommen, G. J. E. J. Hooiveld, G. Bonnema, R. G. F. Visser, M. R. Muller, and J. A. M. Leunissen. 2011. MADMAX - Management and analysis database for multiple ~omics experiments. Journal of integrative bioinformatics 8: 160-160. doi:10.2390/biecoll-jib-2011-160
21. Lin, Y., G. W. Meijer, M. A. Vermeer, and E. A. Trautwein. 2004. Soy protein enhances the cholesterol-lowering effect of plant sterol esters in cholesterol-fed hamsters. J Nutr 134: 143-148.
22. Loonen, L. M. P., E. H. Stolte, M. T. J. Jaklofsky, M. Meijerink, J. Dekker, P. van Baarlen, and J. M. Wells. 2014. REG3gamma-deficient mice have altered mucus distribution and increased mucosal inflammatory responses to the microbiota and enteric pathogens in the ileum. Mucosal Immunol 7: 939-947. doi:mi2013109 [pii]10.1038/mi.2013.109 [doi]

23. Moon, C., K. L. VanDussen, H. Miyoshi, and T. S. Stappenbeck. 2014. Development of a primary mouse intestinal epithelial cell monolayer culture system to evaluate factors that modulate IgA transcytosis. Mucosal Immunol 7: 818-828. doi:mi201398 [pii] 10.1038/mi.2013.98 [doi]
24. Moriyama, T., K. Kishimoto, K. Nagai, R. Urade, T. Ogawa, S. Utsumi, N. Maruyama, and M. Maebuchi. 2004. Soybean beta-conglycinin diet suppresses serum triglyceride levels in normal and genetically obese mice by induction of beta-oxidation, downregulation of fatty acid synthase, and inhibition of triglyceride absorption. Bioscience, biotechnology, and biochemistry 68: 352-359. doi:10.1271/bbb.68.352 [doi]
25. Mustata, R. C., G. Vasile, V. Fernandez-Vallone, S. Strollo, A. Lefort, F. Libert, D. Monteyne, D. Perez-Morga, G. Vassart, and M.-I. Garcia. 2013. Identification of Lgr5-independent spheroid-generating progenitors of the mouse fetal intestinal epithelium. Cell Rep 5: 421-432. doi:S2211-1247(13)00513-5 [pii] 10.1016/j.celrep.2013.09.005 [doi]
26. NRC. 2012. Nutrient Requirements of Swine. National Academic Press, Washington, D.C. 20418 USA.
27. Ootani, A., X. Li, E. Sangiorgi, Q. T. Ho, H. Ueno, S. Toda, H. Sugihara, K. Fujimoto, I. L. Weissman, M. R. Capecchi, and C. J. Kuo. 2009. Sustained *in vitro* intestinal epithelial culture within a Wnt-dependent stem cell niche. Nature medicine 15: 701-706. doi:nm.1951 [pii] 10.1038/nm.1951 [doi]
28. Radcliffe, J. D., V. L. Imrhan, and A. M. Hsueh. 1998. The use of soy protein isolate to reduce the severity of 13-cis retinoic acid-induced hypertriglyceridemia. Cancer Detect Prev 22: 526-532.
29. Randall, K. J., J. Turton, and J. R. Foster. 2011. Explant culture of gastrointestinal tissue: a review of methods and applications. Cell Biol Toxicol 27: 267-284. doi:10.1007/s10565-011-9187-5 [doi]
30. Rosch, C., N. Taverne, K. Venema, H. Gruppen, J. M. Wells, and H. A. Schols. 2017. Effects of *in vitro* fermentation of barley beta-glucan and sugar beet pectin using human fecal inocula on cytokine expression by dendritic cells. Mol Nutr Food Res 61. doi:10.1002/mnfr.201600243 [doi]
31. Sartor, M. A., C. R. Tomlinson, S. C. Wesselkamper, S. Sivaganesan, G. D. Leikauf, and M. Medvedovic. 2006. Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments. Bmc Bioinformatics 7. doi:10.1186/1471-2105-7-538
32. Sato, T., D. E. Stange, M. Ferrante, R. G. J. Vries, J. H. van Es, S. van den Brink, W. J. van Houdt, A. Pronk, J. van Gorp, P. D. Siersema, and H. Clevers. 2011a. Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. Gastroenterology 141: 1762-1772. doi:10.1053/j.gastro.2011.07.050
33. Sato, T., J. H. van Es, H. J. Snippert, D. E. Stange, R. G. Vries, M. van den Born, N. Barker, N. F. Shroyer, M. van de Wetering, and H. Clevers. 2011b. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature 469: 415-+. doi:10.1038/nature09637

34. Sato, T., R. G. Vries, H. J. Snippert, M. van de Wetering, N. Barker, D. E. Stange, J. H. van Es, A. Abo, P. Kujala, P. J. Peters, and H. Clevers. 2009. Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. Nature 459: 262-U147. doi:10.1038/nature07935
35. Schmittgen, T. D., and K. J. Livak. 2008. Analyzing real-time PCR data by the comparative C-T method. Nat Protoc 3: 1101-1108. doi:10.1038/nprot.2008.73
36. Sovran, B., L. M. P. Loonen, P. Lu, F. Hugenholtz, C. Belzer, E. H. Stolte, M. V. Boekschoten, P. van Baarlen, M. Kleerebezem, P. de Vos, J. Dekker, I. B. Renes, and J. M. Wells. 2015. IL-22-STAT3 Pathway Plays a Key Role in the Maintenance of Ileal Homeostasis in Mice Lacking Secreted Mucus Barrier. Inflammatory Bowel Diseases 21: 531-542.
37. Spence, J. R., C. N. Mayhew, S. A. Rankin, M. F. Kuhar, J. E. Vallance, K. Tolle, E. E. Hoskins, V. V. Kalinichenko, S. I. Wells, A. M. Zorn, N. F. Shroyer, and J. M. Wells. 2011. Directed differentiation of human pluripotent stem cells into intestinal tissue *in vitro*. 470: 105-109.
38. Standeven, A. M., R. L. Beard, A. T. Johnson, M. F. Boehm, M. Escobar, R. A. Heyman, and R. A. Chandraratna. 1996. Retinoid-induced hypertriglyceridemia in rats is mediated by retinoic acid receptors. Fundam Appl Toxicol 33: 264-271. doi:S0272059096901645 [pii]
39. Storey, J., and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100. doi:10.1073/pnas.1530509100
40. Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts, and J. A. M. Leunissen. 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research 35: W71-W74. doi:10.1093/nar/gkm306
41. VanDussen, K. L., J. M. Marinshaw, N. Shaikh, H. Miyoshi, C. Moon, P. I. Tarr, M. A. Ciorba, and T. S. Stappenbeck. 2015. Development of an enhanced human gastrointestinal epithelial culture system to facilitate patient-based assays. Gut 64: 911-920. doi:10.1136/gutjnl-2013-306651
42. Wang, F., D. Scoville, X. C. He, M. M. Mahe, A. Box, J. M. Perry, N. R. Smith, N. Y. Lei, P. S. Davies, M. K. Fuller, J. S. Haug, M. McClain, A. D. Gracz, S. Ding, M. Stelzner, J. C. Y. Dunn, S. T. Magness, M. H. Wong, M. G. Martin, M. Helmrath, and L. Li. 2013. Isolation and characterization of intestinal stem cells based on surface marker combinations and colony-formation assay. Gastroenterology 145: 383-395 e381-321. doi:S0016-5085(13)00701-4 [pii] 10.1053/j.gastro.2013.04.050 [doi]
43. Wells, J. M., R. J. Brummer, M. Derrien, T. T. MacDonald, F. Troost, P. D. Cani, V. Theodorou, J. Dekker, A. Meheust, W. M. de Vos, A. Mercenier, A. Nauta, and C. L. Garcia-Rodenas. 2017. Homeostasis of the gut barrier and potential biomarkers. Am J Physiol Gastrointest Liver Physiol 312: G171-G193. doi:ajpgi.00048.2015 [pii] 10.1152/ajpgi.00048.2015 [doi]
44. Xiao, C. W., M. R. L'Abbe, G. S. Gilani, G. M. Cooke, I. H. Curran, and S. A. Papademetriou. 2004. Dietary soy protein isolate and isoflavones modulate hepatic thyroid hormone receptors in rats. J Nutr 134: 743-749.
45. Xiao, C. W., J. Mei, W. Huang, C. Wood, M. R. L'Abbe, G. S. Gilani, G. M. Cooke, and I. H. Curran. 2007. Dietary soy protein isolate modifies hepatic retinoic acid receptor-beta proteins and inhibits their DNA binding activity in rats. J Nutr 137: 1-6. doi:137/1/1 [pii]

## Supplemental Information

**Supplementary Table 1. RT-qPCR primer sequences for microarray validation of significantly regulated genes in protein-treated enteroids compared to non-treated control (blank).**

| Gene | Primer forward | Primer reverse | AT* (°C) |
|---|---|---|---|
| *Stxbp1* | attttcatccttgggggtgt | aagtcggggtgtctcaggt | 60 |
| *TNFsf13b* | tgccttggaggagaaagaga | ccagccgagtagcaggaa | 60 |
| *Gm41* | cctgtcctgtttgctgctct | ctcctttctctcctgccttg | 61 |
| *Cyp1a1* | cagaaggtgatggcagaggt | ggtaacggaggacaggaatg | 60 |
| *Olfr1162* | tggaaagaaatgtgagtgtgg | tgatggttgagtagcagaagtc | 59 |
| *Lims2* | gcggattctgtggtgaattt | cttgaacatgaggggctgtt | 60 |

* Annealing Temperature

**Supplementary Table 2. RT-qPCR validation of microarray results based significantly regulated genes in protein-treated enteroids compared to non-treated control (blank). Gene expression levels were checked by performing quantitative PCR on protein treated RNA (n=3 per treatment) and compared to the blank.**

| | Relative to blank | | Correlation |
|---|---|---|---|
| Genes | Microarray | qPCR | |
| *Stxbp1* | 4.23 | 4.57 ± 1.75 | 0.80 |
| *TNFsf13b* | 2.13 | 2.19 ± 0.83 | |
| *Gm41* | 1.95 | 6.32 ± 2.33 | |
| *Cyp1a1* | 1.92 | 7.00 ± 0.15 | |
| *Olfr1162* | -1.82 | -1.20 ± 0.23 | |
| *Lims2* | -2.38 | -1.41 ± 0.21 | |

# CHAPTER 4

## Soybean meal in mice feed reduces expression of mTOR signaling pathways and strongly increases abundance of *Bacteroidales* family *S24-7*

*Soumya K. Kar[1,2], Alfons J. M. Jansman[3], Nirupama Benis[1,2,6], Javier Ramiro-Garcia[4-6], Dirkjan Schokker[2], Leo Kruijt[2,3], Ellen H. Stolte[1], Johanna J. Taverne-Thiele[1], Jerry M. Wells[1], Mari A. Smits[1,2,7]*

[1]Host Microbe Interactomics Group, Wageningen University & Research, Wageningen, the Netherlands
[2]Wageningen University & Research Animal Breeding and Genomics, Wageningen, The Netherlands
[3]Wageningen Livestock Research, Animal Nutrition, Wageningen University & Research, Wageningen, The Netherlands
[4]Laboratory of Microbiology, Wageningen University & Research, Wageningen, the Netherlands
[5]Top Institute Food and Nutrition, Wageningen, the Netherlands
[6]Laboratory of System and Synthetic Biology, Wageningen University & Research, Wageningen, the Netherland
[7]Wageningen Bioveterinary Research, Wageningen University & Research, Wageningen, the Netherlands

*Manuscript submitted for publication*

## Abstract

Dietary protein sources can have profound effects on host-microbe interactions in the gut that are critically important for immune resilience. However more knowledge is needed to assess the impact of different protein source on gut and animal health. Thirty-six wildtype male C57BL/6J mice of 35 d age (n=6/group; mean ± SEM body weight 21.9 ± 0.25 g) were randomly assigned to groups fed for four weeks with semi synthetic diet prepared with one of the following protein sources containing (300 g/kg as fed basis): soybean meal (SBM), casein, partially delactosed whey powder, spray dried plasma protein, wheat gluten meal and yellow meal worm. At the end of the experiment, mice were sacrificed to collect ileal tissue to acquire gene expression data, ileal digesta to study changes in microbiota and serum to measure cytokines and chemokines. By genome-wide transcriptome analysis, we identified fourteen high level regulatory genes that are strongly affected in SBM-fed mice compared to the other experimental groups. They mostly related to the mammalian (mechanistic) target of rapamycin (mTOR) pathway. In addition, an increased ($P < 0.05$) concentration of granulocyte colony-stimulating factor was observed in serum of SBM-fed mice compared to other dietary groups. Moreover, by 16S rRNA sequencing, we observed that SBM-fed mice had higher ($P < 0.05$) abundances of *Bacteroidales* family *S24-7*, compared to the other dietary groups. We showed that measurements of genome-wide expression and microbiota composition in the mouse ileum reveal divergent responses to diets containing different protein sources, in particular for a diet based on SBM.

**Keywords: dietary protein sources; mTOR pathway, T cells, microbiota, *S24-7*, granulocyte colony-stimulating factor; mice.**

## Introduction

For economical reasons soybean meal (SBM) is commonly used as a protein source in animal feed [1] increasing demand and price for SBM, is stimulating commercial interest in alternative sources of protein. Apart from the predicted nutritional properties of these novel protein sources, nothing is known about their potential effects on gut immunity and health.

Ingredients of animal feed may contain constituents that not only provide nutrients but also impact on diverse physiological functions, among which modulation of immunity is considered as one of the most important. Immune functions are indispensable for an efficient defense against pathogens contributing to the disease resistance and resilience of animals. Recently, mammalian (mechanistic) target of rapamycin (mTOR), a serine/threonine protein kinase, has been recognized as having a key role in the modulation of both innate and adaptive immune responses. mTOR participates in diverse signaling cascades linking physiological events to cell growth, proliferation and differentiation by stimulating *de novo* synthesis of proteins, nucleotides and lipids and by inhibiting autophagy [2]. Recent evidence suggests that mTOR signaling serves to couple metabolic activity to important cellular immune functions, for example, cytokine production by antigen presenting dendritic cells thereby determining the nature of effector T cell responses. Moreover, it has a vital role in the activation of effector T cells and their function as well as proliferation of regulatory T cells (Tregs), thereby having important immuno-regulatory functions [3]. The impact of diet on mTOR signaling in the intestine is not fully understood, not only with regard to the effector pathways, but also with respect mTOR sensing of novel nutritional and microbial metabolites.

The effects of dietary protein on performance and health are dependent on the source(s) of protein (e.g. milk; plasma; plant cereals, insects, algae) included in the diet, their digestibility in the gastro-intestinal tract, the matrix in which the protein fraction is incorporated and the nature and extent of technical processing to which these ingredients are exposed. Furthermore, the amino acid composition and sequences of the individual proteins may influence the kinetics of protein digestion and the generation of bioactive peptides along the gastro-intestinal tract (GIT) [4-6]. In addition, variations in the non-protein components of protein sources may have direct effects on the health and performance of hosts or indirect effects via modulation of the intestinal microbiota [7]. Dietary casein (CAS) and whey proteins are known to stimulate the immune system and promote host protection against allergies [8, 9]. Spray dried plasma protein (SDPP) is recognized as a high-quality feed ingredient for farm animals, including pigs and it has been recommended in animal diet as a source of immunological support due to its high level of globulin proteins, including immunoglobulins [10]. For many mammalian

species, SBM is the common source of dietary protein. It contains a range of proteins, carbohydrates and phytochemicals, such as non-starch polysaccharides (NSP) and isoflavones, which can influence the activity of the immune system [7]. Wheat gluten meal (WGM) is also considered as a good source of dietary proteins for many mammals, as it is highly digestible and regarded as an excellent source of glutamine and glutamic acid, amino acids which are known to play a role in the modulation of gut immunity [11]. Furthermore, yellow meal worm (YMW) seems to be a promising source of protein for human consumption as well as for monogastric animals based on protein digestibility and amino acid composition [12].

The objective of the present pilot study was to investigate and compare the effect of a set of conventional and unconventional dietary protein sources on a variety of intestinal and systemic health associated parameters. We hypothesized that the selected protein sources would differ in their capacities to affect the activity of mTOR regulated immune processes. We have deliberately chosen to study the effect of complex diet because it is known that purified bioactive compounds behave differently when included in complex dietary mixtures [13, 14]. In this study, experimental diets were prepared by replacing the corresponding reference ingredient (i.e. SBM) from the "reference" diet (i.e. SBM-diet) with CAS, partially delactosed whey powder (DWP), WGM, SDPP and YMW as sources of protein. Mice were fed these experimental diets for a period of four weeks. We have focused on the small intestine because it is the location where dietary nutrient digestion and absorption is expected to have most effects on the intestinal mucosal (immune) system. Besides the local effects, we also investigated possible effects of the different diets on several systemic immune markers in the blood.

## Materials and Methods

### *Animals*

All procedures were approved by the animal experimentation board at Wageningen University & Research Center (accession number 2012062.c) and carried out according to the guidelines of the European Council Directive 86/609/EEC dated November, 1986. A schematic representation and detailed description of the experimental design and sample collection is given in online supporting information (see S1 Fig). Briefly, thirty-six 21-day-old wild type male C57BL/6J mice (Harlan Laboratories, Horst, the Netherlands) were stratified according to bodyweight and litter of origin into 6 dietery groups (n=6/group) in the light and temperature-controlled animal facility of Wageningen University (12:12 h reversed light/dark cycle, 20 ± 2°C) upon arrival. The mice were housed in pairs in a specific pathogen-free environment with ad libitum access to diet and water. Prior to the start of the experiment mice were adapted for one week to a standard diet based on AIN-1993 growth (AIN-93G), which included 300 g/kg casein (CAS) as the

protein source (as fed basis). Thereafter, one group continued with the CAS-fed diet and the other five groups received similar semi-synthetic diets containing 300 g/kg (as fed basis) of one of the alternative protein sources (SBM, DWP, WGM, SDPP or YMW) for 28 days. Body weights of animals were measured every week. Thereafter, the animals were anaesthetized with isoflurane and sacrificed to collect samples. From the same location three segments of ileal tissue were taken, one was snap frozen in liquid nitrogen and stored at -80 C for gene expression studies, the second was fixed with methanol-carnoy's fixative for immunohistochemistry and the third tissue segment along with luminal content was collected to perform a community-scale analysis of gut microbiota. Blood samples were collected by orbital puncture and serum was extracted using 500 µl SST tubes (Becton Dickinson, Franklin Lakes, New Jersey) within 30 minutes after collection of the blood. The sera were stored at -20ºC for further analysis of cytokine levels. Soybean meal (SBM) diet served as reference to make comparisons with other experimental diets for all analysis, as it is the most widely used protein source in animal feed [15, 16].

***Diet***

Customized semi-synthetic diets based on AIN-93G were prepared replacing casein with any one of five other protein sources *i.e.* SBM, DWP, WGM, SDPP and YMW at an inclusion level of 300 g/kg [17]. Representative samples of dried and ground diets were chemically analysed for dry matter (DM; NEN-ISO 6496 by 4 hours drying at 104°C), nitrogen (N; NEN-ISO 5983-2 by Kjeldahl method and crude protein calculated as Nx6.25), ash (NEN-ISO 5984 after 3 hours ashing at 550°C), ether extract (EE, NEN-ISO 6492 by extraction with petroleum ether) and gross energy (GE; NEN- EN-ISO 9831 by bomb calorimetry). Ingredient and chemical compositions of the experimental diets are presented in supporting information (S1 Table).

***Gene expression***

Total RNA extraction from ileal tissue samples, labelling, hybridization of individual samples on Affymetrix GeneChip mouse gene 1.1 ST arrays (Affymetrix, Santa Clara, CA, USA), scanning, quality control and normalization of the resulting datasets was performed as described previously [18] and the data is available in the Gene Expression Omnibus from NCBI with the accession number GSE84442. The output was used for Gene Set Enrichment Analysis (GSEA) [19] with human official gene symbols in which each experimental diet was compared to the diet containing SBM with permutations on gene sets. InteractiVenn [20] was used to visualize significant GSEA results (FDR < 0.05). Subsequently, we defined a set of common core genes, *i.e.* the genes that are enriched in the significant differential gene-sets common to all five comparisons. The common core genes were used to build two types of networks in Cytoscape [21]: a network of GO terms restricted to terms with FDR < 0.001 (using the app BINGO) [22]; and a

Functional Interaction (FI) network (using the app Reactome FI) [23]. In the BiNGO network, nodes are GO terms and edges are relations between them, the network was restricted to terms with FDR < 0.001. The FI network has genes as nodes and the edges are interactions between the genes (from literature or predictions). FI Nodes with a degree (number of directly connected nodes) greater than 20 were considered as hubs.

***Microbiota***

DNA was isolated from snap frozen intestinal segments and the bacterial 16S rDNA V3 region was sequenced by targeted-amplicon 16S sequencing on a Illumina Mi-Seq sequences as previously described [24]. The 16S rRNA gene sequencing reads were analyzed using an in-house pipeline [25]. Shortly, paired-end libraries were filtered to contain only sequence read pairs with perfectly matching primer and barcodes. Resulting sequence reads were separated by sample using the barcodes and operational taxonomic units (OTUs) were assigned using an open reference database and a customized SILVA 16S rRNA reference database [26]. Microbial composition was generated using a workflow based on quantitative insights into microbial ecology (QIIME) v1.2 [27]. The microbial groups that had a $P < 0.05$ in one of the diets vs SBM were considered significant.

***Immunohistochemistry***

For positive CD3-positive T cells, paraffin sections (5 µm) of fixed tissue were attached to poly-L-lysine-coated glass slides (Thermo scientific, Germany). Sections were heated for 20 minutes in 0.01 M sodium citrate (pH 6.0) at 100˚C, washed 2 times for 15 minutes with TRIS-Buffered Saline-triton (TBS-t) and then incubated for 30 minutes at room temperature in 5% (v/v) goat serum (Invitrogen, Life technologies Ltd, Paisley, UK) in TBS. Briefly, the T cell marker CD3 was detected by incubating the sections with anti-CD3 antibody (Invitrogen, Life technologies Ltd, Paisley, UK) diluted 1:100 in TBS-t, overnight at 4°C, washing 2 times for 10 minutes with TBS-t and incubation with secondary anti-body GhR/HRP diluted 1:200 in TBS-t for 60 minutes at room temperature. To detect the secondary antibody, the sections were washed for 10 minutes with two changes in TBS-t followed by 10 minutes rinsing in TRIS-HCl buffer (pH 7.6) and incubated with diamino-benzidine. Finally the slides were immersed for one second in hematoxylin (1:1) and then immediately rinsed under running tap water for 10 minutes. Digital images of transverse sections of the ileum were used to enumerate the number of brown-stained CD3-positive T cells using ImageJ software (NIH, Maryland, USA). T cells were counted in 10 villi per section using 2 sections per animal and 4 animals per dietary group.

For mTOR protein detection, we followed a similar protocol that was followed for CD3-positive cells with the following exceptions. Briefly, sections were heated for 20 minutes in 0.01 M sodium citrate (pH 6.0) at 100˚C, washed 2 times for 15 minutes with TRIS-Buffered Saline-triton (TBS-t) and then incubated for 30 minutes at room

temperature in 5% (v/v) rabbit serum (Invitrogen, Life technologies Ltd, Paisley, UK) in TBS. The mTOR protein was detected by incubating the sections with anti-mTOR antibody (Abcam, Cambridge, UK) diluted 1:500 in TBS-t, overnight at 4°C, washing 2 times for 10 minutes with TBS-t and incubation with secondary anti-body goat-anti-rabbit-biotin (Invitrogen, Life technologies Ltd, Paisley, UK) for 60 minutes at room temperature. Avidin-HRP diluted 1:200 in TBS-t was used and kept for 60 minutes at room temperature to detect the secondary antibody. Thereafter the sections were washed for 10 minutes with two changes in TBS-t followed by 10 minutes rinsing in TRIS-HCl buffer (pH 7.6) and incubated with diamino-benzidine. Finally, the slides were immersed in hematoxylin (1:1) and then immediately rinsed under running tap water for 10 minutes. Digital images of transverse sections of the ileum were used to enumerate the reaction in the image.

***TLR signalling assay***

Endotoxins were extracted from fecal samples as described before [28] and stored at -20 °C. TLR signalling assays were performed using stably transformed cell lines of human embryonic kidney cells (HEK293) expressing TLR 4 and transfected with a reporter plasmid (pNiFTY) [29]. Briefly, human embryonic kidney (HEK) cell lines were seeded at approximately 5 × 105 cells/cm2 into 96-well plates and incubated overnight under standard culture conditions. Samples of purified LPS or soluble LPS, extracted from mouse faeces (325 ng/ml) were incubated with the cell reporter lines for 6 h, after which time the luciferase activity was measured using the Bright-glo luciferase assay (Promega Benelux b.v. NL).

***Cytokine and chemokine profiles***

Serum cytokine and chemokine concentrations (pg/ml) were measured using a Bio-Rad Mouse 23-plex kit (Bio-Rad, Hercules, CA, USA). Calibration curves from recombinant cytokine and chemokine standards were prepared for the 8-point standard dilution set with 4-fold dilution steps in sterile PBS. The samples were measured using a Bio-Plex MagPix Multiplex Reader (Bio-Rad Laboratories Inc. by the Luminex Corporation, The Netherlands). The Bio-Plex Manager software's five-parameter logistic curve fitting (5PL) method was used for raw data analysis and calculation of cytokine concentrations.

***Statistical analysis***

Results of immunohistochemistry, TLR assay and cytokines are presented as means ± SEM. Statistical analysis was performed by one-way ANOVA followed by post hoc test (Dunnett test: compared all treatment vs. SBM group as control) using GraphPad prism version 5.03 for Windows Vista (GraphPad Software, San Diego, California, USA). *P* value < 0.05 was considered significant.

## Results

### *Diet, animal characteristics and feed intake*

Nutrient requirements of the mice were based on AIN-93G and varied in nutrient composition, particularly for crude protein, crude fat, sugar, NSP, Ca, P, K, Na, Cl and electrolyte balance (S1 Table). The animals appeared to be healthy throughout the experimental period of 4 weeks. At the start of the dietary intervention (age 28 days), mean body weights of other experimental groups were not significantly different compared with the SBM group. At the end of the experiment, mean body weight of SBM fed mice deviated strongly from the DWP group and not from the other experimental groups (S2 Fig). Strikingly, in the YMW-fed group, we measured a significantly lower feed intake compared to the SBM fed group (S2 Fig), while the body weight over the experimental period did not differ compared to the SBM fed group.

### *Differential gene expression in the ileum*

We started the gene expression analyses in GSEA by a one-to-one comparison of all possible dietary combinations. When we took SBM group as "reference", the highest number of significantly different gene sets were observed in each observed group. The number of significantly regulated gene sets as determined by GSEA is presented in Table 1. When compared to SBM-fed mice, we found 4, 23, 8 and 4 unique differently expressed gene-sets (FDR < 0.05) in CAS, DWP, SDPP and YMW, respectively. These gene sets mainly correspond to essential cellular processes such as cell cycle, cellular metabolism (anabolic and catabolic) and immune response (Fig 1; S2 Table).

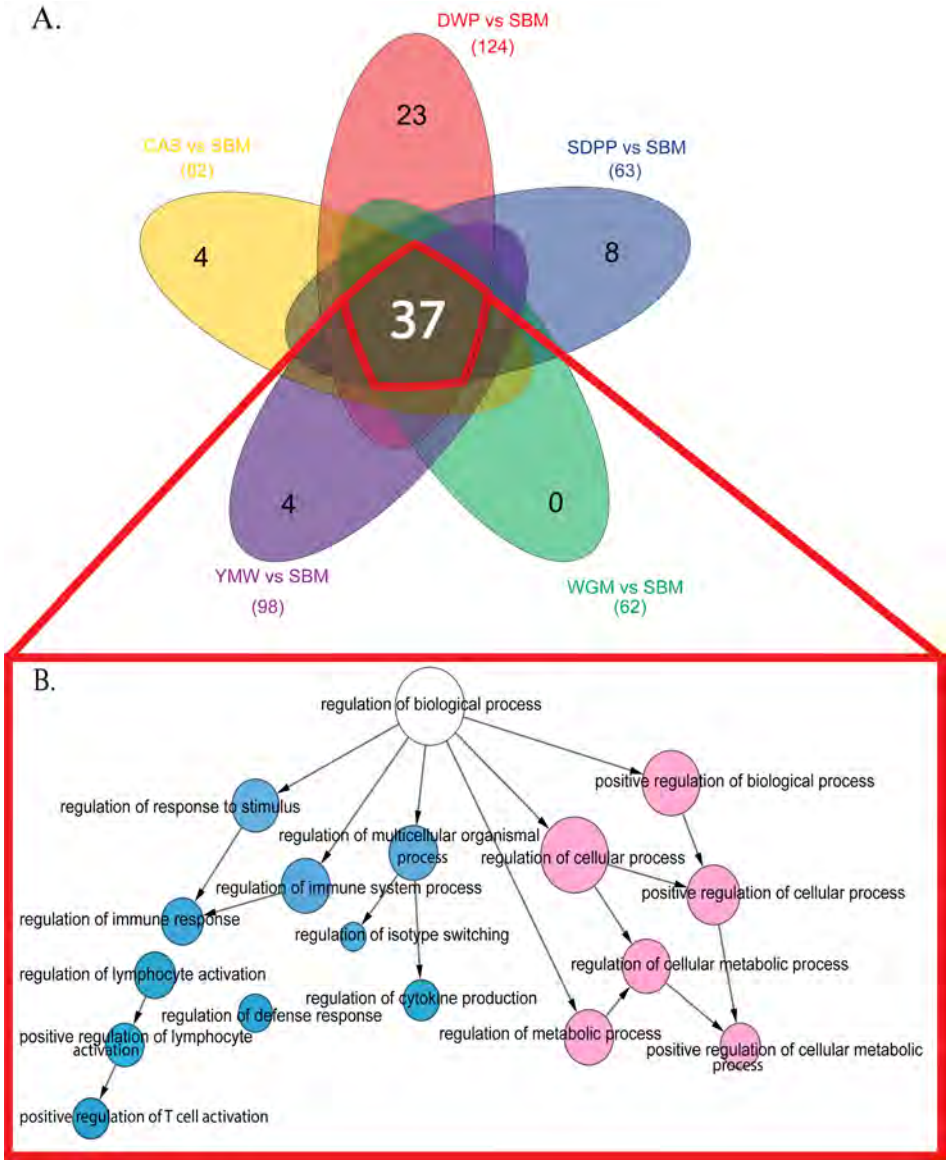



**Fig 1 Venn diagram featuring common and unique enriched gene-set expression patterns.**
(A) Colored spheres represent different experimental diets. The white number at the core of the Venn diagram represents the number of common overlapping gene-sets and, black numbers towards the periphery of each sphere represent the number of unique gene-sets. (B) Functional interaction network for the common overlapping gene-sets. The colored nodes denote the GO term significantly (FDR < 0.001) overrepresented in the enriched gene-sets. The edges represent interactions between gene-set as determined by BiNGO. Sky blue color nodes relate to immune processes and light pink color nodes relate to metabolic processes. Arrows represent directed interactions. The diameter of the nodes represents the number of genes associated with that particular GO term. SBM, soybean meal; CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.

**Table 1. Differential gene expression and number of enriched gene-sets in ileal mucosa of mice fed experimental diets with indicated protein sources relative to a diet with SBM.**

| Experimental diets vs SBM | Upregulated gene-sets | Down regulated gene-sets | Significantly enriched gene-sets (FDR < 0.05)* |
|---|---|---|---|
| CAS | 559 | 0 | 82 |
| DWP | 561 | 0 | 124 |
| SDPP | 591 | 0 | 63 |
| WGM | 516 | 0 | 62 |
| YMW | 559 | 0 | 98 |

CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm; SBM, soybean meal.

* FDR value was calculated in gene set enrichment analysis (GSEA) for indicated protein sources relative to a diet with SBM.

We identified a large number (thirty-seven) of overlapping gene-sets, which were significantly (FDR < 0.05) differentially expressed across all the experimental diets in comparison with SBM (Fig 1). We calculated an interaction network from these 37 gene-sets and observed down regulation (FDR < 0.001) of a number of immune and metabolic processes in ileal mucosae of the mice kept on SBM-fed diet relative to all other diets (Fig 1). In the next step, we analysed the degree distribution of the Reactome FI network and identified a total of 14 hub genes, which are indicated in Table 2 along with their pathways and biological functions involved. Strikingly, the mTOR pathway and biological processes related to T cell functioning and antigen presentation are heavily represented.

**Table 2. Key or hub genes of the functional network downregulated in mice fed soybean meal (SBM)-based diet.**

| Hub gene | Degree* | Related pathway | Biological relevance | References |
|---|---|---|---|---|
| Lck | 38 | PI3K/AKT/mTOR, MAPK | Selection and maturation of developing T cells. | [30] |
| Fyn | 36 | PI3K/AKT/mTOR, MAPK | Regulation of cell growth and adhesion humoral immune response. | [31] |
| Il2rg | 27 | PI3K/AKT/mTOR | Critical for intestinal T cell reconstitution | [32] |
| Cd4 | 27 | TGF β | Augment the early phase of T cell activation. Also found in B cells, macrophages and granulocytes. | [33] |

| Hub gene | Degree* | Related pathway | Biological relevance | References |
|---|---|---|---|---|
| Cd3e | 27 | Class I MHC mediated antigen processing and presentation, NF-κB Family Pathway | Antigen recognition by T cell and T cell development. | [34] |
| Cd3g | 25 | Class I MHC mediated antigen processing and presentation, PI3K/AKT/mTOR | Antigen recognition by T cell and T cell development. | [35] |
| Cd3d | 24 | Class I MHC mediated antigen processing and presentation, PI3K/AKT/mTOR | Antigen recognition by T cell and T cell development. | [36] |
| Pik3cg | 24 | PI3K/AKT/mTOR | T cell activation and differentiation. Structural and functional integrity of epithelium. | [37] |
| Stat1 | 24 | PI3K/AKT/mTOR, Signaling by FGFR | Important for cell viability in response to different cell stimuli and pathogens. | [38] |
| Lcp2 | 23 | PI3K/AKT/mTOR, Signaling by FGFR | Promotes T cell activation and development as well as mast cell and platelet function. | [39] |
| Cd8a | 21 | Class I MHC mediated antigen processing and presentation, TCR signalling | Selection and maturation of developing T cells. | [40] |
| Ptprc | 21 | Class I MHC mediated antigen processing and presentation | Regulates cell growth, differentiation, mitosis. Essential regulator of T- and B-cell antigen receptor signaling. | [41] |
| Zap70 | 21 | Class I MHC mediated antigen processing and presentation, Ras signaling pathway | T cell development and lymphocyte activation. | [42] |
| Btk | 21 | PI3K/AKT/mTOR, Signaling by FGFR | B-cell development. | [43] |

* Degree is the number of directly connected nodes within functional interaction (FI) network. PI3K/AKT/mTOR, Phosphoinositide 3-kinase/Protein kinase B/mammalian (mechanistic) target of rapamycin; MAPK, mitogen-activated protein kinases; TGF, transforming growth factor; MHC, major histocompatibility complex; NF-κB, nuclear factor kappa-light-chain-enhancer of activated B cells; FGFR, fibroblast growth factor receptors; TCR, T cell receptor.

### *Microbiota profiling in the terminal ileum*

To investigate the dietary effects on intestinal microbiota, we performed 16S rRNA gene sequencing on ileal digesta of six mice per group. Details of sequence reads count and the number operational taxonomic unit (OTU) are given in online supporting material (S3 Fig). Further analysis revealed that the microbial taxon profiles were considerably different among all dietary treatments (Fig 2 and 3). Hierarchical clustering analysis indicated that the microbial profile of SBM-fed mice was significantly different ($P < 0.05$) from the other dietary groups (Fig 2). Furthermore, the *Firmicutes* to *Bacteroidetes* ratio of SBM-fed mice was inverted as compared to all the other experimental diets fed mice (Fig 3C). Since these phyla constitute up to 90% of the microbial population, we analysed their family structure in more detail.

Within the *Bacteroidetes*, the proportion of the *S24-7* family was largely increased (Fig 3D) in SBM and WGM-fed mice. Varied level of percentage of total microbiota among different taxon were observed in mice fed different diets. The *Erysipelotrichaceae* family (*Firmicutes*) was reduced by 16.4% and 34.5% in the SBM group as compared to the DWP and SDPP groups, respectively (Fig 3E). The *Bifidobacteriaceae* family

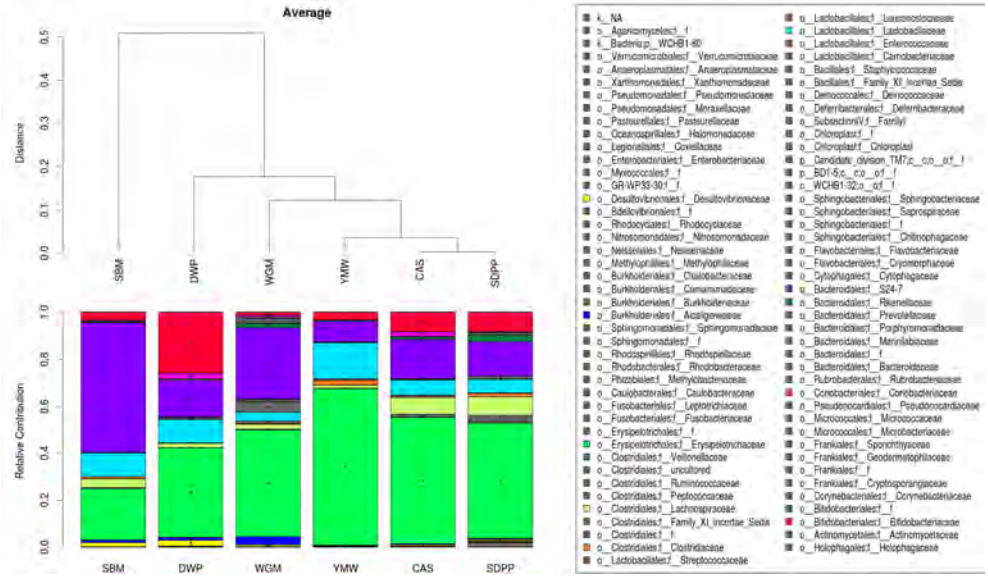



**Fig 2. The ileal microbiome composition profiles based on the relative abundance at the family level.**

Bacterial families are represented with different colors (shown in key). Hierarchical clustering of the microbial family composition is indicated above the composition profile. * Represents significant difference ($P < 0.05$) of microbiota at family level compared to SBM-fed mice. Members of microbial family belonging to 'others' are listed in S3 Table. SBM, soybean meal; CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.

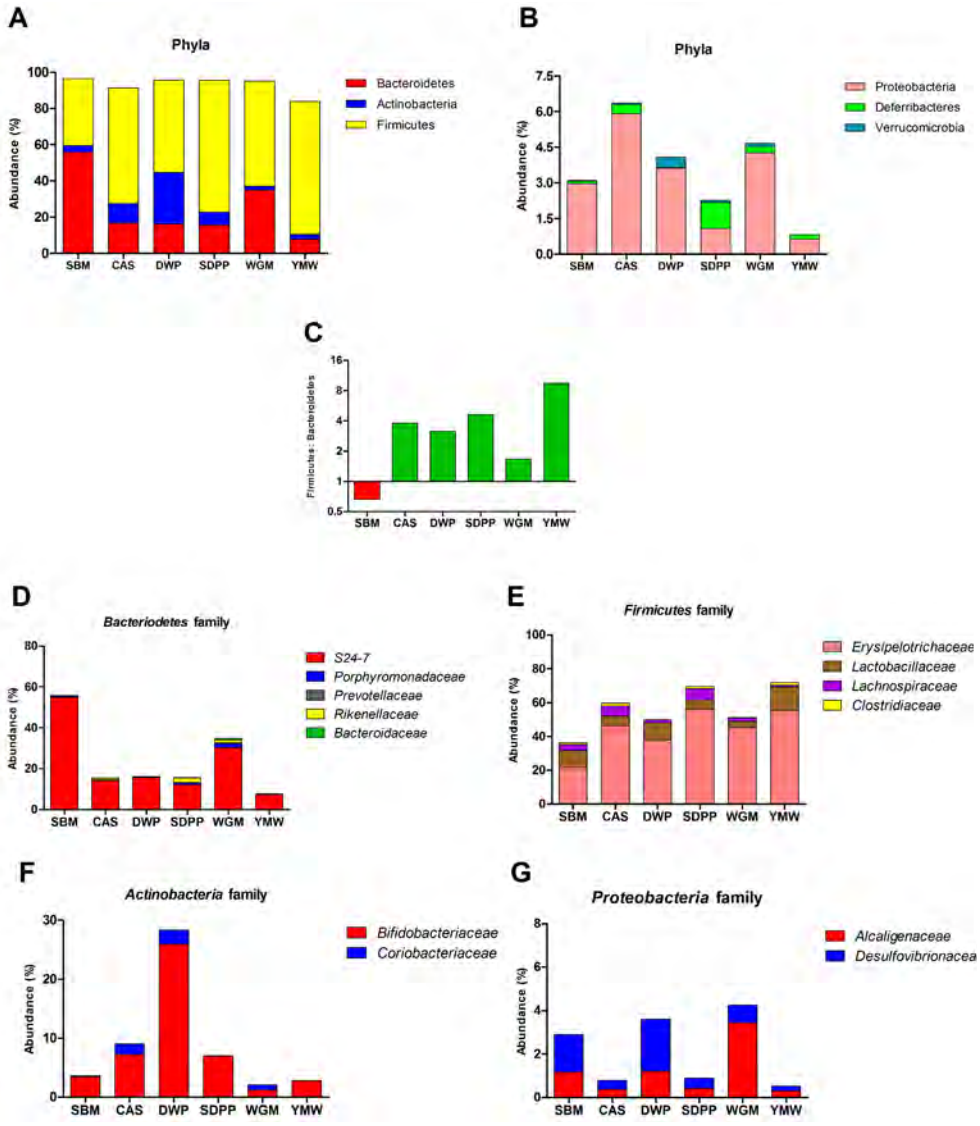



**Fig 3. Ileal microbial profiles of mice fed diets containing protein from different sources.**
(A) and (B) show the microbiota distribution at the phylum level; (C) shows the *Firmicutes/Bacteroidetes* ratio in the various groups of mice; (D-G) show the microbiota distribution at the family level. Bacterial abundance is shown as a percentage of the total 16S RNA gene sequences per group. SBM, soybean meal; CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.

(*Actinobacteria*) was increased by 22.5%, 3.8% and 3.5% in the DWP, CAS and SDPP groups as compared to the SBM group (Fig 3F). In the phylum *Proteobacteria*, the *Alcaligenaceae* family only increased in the WGM group by 2.3% compared to the SBM group (Fig 3G). The proportions of several other families were also affected by the various diets (Fig 2 and 3).

***Ileal immunohistochemistry***

In ileal sections, the WGM-fed mice had a significant higher count of positive $CD3^+$ T cell (S4 Fig) compared to SBM-fed mice. The lowest number of CD3+ T cell cells were found in the ileum of mice fed the SBM diet but this was not significant ($P = 0.06$) compared to other diets, except for the mice fed WGM diet. By histological staining mTOR was not detectable in the ileal section of mice fed SBM diet and was comparable to control tissues processed without the primary antibody. In contrast the tissue from CAS, SDPP, DWP and YMW groups stains strongly for mTOR; only weak staining was evident in tissue from WGM fed mice (Fig 4).

***Systemic cytokines and chemokines***

The concentrations of serum cytokines and chemokines were considerably differed among all the dietary treatments (S5 Fig). Importantly, that received the other experimental diets (Fig 5). The serum concentration of EOTAXIN, granulocyte-colony stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), interferon gamma (IFN-γ), interleukin-2 (IL-2), IL-5, IL-6, IL-12p70, IL-13, monocyte chemotactic and activating factor-1 (MCP-1) and macrophage inflammatory protein-1β (MIP-1β) were significantly different ($P < 0.05$) in at least one of the experimental dietary groups as compared to SBM (S5 Fig). Furthermore, significant lower ($P < 0.05$) concentration of GM-CSF, IL-6, IL-13, MCP and MIP-1β were observed in mice fed with WGM based diet compared to SBM.

***TLR4 assay***

The SBM group had the lowest ratio of *Firmicutes* to *Bacteroidetes* phylum. It is known that the *Firmicutes* phylum contains mostly Gram-positive bacteria, suggesting a difference in the abundance of LPS between the experimental groups. Therefore we investigated whether the feces of the dietary groups differed in their TLR4 signaling capacity using a TLR reporter cell assay. The results showed that only the TLR4 signaling activity of the feces of the YMW group significantly differed from the other dietary treatment groups (S6 Fig).

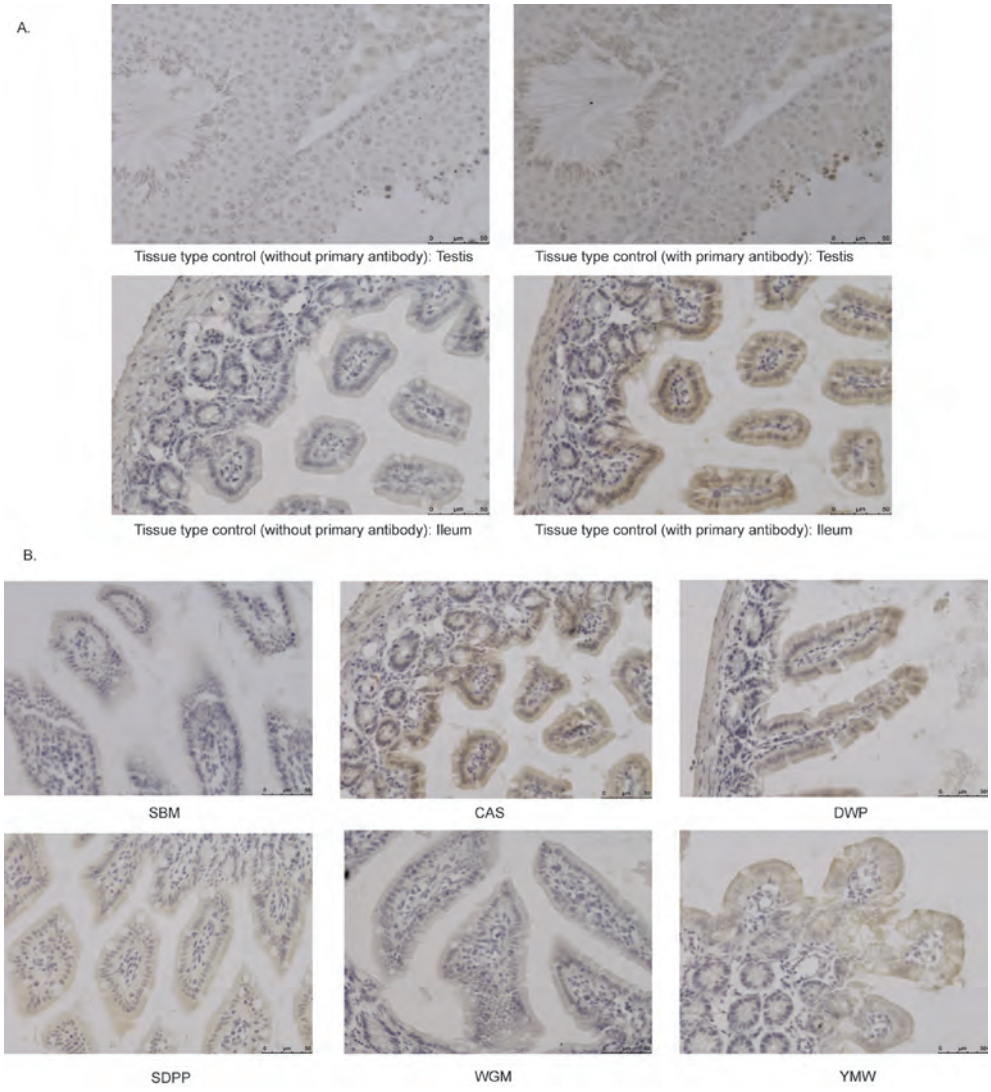



**Fig 4: Immunohistochemistry of PFA-fixed paraffin-embedded tissue sections with anti-mTOR antibody in ileal tissue of mice fed with different experimental diet.**

(A) Brown color indicates positive reaction in the image of tissue type control (testis) and a positive ileal tissue section (CAS). No positivity was observed in the same tissue by withholding the primary antibody (i.e. anti-mTOR antibody) during the staining procedure. (B) Immunohistochemistry of PFA-fixed paraffin-embedded ileum sections with anti-mTOR antibody of mice fed with different experimental diet. Brown color in the tissue section indicates positive reaction. SBM, soybean meal; CAS, casein; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm. Scale bar: 50 µm

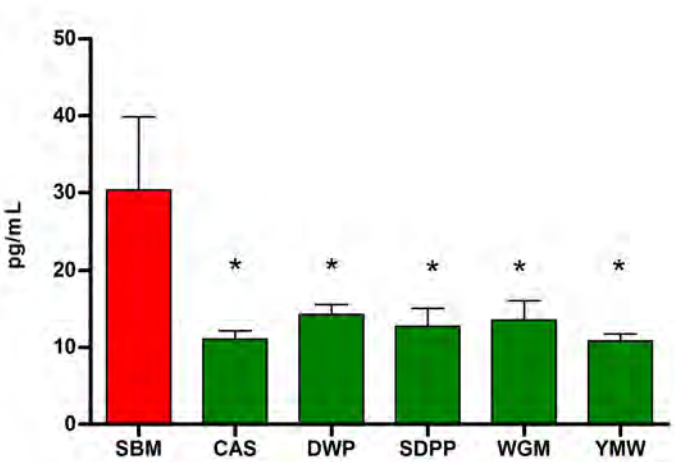



**Fig 5. Concentrations of serum granulocyte-colony stimulating factor (G-CSF) of mice fed diet containing different dietary proteins.**

Bars and whiskers represents means values ± SEM (n = 6), recorded at the end of experiment. *$P < 0.05$ compared to SBM-fed mice fed. SBM, soybean meal; CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.

## Discussion

The variability in the nutrient composition of the experimental diets are related to the differences in nutrient composition of the protein sources, included at fixed proportions in the diets [17]. As the experimental diets were not formulated to be iso-energetic and equal in all other nutrients, we cannot draw any firm conclusion on performance parameters as measured in this study, although there were some noteworthy differences. DWP-fed mice showed a significantly lower feed intake compared to SBM-fed mice and had a lower body weight gain over the experimental period as compared to the other groups. This agrees with previous observations that whey proteins, compared to CAS, increase the plasma levels of cholecystokinin and glucagon- like peptide-1 and thereby induce satiety [44]. A possible explanation for the reduced body weight in the DWP-fed group could be the high electrolyte balance compare to other experimental diets, related to a high concentration of potassium and high concentrations of other mineral, especially calcium, in the DWP-based diet. Similarly Pilvi and colleagues [45] showed that a high calcium diet with whey protein decreased body weight gain in high-fat-fed C57Bl/6J mice.

Funkat and colleague reported that mice fed high fat diet gain more body weight than mice fed a standard chow diet [46]. We speculate that the higher proportion of fat in YMW (160 g/kg vs. < 90 g/kg for the diet in the other treatments), may have promoted

satiety levels via one of the gut lipid sensing system resulting in the observed lower feed intake [47]. This might explain the result of YMW-fed mice observed in our study. However, further research is needed to confirm these results.

In this study we have shown that diets prepared with different protein sources, differ in their ability to modulate host physiology including immunological parameters. Apart from the diet-specific ileal responses, it was remarkable to observe that the ileal transcriptome and microbiota composition of the SBM-fed mice deviated strongly from the groups of mice given the other five protein sources. In comparison to the other protein sources, SBM-based diet down-regulated 14 hub genes in gene networks associated with antigen presentation, mTOR signalling and TGF $\alpha$ expression. Functionally these genes are of particular relevance to the activation, differentiation and proliferation of T cells along with B cells and antigen presenting cells (Table 2).

Activation of mTOR pathway plays a key role in shaping and controlling the effector responses of immune cells associated with innate and adaptive immune responses through coupling these events to intracellular metabolic status and environmental nutrients [3]. For example, TLR signalling activates the mTOR pathway in monocytes, macrophages and DCs, leading to activation, migration and production of cytokines and chemokines, which are all metabolically demanding processes. Similarly, antigen presentation to T cells activates mTOR which, in turn, is important for activation, proliferation and cell fate. Thus the concomitant down-regulation of Cd3 subunit genes, which are important in the formation of T cell receptor-CD3 complexes [34] as well as T helper (CD4) and cytotoxic T cell markers (CD8), seems to be a logical consequence of the reduction of mTOR signalling in the SBM group of mice. Moreover, we have observed no positive reaction when stained for mTOR protein with anti-mTOR antibody in the ileal section of mice fed the SBM diet. All these observations together, suggest that the SBM-based diet inhibited the expression of the mTOR protein, consequently dampening the local development of effector T cell subsets. This also explains the result in the ileum of mice fed the YMW diet were we noticed relatively low (in contrast to other diets but higher than SBM) numbers of CD3+ T cells along with weak positive reaction for mTOR protein. There was indeed a trend for reduced numbers of $CD3^+$ T cells in intestinal sections of the SBM-fed mice compared to mice fed the other diets ($P = 0.06$, except for the mice fed WGM diet), which is consistent with generated microarray gene expression data of this segment.

To investigate whether the effects of dietary protein sources on antigen presentation, mTOR signalling pathways in the mucosa also affected systemic immune markers in the serum, we measured the concentration of 23 cytokines and chemokines. Mice fed the SBM based diet had significantly higher levels of granulocyte colony-stimulating factor

(G-CSF) in the serum than any of the other dietary groups (Fig 5). G-CSF is mainly produced by bone marrow cells, innate cells such as tissue macrophages, but is also expressed constitutively in many tissues including the intestine, suggesting that it is of relevance to intestinal homeostasis [48]. G-CSF has been reported to have immuno-regulatory functions on monocytes, macrophages and dendritic cells by down-regulating the expression of inflammatory cytokines such as TNF and IL-12 [49]. Both $CD4^+$ and $CD8^+$ T cells also express a functional G-CSF receptor and G-CSF is a strong immune regulator modulating the expression of GATA-3, the positive regulator of Th2 responses and IL-4 expression. Moreover, the cytokine IL-4 inhibits production of the Th1 cytokine IFN-γ [50]. G-CSF has also been shown to favour the generation of Tr1-like Tregs *in vitro* and *in vivo* in mice [51, 52]. Recently, G-CSF was shown to promote the generation of Gr-1high/F4/80- macrophage-like cells in bone marrow cell cultures *in vitro*, which have an "M2-like" non-inflammatory phenotype [53]. When these cells were adoptively transferred to mice they showed a gut-homing phenotype. Moreover, the lamina propria of the gut of G-CSF-/- mice has reduced numbers of Gr-1high/F4/80- macrophages suggesting a key role for circulating G-CSF in generation and homing of M2-like macrophages to the gut which play a key a role in maintaining an anti-inflammatory tone in the intestine [53].

Based on the known functions of G-CSF and the observation of significantly elevated serum concentration of G-CSF in the SBM diet only, we speculate that the lamina propria of the small intestine of these mice may comprise more T regs and reduced activity of pro-inflmmatory cells, thereby potentially explaining the reduced expression of immune response gene sets observed in the SBM-fed mice transcriptome. Altered activity of these cellular immune pathways would also account for the observed reduced activity of the mTOR and lymphocyte maturation and differentiation pathways in the small intestinal transcriptome along with lower positive T cell count in the ileum of the mice on the SBM diet. Although we anticipated the effects of the SBM diet to be most pronounced in the mucosa, there were some indications that the significantly elevated G-CSF in the serum might impact on systemic immunity. For example, in mice fed DWP and YMW, IFN-γ a Th1 cytokine was significantly increased in the serum compared to SBM fed mice (S5 Fig). Whether the systemic G-CSF is produced in the bone marrow or in the ileal tissue of the SBM fed mice need to be further explored.

As the effects of the SBM diet on mTOR and immunity pathways in the ileal mucosal might be influenced by crosstalking microbiota, we compared the ileal microbiota composition of each dietary group. The composition of the ileal microbiota was strongly influenced by the inclusion of different protein sources in the diet, at least at the microbial phylum and family level. The most substantial and significant difference among the dietary groups was observed within the SBM group of mice where the ratio of *Firmicutes* to *Bacteroidetes*

phyla was inverted compared to the other dietary groups (Fig 3C). *Bacteroidetes* was the most abundant phylum (~55.9%) in the SBM fed group of mice followed by WGM (34.8%) and lowest abundance was seen in the YMW group (7.7%) (Fig 3D). Within the *Bacteroidetes* phylum, the genus *Bacteroides* are known to possess a large number of genes encoding for enzymes involved in the degradation and fermentation of a variety of different carbohydrates [54]. Traditionally, SBM is considered to contain several complex carbohydrates, which includes NSP. Thus the NSP from SBM may have caused the blooming of members in *Bacteroidetes* phylum, the most significant being the *S24-7* family.

Although the mice fed with YMW based diet consumed significantly less feed compared to mice in the SBM group, the body weight of the mice in both groups were similar. An increase of the *Firmicutes* to *Bacteroides* ratio in YMW fed mice (Fig 3C) is a signature pattern found in studies on obese mice [55]. An increased *Firmicutes* to *Bacteroides* ratio in mice has been associated with enhanced energy extraction from the diet leading to adiposity and weight gain in weaned and adult mice [56]. The DWP-fed mice showed an increase of microbes belonging to the *Actinobacteria* phylum and mainly driven by bacteria belonging to *Bifidobacteriaceae* family. This is most probably due to the high sugar (lactose and oligosaccharides) content of the diet prepared with DWP. The characteristic of bifidobacterial phylotypes in that they have a high capacity in utilizing milk oligosaccharides as an energy source for their growth. It has already been shown that infant-type *Bifidobacteria* efficiently metabolize several small mass milk oligosaccharides and that they possess large gene clusters encoding for enzymes involved in (milk) oligosaccharides metabolism [57].

As the abundance of *Bacteroidetes* phylum, comprising mainly of Gram-negative bacteria, was substantially increased in the SBM-fed mice, we investigated whether this might lead to differences in TLR4 signaling activity in the intestinal lumen of the different dietary groups. Despite the higher abundance of ileal Gram negatives (*Bacteroides*) in the SBM-fed mice, we did not observe a higher TLR4 stimulating activity in their feces. One explanation could be that the structure of the LPS derived from certain members of the *Bacteroidetes* phylum does not signal through TLR4 or may actively inhibit TLR4 mediated NF-κB activation [58]. Another possibility is that fecal microbiota do not mimic the ileal microbiota with respect to their *Firmicutes* to *Bacteroides* ratio.

Although transcriptomics on intestinal tissue in SBM-fed mice showed that there was a relative decrease in activation of the mTOR pathway, which are mainly involved in coupling of metabolic activity to important cellular immune functions, we were not able to determine whether this is caused by specific components in SBM or indirectly by the effects of the SBM based diet on the intestinal microbiota. The increased serum levels

of G-CSF, however, may be linked to the reduced expression of immune response gene sets observed in the transcriptome of SBM-fed mice due to its known role in promoting Tregs and non-inflammatory cells. In future, microbiota transfer experiments in germ-free mice and studies with specific fractions of the SBM diet may help to elucidate the cause of the observed effects.

## Conclusion

In this study we clearly demonstrated that the protein source used for the preparation of mice diets greatly affects local and systemic immune parameters in the host. Especially the SBM-based diet deviated strongly from the others in its capacity to down-regulate expression of genes related to mTOR signaling regulating the expression of a variety of response genes, including those with T cell related functions. From our studies, it is not clear yet whether the use of particular protein sources also affects the long-term health of the host. Currently "quality of protein sources" in both human and animal nutrition sectors is mainly based on their capacity to deliver energy and nutrients, including (essential) amino acids. However, our results highlight the importance of assessing parameters related to health and immune competence. Thus with regard to the selection of protein sources in animal or human diet, the stakeholders (nutritional researchers or industries) should also consider potential health related effects, apart from their nutritional value, as both could potentially impact vitality in humans and productivity in animals.

## Acknowledgments

We thank Mark Boekschoten and Guido Hooiveld of Human Nutrition, Wageningen University and Research for expertise and help with microarray. Additionally we want to thank Frank Harders from Central Veterinary Institute, Lelystad, The Netherlands for expertise and help with the Illumina Mi-Seq sequencing. Marjolein Meijerink from the Host-Microbe Interactomics Group, Wageningen University and Research for extending her help with ethical conduct of this experiment.

## Author Contributions

Conceived and designed the experiments: SKK AJMJ JMW MAS. Performed the experiments: SKK AJMJ DS LK EHS JJTT. Analysed the data: SKK NB JRG DS EHS. Performed immunohistochemistry: SKK JJTT. Wrote the paper: SKK AJMJ DS EHS JMW MAS.

## References


1. Capper J, Berger L, Brashears M, Jensen H. Animal Feed vs. Human Food: Challenges and Opportunities in Sustaining Animal Agriculture Toward 2050. Council for Agricultural Science and Technology (CAST). 2013;(Issue paper 53. CAST).
2. Dibble CC, Manning BD. Signal integration by mTORC1 coordinates nutrient input with biosynthetic output. Nat Cell Biol. 2013;15(6):555-64.
3. Weichhart T, Hengstschlager M, Linke M. Regulation of innate immune cell function by mTOR. Nat Rev Immunol. 2015;15(10):599-614.
4. Jahan-Mihan A, Luhovyy BL, Khoury DE, Anderson GH. Dietary Proteins as Determinants of Metabolic and Physiologic Functions of the Gastrointestinal Tract. Nutrients. 2011;3(5): 574-603.
5. Hernández-Ledesma B, del Mar Contreras M, Recio I. Antihypertensive peptides: production, bioavailability and incorporation into foods. Advances in colloid and interface science. 2011;165(1):23-35.
6. Chobert J-M. Milk protein modification to improve functional and biological properties. Advances in food and nutrition research. 2003;47:1-73.
7. Karr-Lilienthal LK, Kadzere CT, Grieshop CM, Fahey GC. Chemical and nutritional properties of soybean carbohydrates as related to nonruminants: A review. Livestock Production Science. 2005;97(1):1-12. doi: 10.1016/j.livprodsci.2005.01.015.
8. Madureira AR, Pereira CI, Gomes AMP, Pintado ME, Malcata FX. Bovine whey proteins - Overview on their main biological properties. Food Research International. 2007;40(10):1197-211.
9. van Neerven RJJ, Knol EF, Heck JML, Savelkoul HFJ. Which factors in raw cow's milk contribute to protection against allergies? Journal of Allergy and Clinical Immunology. 2012;130(4): 853-8.
10. Gao YY, Jiang ZY, Lin YC, Zheng CT, Zhou GL, Chen F. Effects of spray-dried animal plasma on serous and intestinal redox status and cytokines of neonatal piglets. J Anim Sci. 2011;89(1):150-7. doi: 10.2527/jas.2010-2967.
11. Ruth MR, Field CJ. The immune modifying effects of amino acids on gut-associated lymphoid tissue. Journal of Animal Science and Biotechnology. 2013;4. doi: 10.1186/2049-1891-4-27.
12. Ghaly AE, Alkoaik F. The yellow mealworm as a novel source of protein. American Journal of Agricultural and Biological Sciences. 2009;4(4):319-31.
13. Pingitore A, Lima GPP, Mastorci F, Quinones A, Iervasi G, Vassalle C. Exercise and oxidative stress: potential effects of antioxidant dietary strategies in sports. Nutrition (Burbank, Los Angeles County, Calif ). 2015;31(7-8):916-22. doi: 10.1016/j.nut.2015.02.005.
14. Liu RH. Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. Am J Clin Nutr. 2003;78(3 Suppl):517S-20S.
15. Heuzé V, Tran G, Kaushik S. Soybean meal. Feedipedia, a programme by INRA, CIRAD, AFZ and FAO. 2016 27.10.2016.

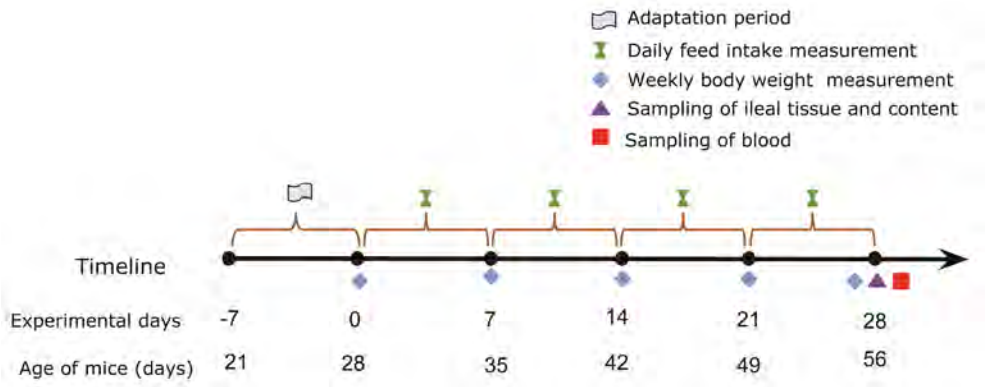
16. Gilbert R. World Animal Feed Industry. Protein sources for the animal feed industry. FAO,UN, Rome: Agriculture and consumer protection; 2004. p. 1-8.
17. Kar SK, Jansman AJM, Boeren S, Kruijt L, Smits. MA. Approximation of the amino acid composition and bio-functional properties of current and novel protein sources for pigs. J Anim Sci.In press.
18. Sovran B, Loonen LMP, Lu P, Hugenholtz F, Belzer C, Stolte EH, et al. IL-22-STAT3 Pathway Plays a Key Role in the Maintenance of Ileal Homeostasis in Mice Lacking Secreted Mucus Barrier. Inflammatory Bowel Diseases. 2015;21(3):531-42.
19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
20. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. Bmc Bioinformatics. 2015;16.
21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research. 2003;13(11):2498-504.
22. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21(16):3448-9.
23. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Research. 2014;42(D1):D472-D7.
24. Schokker D, Veninga G, Vastenhouw SA, Bossers A, de Bree FM, Kaal-Lansbergen LMTE, et al. Early life microbial colonization of the gut and intestinal development differ between genetically divergent broiler lines. BMC Genomics. 2015. doi: DOI 10.1186/s12864-015-1646-6.
25. Ramiro-Garcia J, Hermes G, Giatsis C, Sipkema D, Zoetendal E, Schaap P, et al. NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes F1000Research. 2016;5. doi: 10.12688/f1000research.9227.1.
26. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. 2013. doi: 10.1093/nar/gks1219.
27. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 2010;7(5):335-6. doi: 10.1038/nmeth.f.303.
28. Rodes L, Saha S, Tomaro-Duchesneau C, Prakash S. Microencapsulated Bifidobacterium longum subsp. infantis ATCC 15697 favorably modulates gut microbiota and reduces circulating endotoxins in F344 rats. BioMed research international. 2014;2014:602832. doi: 10.1155/2014/602832.
29. Karczewski J, Troost FJ, Konings I, Dekker J, Kleerebezem M, Brummer RJM, et al. Regulation of human epithelial tight junction proteins by Lactobacillus plantarum *in vivo* and protective effects on the epithelial barrier. American Journal of Physiology-Gastrointestinal and Liver

Physiology. 2010;298(6):G851-G9. doi: 10.1152/ajpgi.00327.2009. PubMed PMID: WOS:000277809200006.

30. Van Laethem F, Tikhonova AN, Pobezinsky LA, Tai X, Kimura MY, Le Saout C, et al. Lck availability during thymic selection determines the recognition specificity of the T cell repertoire. Cell. 2013;154(6):1326-41.
31. Chaimowitz NS, Falanga YT, Ryan JJ, Conrad DH. Fyn Kinase Is Required for Optimal Humoral Responses. PLoS One. 2013;8(4).
32. Scheeren FA, Kuo AH, van Weele LJ, Cai S, Glykofridis I, Sikandar SS, et al. A cell-intrinsic role for TLR2-MYD88 in intestinal and breast epithelia and oncogenesis. Nat Cell Biol. 2014;16(12):1238-48.
33. Fujiwara D, Chen L, Wei B, Braun J. Small intestine CD11c+ CD8+ T cells suppress CD4+ T cell-induced immune colitis. Am J Physiol Gastrointest Liver Physiol. 2011;300(6):G939-47.
34. Bettini ML, Guy C, Dash P, Vignali KM, Hamm DE, Dobbins J, et al. Membrane association of the CD3ε signaling domain is required for optimal T cell development and function. J Immunol. 2014;193(1):258-67
35. Haks MC, Pépin E, van den Brakel JHN, Smeele SAA, Belkowski SM, Kessels HWHG, et al. Contributions of the T Cell Receptor–associated CD3γ–ITAM to Thymocyte Selection. J Exp Med. 2002;196(1):1-13.
36. Boding L, Nielsen MW, Bonefeld CM, von Essen MR, Nielsen BL, Lauritsen JPH, et al. Polymorphisms of the T cell receptor CD3delta and CD3epsilon chains affect anti-CD3 antibody binding and T cell activation. Mol Immunol. 2010;47(15):2450-7.
37. Ladygina N, Gottipati S, Ngo K, Castro G, Ma J-Y, Banie H, et al. PI3Kγ kinase activity is required for optimal T-cell activation and. Eur J Immunol. 2013;43(12):3183-96.
38. Leopold Wager CM, Hole CR, Wozniak KL, Olszewski MA, Wormley FL, Jr. STAT1 signaling is essential for protection against Cryptococcus neoformans infection in mice. J Immunol. 2014;193(8):4060-71.
39. Siggs OM, Miosge LA, Daley SR, Asquith K, Foster PS, Liston A, et al. Quantitative reduction of the TCR adapter protein SLP-76 unbalances immunity and immune regulation. J Immunol. 2015;194(6):2587-95.
40. Wang X, Sumida H, Cyster JG. GPR18 is required for a normal CD8alphaalpha intestinal intraepithelial lymphocyte compartment. J Exp Med. 2014;211(12):2351-9.
41. Germena G, Volmering S, Sohlbach C, Zarbock A. Mutation in the CD45 inhibitory wedge modulates integrin activation and leukocyte recruitment during inflammation. J Immunol. 2015;194(2):728-38.
42. Schim van der Loeff I, Hsu L-Y, Saini M, Weiss A, Seddon B. Zap70 is essential for long-term survival of naive CD8 T cells. J Immunol. 2014;193(6):2873-80.
43. Ni Gabhann J, Hams E, Smith S, Wynne C, Byrne JC, Brennan K, et al. Btk regulates macrophage polarization in response to lipopolysaccharide. PLoS One. 2014;9(1):e85834.

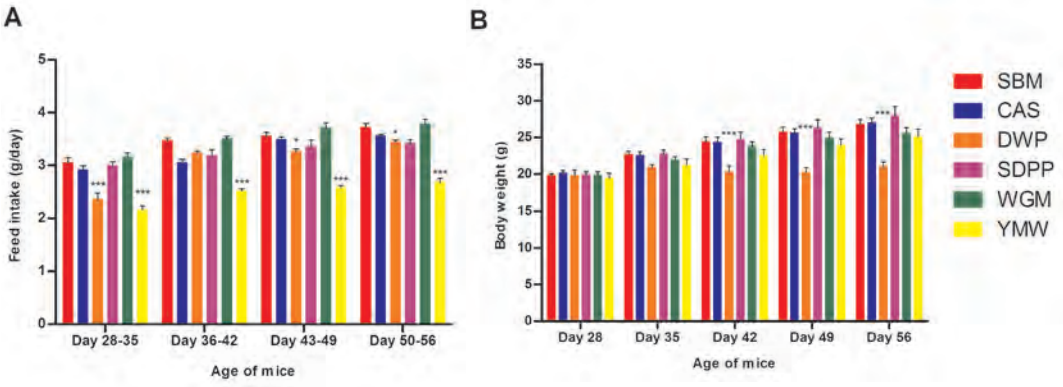
44. Hall WL, Millward DJ, Long SJ, Morgan LM. Casein and whey exert different effects on plasma amino acid profiles, gastrointestinal hormone secretion and appetite. Br J Nutr. 2003;89(2):239-48.
45. Pilvi TK, Korpela R, Huttunen M, Vapaatalo H, Mervaala EM. High-calcium diet with whey protein attenuates body-weight gain in high-fat-fed C57Bl/6J mice. Br J Nutr. 2007;98(5):900-7.
46. Funkat A, Massa CM, Jovanovska V, Proietto J, Andrikopoulos S. Metabolic adaptations of three inbred strains of mice (C57BL6, DBA/2, and 129T2) in response to a high-fat diet. Journal of Nutrition. 2004;134(12):3264-9.
47. Lee CY. The Effect of High-Fat Diet-Induced Pathophysiological Changes in the Gut on Obesity: What Should be the Ideal Treatment? Clinical and Translational Gastroenterology. 2013. doi: 10.1038/ctg.2013.11.
48. Andoh A, Yasui H, Inatomi O, Zhang Z, Deguchi Y, Hata K, et al. Interleukin-17 augments tumor necrosis factor-alpha-induced granulocyte and granulocyte/macrophage colony-stimulating factor release from human colonic myofibroblasts. J Gastroenterol. 2005;40(8):802-10.
49. Boneberg EM, Hareng L, Gantner F, Wendel A, Hartung T. Human monocytes express functional receptors for granulocyte colony-stimulating factor that mediate suppression of monokines and interferon-gamma. Blood. 2000;95(1):270-6.
50. Franzke A, Piao W, Lauber J, Gatzlaff P, Konecke C, Hansen W, et al. G-CSF as immune regulator in T cells expressing the G-CSF receptor: implications for transplantation and autoimmune diseases. Blood. 2003;102(2):734-9.
51. Rutella S, Pierelli L, Bonanno G, Sica S, Ameglio F, Capoluongo E, et al. Role for granulocyte colony-stimulating factor in the generation of human T regulatory type 1 cells. Blood. 2002;100(7):2562-71.
52. Morris ES, MacDonald KPA, Rowe V, Johnson DH, Banovic T, Clouston AD, et al. Donor treatment with pegylated G-CSF augments the generation of IL-10-producing regulatory T cells and promotes transplantation tolerance. Blood. 2004;103(9):3573-81.
53. Meshkibaf S, Gower MW, Dekaban GA, Kim SO. G-CSF preferentially supports the generation of gut-homing Gr-1high macrophages in M-CSF-treated bone marrow cells. J Leukoc Biol. 2014;96(4):549-61.
54. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. Microbial degradation of complex carbohydrates in the gut. Gut microbes. 2012;3(4):289-306.
55. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A. 2005;102(31):11070-5. doi: 0504978102.
56. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature. 2006;444(7122):1022-3.
57. Sela DA, Mills DA. Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. Trends in microbiology. 2010;18(7):298-307.
58. Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell. 2016;165(4):842-53.

## Supplemental Information

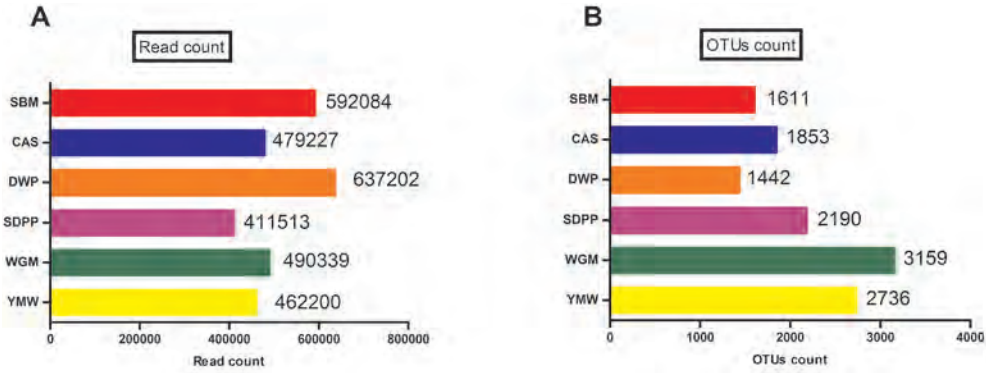



**S1 Fig. Design of the experiment.**
The solid black dot in the timeline represents the corresponding experimental days/age of the mice (days). Ileal tissue and its digesta were used for transcriptome and microbiota analysis. Blood was collected for analysis of systemic immune signaling molecules.



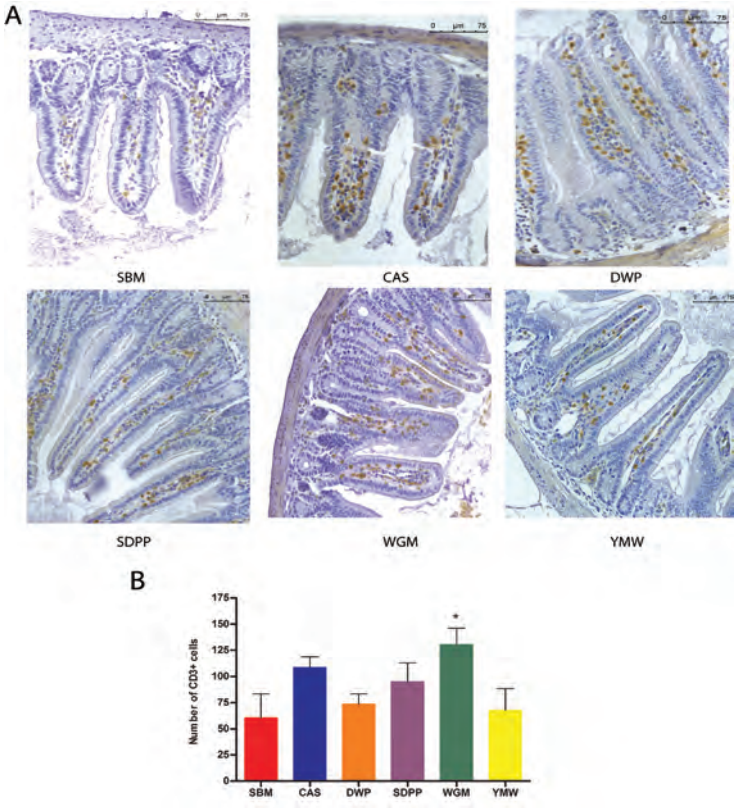



**S2 Fig. Feed intake (A) and bodyweight (B) of mice fed with different experimental diets.**
Bars and whiskers represent mean values ± SEM (n = 6), respectively for feed intake (left) and body weight (right) recorded throughout the experimental period. *$P$ < 0.05, ***$P$ < 0.001 compared with SBM-fed mice fed. Here, SBM, soybean meal; CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.



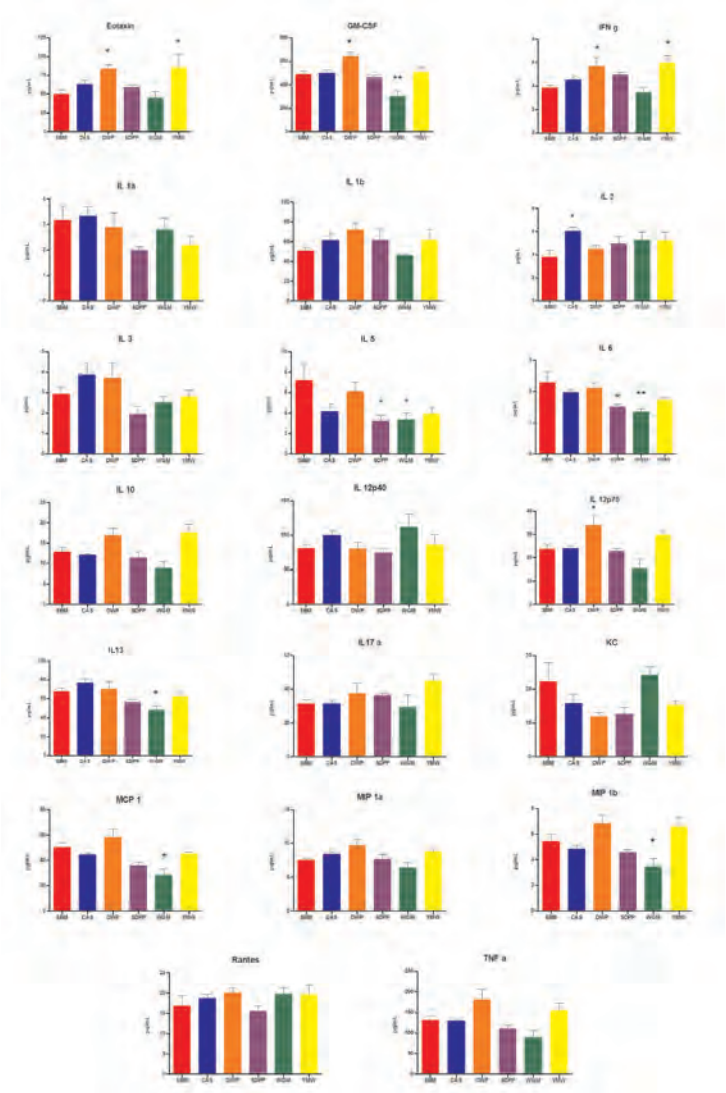

**S3 Fig. The mean number of 16S rRNA sequence reads (A) and the number of OTU (B) counts detected in the ileal samples of mice fed with different experimental diets.**

SBM, soybean meal; CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.



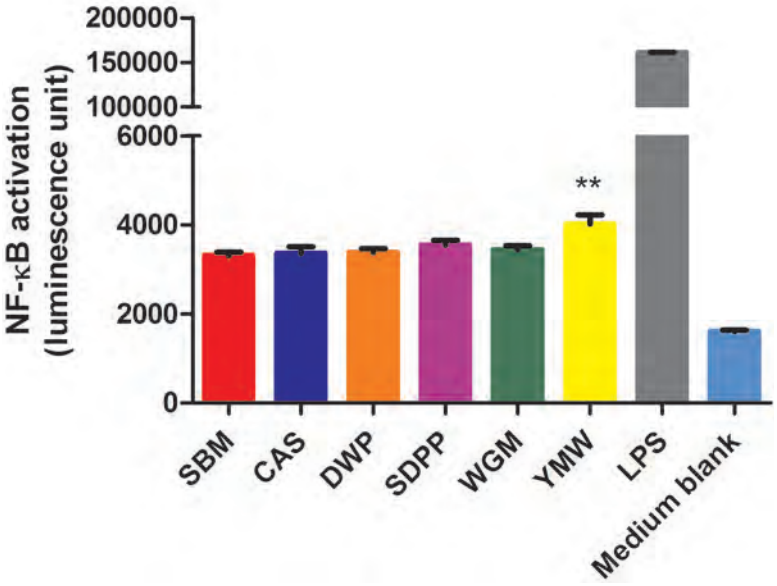

**S4 Fig. Immunohistochemistry of CD3+ T cell response in ileal tissue of mice fed with different experimental diets.**

(A): Scale bar: 75 μm. Brown colored cell represent $CD3^+$ cells (marker of T cell). (B): Pooled count of $CD3^+$ cells in ileum of mice fed with the various diets. SBM, soybean meal; CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.




**S5 Fig. Concentrations of serum cytokines and chemokines in response to dietary treatment in mice.**

Bars and whiskers represent mean values ± SEM (n = 6), for systemic chemokines and cytokines recorded at the end of experiment. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 compared with SBM-fed mice fed. Here, SBM, soybean meal; CAS, casein preparation intended for animal use; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.




**S6 Fig. Toll like receptor (TLR) 4 response of supernatant isolated from the faecal samples of mice fed with diets containing protein from various sources, LPS and blank.**

Error bars indicate SEM; n=6. ***P* < 0.01 compared to SBM-fed mice fed. SBM, soybean meal; CAS, casein; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.

**S1 Table. Ingredient and calculated or analysed nutrient composition of the experimental diets for mice, as fed basis[1].**

| | Diets[2] | | | | | |
|---|---|---|---|---|---|---|
| **Item** | **SBM** | **CAS** | **DWP** | **SDPP** | **WGM** | **YMW** |
| Ingredient, *g/kg* | | | | | | |
| Maize | 297.5 | 297.5 | 297.5 | 297.5 | 297.5 | 297.5 |
| Dextrose | 132 | 132 | 132 | 132 | 132 | 132 |
| Sugar | 100 | 100 | 100 | 100 | 100 | 100 |
| Arbocell | 50 | 50 | 50 | 50 | 50 | 50 |
| Soybean Oil | 70 | 70 | 70 | 70 | 70 | 70 |
| AIN-93G MX | 35 | 35 | 35 | 35 | 35 | 35 |
| AIN-93-VX | 10 | 10 | 10 | 10 | 10 | 10 |
| Choline chloride | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
| DL-Methionine | 3 | 3 | 3 | 3 | 3 | 3 |
| Soybean meal | 300 | 0 | 0 | 0 | 0 | 0 |
| Casein | 0.0 | 300 | 0 | 0 | 0 | 0 |
| Delactosed whey powder | 0.0 | 0 | 300 | 0 | 0 | 0 |
| Spray dried plasma protein | 0.0 | 0 | 0 | 300 | 0 | 0 |
| Wheat gluten meal | 0.0 | 0 | 0 | 0 | 300 | 0 |
| Yellow meal worm | 0.0 | 0 | 0 | 0 | 0 | 300 |
| Composition, *g/kg*[3] | | | | | | |
| Dry matter | 914 | 957 | 930 | 924 | 917 | 929 |
| Crude protein | 153 | 268 | 80 | 238 | 252 | 148 |
| Ash | 43 | 29 | 77 | 28 | 47 | 35 |
| Crude fibre | 55 | 44 | 44 | 44 | 46 | 59 |
| Crude fat | 76 | 65 | 74 | 87 | 70 | 160 |
| Starch | 251 | 249 | 249 | 249 | 268 | 261 |
| Sugar | 295 | 263 | 403 | 263 | 271 | 263 |
| NSP[4] | 71 | 8 | 17 | 15 | 2 | 18 |
| **Gross energy, *KJ/g*** | **17** | **19** | **16** | **18** | **18** | **20** |
| Ca | 5.9 | 5.5 | 10.0 | 5.2 | 5.2 | 6.1 |
| P | 3.7 | 3.3 | 6.2 | 1.9 | 2.3 | 4.0 |
| K | 10.4 | 4.0 | 16.5 | 4.5 | 4.0 | 3.6 |
| Na | 1.1 | 1.2 | 5.9 | 8.2 | 1.3 | 1.0 |
| Cl | 1.7 | 2.1 | 10.4 | 12.7 | 1.9 | 1.6 |
| Linoleic acid | 38 | 36 | 36 | 36 | 36 | 36 |
| Electrolyte balance, *Meq/kg* | 266 | 94 | 388 | 115 | 106 | 92 |

[1]Analysed composition is presented in bold.

[2]Diet: CAS is casein preparation intended for animal use, DWP is partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal , WGM is wheat gluten meal and YMW is yellow meal worm.

[3]Diet were formulated using data on ingredient nutrient composition and nutrient digestibility coefficients according to the Central Bureau for Livestock Feeding (CVB, Lelystad, the Netherlands).

[4]NSP: Non-starch polysaccharides.

**S2 Table. Unique significantly expressed gene-sets (FDR <0.05) in ileum of mice fed with different protein sources compared to SBM-fed diet as shown in Fig 1.**

| Treatments vs SBM | Serial numbers | Unique gene sets |
|---|---|---|
| CAS | | |
| | 1 | RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS_AND_ASSEMBLY |
| | 2 | KEGG_PYRIMIDINE_METABOLISM |
| | 3 | RIBOSOME_BIOGENESIS_AND_ASSEMBLY |
| | 4 | RHYTHMIC_PROCESS |
| DWP | | |
| | 1 | PEPTIDYL_TYROSINE_PHOSPHORYLATION |
| | 2 | DNA_DEPENDENT_DNA_REPLICATION |
| | 3 | ORGANELLE_LOCALIZATION |
| | 4 | CELL_CYCLE_GO_0007049 |
| | 5 | PEPTIDYL_AMINO_ACID_MODIFICATION |
| | 6 | INTERPHASE |
| | 7 | PEPTIDYL_TYROSINE_MODIFICATION |
| | 8 | SMALL_GTPASE_MEDIATED_SIGNAL_TRANSDUCTION |
| | 9 | POSITIVE_REGULATION_OF_SIGNAL_TRANSDUCTION |
| | 10 | KEGG_LYSOSOME |
| | 11 | DNA_METABOLIC_PROCESS |
| | 12 | CHROMOSOME_ORGANIZATION_AND_BIOGENESIS |
| | 13 | DNA_REPAIR |
| | 14 | I_KAPPAB_KINASE_NF_KAPPAB_CASCADE |
| | 15 | KEGG_NON_SMALL_CELL_LUNG_CANCER |
| | 16 | INTERPHASE_OF_MITOTIC_CELL_CYCLE |
| | 17 | REGULATION_OF_SIGNAL_TRANSDUCTION |
| | 18 | REGULATION_OF_PROGRAMMED_CELL_DEATH |
| | 19 | KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION |
| | 20 | REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS |
| | 21 | REGULATION_OF_APOPTOSIS |
| | 22 | REGULATION_OF_MAPKKK_CASCADE |
| | 23 | KEGG_P53_SIGNALING_PATHWAY |

| Treatments vs SBM | Serial numbers | Unique gene sets |
|---|---|---|
| SDPP | | |
| | 1 | KEGG_PPAR_SIGNALING_PATHWAY |
| | 2 | KEGG_BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS |
| | 3 | KEGG_COMPLEMENT_AND_COAGULATION_CASCADES |
| | 4 | MITOTIC_CELL_CYCLE_CHECKPOINT |
| | 5 | KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION |
| | 6 | KEGG_THYROID_CANCER |
| | 7 | POSITIVE_REGULATION_OF_T_CELL_ACTIVATION |
| | 8 | REGULATION_OF_MITOSIS |
| YMW | | |
| | 1 | KEGG_PRION_DISEASES |
| | 2 | DNA_REPLICATION_INITIATION |
| | 3 | HOMOPHILIC_CELL_ADHESION |
| | 4 | REGULATION_OF_DEVELOPMENTAL_PROCESS |

Here, SBM, soybean meal; CAS, casein preparation intended for animal use; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.

**S3 Table. Members of microbial family belonging to 'others' as shown in Fig 2.**

| Sl No. | Taxa | Sl No. | Taxa |
|---|---|---|---|
| 1 | o__Holophagales;f__Holophagaceae | 25 | c__WCHB1-32;o__o;f__f |
| 2 | o__Actinomycetales;f__ Actinomycetaceae | 26 | p__BD1-5;c__c;o__o;f__f |
| 3 | o__Bifidobacteriales;f__f | 27 | p__Candidate_division_ TM7;c__c;o__o;f__f |
| 4 | o__Corynebacteriales;f__ Corynebacteriaceae | 28 | o__Chloroplast;f__Chloroplast |
| 5 | o__Frankiales;f__Cryptosporangiaceae | 29 | o__Chloroplast;f__f |
| 6 | o__Frankiales;f__f | 30 | o__SubsectionIV;f__FamilyI |
| 7 | o__Frankiales;f__Geodermatophilaceae | 31 | o__Deferribacterales;f__ Deferribacteraceae |
| 8 | o__Frankiales;f__Sporichthyaceae | 32 | o__Deinococcales;f__Deinococcaceae |
| 9 | o__Micrococcales;f__Microbacteriaceae | 33 | o__Bacillales;f__Family_XI_Incertae_ Sedis |
| 10 | o__Micrococcales;f__Micrococcaceae | 34 | o__Bacillales;f__Staphylococcaceae |
| 11 | o__Pseudonocardiales;f__ Pseudonocardiaceae | 35 | o__Lactobacillales;f__Carnobacteriaceae |
| 12 | o__Rubrobacterales;f__ Rubrobacteriaceae | 36 | o__Lactobacillales;f__Enterococcaceae |
| 13 | o__Bacteroidales;f__Bacteroidaceae | 37 | o__Lactobacillales;f__Streptococcaceae<br>o__Lactobacillales;f__Leuconostocaceae |
| 14 | o__Bacteroidales;f__f | 38 | o__Clostridiales;f__f |
| 15 | o__Bacteroidales;f__Marinilabiaceae | 39 | o__Clostridiales;f__Family_XI_ Incertae_Sedis |
| 16 | o__Bacteroidales;f__ Porphyromonadaceae | 40 | o__Sphingomonadales;f__ Sphingomonadaceae |
| 17 | o__Bacteroidales;f__Prevotellaceae | 41 | o__Clostridiales;f__Peptococcaceae |
| 18 | o__Cytophagales;f__Cytophagaceae | 42 | o__Clostridiales;f__uncultured |
| 19 | o__Flavobacteriales;f__ Cryomorphaceae | 43 | o__Clostridiales;f__Veillonellaceae |
| 20 | o__Flavobacteriales;f__ Flavobacteriaceae | 44 | o__Erysipelotrichales;f__f |
| 21 | o__Sphingobacteriales;f__ Chitinophagaceae | 45 | o__Fusobacteriales;f__Fusobacteriaceae |
| 22 | o__Sphingobacteriales;f__f | 46 | o__Fusobacteriales;f__Leptotrichiaceae |
| 23 | o__Sphingobacteriales;f__ Saprospiraceae | 47 | o__Caulobacterales;f__Caulobacteraceae |
| 24 | o__Sphingobacteriales;f__ Sphingobacteriaceae | 48 | o__Rhizobiales;f__Methylobacteriaceae |

| | | | |
|---|---|---|---|
| 49 | o__Rhodobacterales;f__Rhodobacteraceae | 62 | o__Myxococcales;f__f |
| 50 | o__Rhodospirillales;f__Rhodospirillaceae | 63 | o__Enterobacteriales;f__Enterobacteriaceae |
| 51 | o__Sphingomonadales;f__f | 64 | o__Legionellales;f__Coxiellaceae |
| 52 | o__Sphingomonadales;f__Sphingomonadaceae | 65 | o__Oceanospirillales;f__Halomonadaceae |
| 53 | o__Burkholderiales;f__Burkholderiaceae | 66 | o__Pasteurellales;f__Pasteurellaceae |
| 54 | o__Burkholderiales;f__Comamonadaceae | 67 | o__Pseudomonadales;f__Moraxellaceae |
| 55 | o__Burkholderiales;f__Oxalobacteraceae | 68 | o__Pseudomonadales;f__Pseudomonadaceae |
| 56 | o__Methylophilales;f__Methylophilaceae | 69 | o__Xanthomonadales;f__Xanthomonadaceae |
| 57 | o__Neisseriales;f__Neisseriaceae | 70 | o__Anaeroplasmatales;f__Anaeroplasmataceae |
| 58 | o__Nitrosomonadales;f__Nitrosomonadaceae | 71 | o__Verrucomicrobiales;f__Verrucomicrobiaceae |
| 59 | o__Rhodocyclales;f__Rhodocyclaceae | 72 | k__Bacteria;p__WCHB1-60 |
| 60 | o__Bdellovibrionales;f__f | 73 | o__Agaricomycetes;f__f |
| 61 | o__GR-WP33-30;f__f | 74 | k__NA |

# 5

# CHAPTER 5

## Amine Metabolism is Influenced by Dietary Protein Source

*Soumya K. Kar[1,2], Alfons J.M. Jansman[3], Dirkjan Schokker[2], Amy C. Harms[4,5], Leo Kruijt[2,3], Jerry M. Wells[1], Mari A. Smits[1,2,6]*

[1]Host Microbe Interactomics Group, Wageningen University & Research, Wageningen, the Netherlands
[2]Wageningen University & Research Animal Breeding and Genomics, Wageningen, The Netherlands
[3]Wageningen Livestock Research, Animal Nutrition, Wageningen University & Research, Wageningen, The Netherlands
[4]Netherlands Metabolomics Centre, Leiden University, Leiden, the Netherlands
[5]Department of Analytical Biosciences, Leiden University, Leiden, the Netherlands
[6]Wageningen Bioveterinary Research, Wageningen University & Research, Wageningen, the Netherlands

*Manuscript submitted for publication*

## Abstract

Growth in world population will inevitably lead to increased demand for protein for humans and animals. Protein from insects and blood plasma are being considered as possible alternatives but more research on their nutritional quality and health effects is needed. Here we studied the effect of dietary protein source on metabolism and metabolic amine profiles in serum and urine of mice. Groups of mice were fed semi-purified diets containing 300 g/kg of either soybean meal (SBM), casein (CAS), partially de-lactosed whey powder (DWP), spray dried plasma protein (SDPP), wheat gluten meal (WGM) or yellow meal worm (YMW). Feed and water intake as well as body weight gain were measured for 28 days. After 14 and 28 days serum and urine samples were collected for measurement of a large panel of amine metabolites. Metaboanalyst 3.0 was used for analysis of the raw metabolic data. Out of 68 targeted amine metabolites, we could detect 54 in urine and 41 in blood serum. Dietary protein sources were found to have profound effects on host metabolism, particularly in systemic amine profiles, considered here as an endo-phenotype. We recommend serum over urine to screen for the amine metabolic endo-phenotype based on partial least squares - discriminant analysis. We concluded that metabolites like alpha-aminobutyric acid and 1-methylhistidine are sensitive indicators of too much or too little availability of specific amino acids in the different protein diets. Furthermore, we concluded that amine metabolic profiles can be useful for assessing the nutritional quality of different protein sources.

**Key words: amine metabolites, dietary protein source, endo-phenotype, metabolomics, mice**

## Introduction

If growth in the human population and food production continue at their present rate, there will be more than 2 billion more people to feed in 2050 and calorific deficiencies are likely to be more prevalent than they are today. One of the main concerns about the world food supply is the production of proteins for humans and livestock. Plant proteins (e.g. from soy) will continue to dominate in the next 10 years but inevitably alternative sources of protein will need to be found because arable land cannot be increased in proportion. The strategy of partly replacing current protein sources for animal feed with economically viable and sustainable alternatives e.g. protein from insects or livestock blood plasma, is one possible solution. In recent years, significant efforts have been made to introduce new protein containing feed ingredients in the diet of livestock (1, 2). Unfortunately, information on the functional properties of (new) protein sources towards their consumers is scarce.

In general practice, animal diets are formulated based on the provision of ileal or faecal digestible nutrients and derived net energy by feed ingredients. The nutrients relate to proteins/amino acids (AA), starch and sugars, fat, fermentable non-starch-polysaccharides, minerals and vitamins, which can be characterised as the "strict-nutritional" value of feed ingredients. However, apart from the "strict-nutritional" value, diets and their ingredients have other "non-strict-nutritional", functional properties in relation to e.g. feed intake (satiety), passage rate through the gastro intestinal tract (GIT), pro- and antimicrobial properties, antioxidative and oxidative effects, immune signalling and metabolic effects (3, 4). From the context of protein ingredients, it has the capacity to deliver AAs (both essential and non-essential) that are essential precursors for the synthesis of organic nitrogen compounds, including amines, that influence processes related to protein metabolism. Glutathione, creatinine and nitric oxide have protective affect against oxidative stress and toxicity (5, 6). Other examples of biologically active amines synthesised from amino acids are dopamine and serotonin which are neurotransmitters exerting behavioural effects (7, 8). Further, we have previously predicted that polypeptides with antimicrobial activity, inhibition of angiotensin I-converting enzyme, as well as anti-oxidative, antithrombotic and anti-amnestic activities could be generated during the digestion of several protein ingredients (9). The current practice of feed formulation for farm animals largely ignores the requirement for (precursors of) such functional or bioactive compounds that can be important for health and performance phenotypes. It would be beneficial to not only evaluate the strict-nutritional values of protein sources, but also consider their potential non-strict-nutritional functional properties and effects *in vivo*. However, approaches to evaluate the functional properties of proteins sources beyond their capacity to provide (essential) AA and other nutrients (10) have not been reported.

Measuring amine metabolites in serum or urine is considered a useful approach to assess the host and microbiota metabolism of proteins in the diets. On one hand, a profile of blood amine metabolites would reveal information about the nutritional efficiency of different protein sources and on the other hand, the metabolic and absorptive capacity of the gut, including the influence of the microbiome (11). Systemic amine metabolic profiles are anticipated to be good candidate biomarkers to predict dietary protein associated phenotypes in relation to health (12-14) but have not been previously studied in relation to different protein sources. Ultimately amine metabolic profiles might reveal biomarkers that could be used to assess health status and nutritional quality of different protein sources.

The aim of this research study was to investigate the effect of mice fed different protein sources on the metabolic amine profiles in serum and urine and on host metabolism. Groups of mice were fed semi-purified diets containing different protein sources (300 g/kg) as follows: soybean meal (SBM), casein (CAS), partially delactosed whey powder (DWP), spray dried plasma protein (SDPP), wheat gluten meal (WGM) and yellow meal worm (YMW). Serum and urine were collected after 14 and 28-days for measurement of a large panel of amine metabolites.

## Materials and Methods

### *Diet*

Customised semi-synthetic diets based on AIN-93G were prepared replacing CAS with one of following alternative protein sources: SBM, DWP, WGM, SDPP and YMW at an inclusion level of 300 g/kg. Representative samples of dried and ground diets were chemically analysed for dry matter (DM; NEN-ISO 6496 by 4 hours drying at 104°C), nitrogen (N; NEN-ISO 5983-2 by Kjeldahl method and crude protein calculated as Nx6.25), ash (NEN-ISO 5984 after 3 hours ashing at 550°C), ether extract (EE, NEN-ISO 6492 by extraction with petroleum ether) and gross energy (GE; NEN- EN-ISO 9831 by bomb calorimetry). The ingredient and chemical composition of the experimental diets is presented in Supplementary Table 1.

### *Animal and design of experiment*

All procedures were approved by the Wageningen Animal Ethics Committee (Wageningen, the Netherlands; accession number 2012062.c) and carried out according to the guidelines of the European Council Directive 86/609/EEC dated November, 1986. A schematic representation and detailed description of the experimental design and sample collection is given in Supplementary Figure 1. Briefly, upon arrival 72 twenty-one-day-old wild type male C57BL/6J mice (Harlan Laboratories, Horst, the Netherlands) were stratified according to bodyweight and litter of origin into six dietary groups in a

light and temperature-controlled animal facility of Wageningen University & Research (12:12 h reversed light/dark cycle, 20 ± 2°C). The mice were housed in pairs in a specific pathogen-free environment with ad libitum access to diet and water. Prior to the start of the experiment mice were adapted for one week to a standard diet based on AIN-1993 growth (AIN-93G), which included 300 g/kg casein as the only protein source (as fed basis, CAS). Thereafter, one group continued with the CAS diet and the other five groups received similar semi-synthetic diets containing 300 g/kg (as fed basis) of one of the alternative protein sources (diet containing soybean meal, SBM; diet containing partially delactosed whey powder, DWP; diet containing spray dried porcine plasma, SDPP; diet containing wheat gluten meal, WGM and diet containing yellow meal worm, YMW) for 28 d. Feed intake and water consumption was measured every day and body weight of the animals was measured every week. Urine samples were collected just before euthanasia, from each animal by mechanical stimulation of the ventral body section. On d 14 and 28, six mice from each group were anaesthetized with isoflurane and sacrificed to collect blood and urine samples. After euthanasia of the mice, blood samples were collected by orbital puncture and serum was extracted using 500 µl SST tubes (Becton Dickinson, Franklin Lakes, New Jersey) within 30 minutes after collection of the blood. Urine and blood samples were stored at -80°C for further analysis of metabolites.

## Metabolomics profiling

### *Assay description*

The amine profiling was performed as described previously (15). Briefly, 5 µL of each sample was spiked with an internal standard solution (Table S2), thiol amines were released from proteins and converted to reduced form using Tris-(2-Carboxyethyl) phosphine . Then proteins were precipitated by the addition of methanol. The supernatant was transferred to an Eppendorf tube (Eppendorf, Germany) and dried in a speedvac (Eppendorf, Germany). The residue was reconstituted in borate buffer (pH 8.5) with AQC reagent (Waters, Etten-Leur, The Netherlands). After reaction, the vials were transferred to an autosampler tray (Waters, Etten-Leur, The Netherlands) and cooled to 10°C prior to injection. For amine metabolite analysis, 1 µL of the reaction mixture was injected into the UPLC-MS/MS system using an Accq-Tag Ultra column (Waters, Etten-Leur, The Netherlands).

### *Equipment*

We employed an ACQUITY UPLC system with autosampler (Waters, Etten-Leur, The Netherlands) was online coupled with a Xevo Tandem Quadrupole (TQ) mass spectrometer (Waters, Etten-Leur, The Netherlands) operated using QuanLynx data acquisition software (version 4.1; Waters, Etten-Leur, The Netherlands). The Xevo

TQ was used in the positive-ion electrospray mode and all analytes were monitored in multiple reaction monitoring (MRM) using nominal mass resolution.

***Data processing and quality check of metabolomics data***

Acquired data were evaluated using TargetLynx software (Waters, Etten-Leur, The Netherlands), by integration of assigned MRM peaks and normalization using proper internal standards. For analysis of amino acids, their 13C15N-labeled analogues were used. For other amines, the closest-eluting internal standard was employed (Supplementary Table 2). Blank samples were used to correct for background and in-house developed algorithms were applied using the pooled quality check (QC) samples to compensate for shifts in the sensitivity of the mass spectrometer over the batch analysis (16). Out of 68 targeted amine metabolites, we could detect 41 amines in serum and 53 amines in urine that comply with the acceptance criteria of QC corrections (16). These metabolites were detected in both serum and urine in each six mice per treatment for d 14 and 28 samples. However, only in CAS, DWP and SDPP urine samples from d 28, metabolites were detected in five mice per treatment group due to lack of sufficient sample from one animal in each of these groups. The data are represented as relative response ratios (amine target area/area of internal standard; unit free) of these metabolites (after QC) are available in the supplementary material 1 (available on request).

***Data analysis***

Univariate statistical analysis was employed to experimental data to examine the effect of dietary treatment. In addition, MetaboAnalyst 3.0 (17), a web-based tool was employed to carry out comprehensive metabolomics data analysis and visualization (http://www.metaboanalyst.ca/).

***Univariate statistical analysis***

Results of feed intake, water consumption and body weight are presented as means ± standard error of the mean. Statistical analysis was performed by one-way ANOVA followed by a post hoc test (Dunnett test: compared all treatment vs. SBM group as control) using GraphPad prism version 5.03 for Windows Vista (GraphPad Software, San Diego, California, USA). Statistical significance was defined as P value < 0.05. Soybean meal (SBM) diet served as reference to make comparisons with other dietary treatments for the univariate analysis (18, 19). Pearson's linear correlation was used to calculate the *r* value between the concentrations of calculated according to CVB table apparent ileal digestible (AID) essential amino acids (EAA) in the experimental diets and the amino acid concentrations in serum on d 28.

***Amine metabolomics data analysis***

To analyse the amine metabolomics data, MetaboAnalyst 3.0 was employed. Within MetaboAnalyst 3.0, we used two modules i.e. exploratory (multivariate and clustering) statistical analysis and functional (pathway) analysis. Within the exploratory statistical analysis module, partial least squares - discriminant analysis (PLS-DA) method was used for clustering and classification of the treatments based on the amine profile in serum and urine samples from d 14 and 28 of the experimental animals. Unlike principal component analysis (PCA), PLS-DA is a robust form of analysis, directed towards factor space that are associated with high variation in the responses but biased towards directions that are accurately predicted (in this case, sample groups i.e. experimental diets) (20). Here, we have used leave-one-out cross-validation method to measure the predictability performance ($Q^2$) to validate the PLS-DA model. $Q^2$ has no standard of comparison or critical value for inferring significance, aside from its theoretical maximum of 1 or an empirically inferred acceptable value of ≥ 0.4 for a biological model (21). The value of $Q^2$ closer to 1 is best, whereas above 0.5 is considered good (22). In addition, we have employed permutation tests to overcome the problem of PLS-DA's propensity to data overfitting that cannot be detected through cross-validation (21). The aim of the permutation-based validation is to measure the performance of the predictor model by determining the probability (*P* value) of observing an equal or better performance by pure chance. For example, if none of the permuted classes is better than the observed one in 2000 permutations, the P value is reported as $P < 0.0005$ (less than 1/2000). Permutation test was carried out with separation distance (between/within) with 2000 permutations as featured in the MetaboAnalyst 3.0. Briefly, we used the amine metabolic profiles of individual mice from each treatment groups in both serum and urine samples from d 14 and 28 of the experimental period. The SBM diet served as reference for this study and hence, data was normalized by a pooled sample from the SBM group and log transformation was carried out.

Two outputs i.e. the score and the loading plots of PLS-DA were used for clustering and classification of the treatments based on the amine profile in d 28 serum samples from the experimental animals. The scores plot provides an intuitive summary of the sample clustering patterns by projecting high-dimensional metabolomics data into two dimensions in a way that explains the maximal co-variance (PLS-DA) of the data; while the loading plot shows the underlying compounds responsible for such separation patterns. Two-dimensional (2D) scores plot with 95% confidence region of treatment specific cluster was used to visualized the score plots. In addition, clustered heatmap was used as a tool to find cluster of amine metabolites associated with treatment groups. This provide a visual description of the evolution of the clusters based on the concentration of the metabolites in the treatment groups (23). We choose "do not recognize-samples"

to show the natural contrast among treatment groups and rest other selection criteria or options are left at default for MetaboAnalyst 3.0.

Pathway analysis module of MetaboAnalyst 3.0 was employed to determine the amine metabolism that were affected by the treatments compared to the SBM diet. To perform the pathway analysis, we used the amine metabolic profile of individual mice in serum of d 28 of the experimental period, to prepare the data matrix of individual comparisons of experimental diets with SBM. All the compound names of the metabolites were matched with the human metabolome database. Normalization was performed as described for exploratory statistical analysis. Thereafter, *Mus musculus* pathway library and a reference metabolome based on our technical platform were uploaded (Supplementary document 1). The analysis includes pathway enrichment analysis and topological analysis were involved. The impact-value threshold calculated from topology analysis was set at 0.4 and –log (p) value calculated from pathway enrichment was set to 4 to identify the most related metabolic pathway.

Variable importance in the projection (VIP) score, another output from above mentioned PLS-DA (based on serum of d 28 of the experimental period) was used to determine the amine metabolites that can discriminate all treatments. The VIP score positively reflects the metabolite's influence on the classification. Amine metabolomics data analysis were performed on the data as relative response ratios (amine target area/area of internal standard; unit free) of each metabolite retrieved after QC correction. These amine metabolomics data analysis is available in the Supplementary document 1.

## Results

### *Composition of the diet*

The experimental diets were formulated to contain adequate levels of EAA, minerals, vitamins and fatty acids, as recommended for a rodent diets by the American Institute of Nutrition (AIN-93, growth) (24) (Supplementary Table 1). The concentrations of AID AAs (g/kg) in the diets were calculated using data on the AID of AAs or protein of the protein sources in pigs (Table 1) (25).

**TABLE 1: Calculated concentrations of apparent ileal digestible amino acids[1] (g/kg) in the experimental diets.[2]**

| | Diets[2] | | | | | |
|---|---|---|---|---|---|---|
| | SBM | CAS | DWP | SDPP | WGM | YMW |
| ALA | 4.9 | 7.6 | 2.6 | 7.6 | 6.0 | NA[3] |
| ARG | 10.0 | 9.1 | 1.5 | 12.5 | 8.3 | 8.6 |
| ASP | 13.6 | 18.0 | 5.7 | 18.0 | 7.3 | NA |
| CYS | 1.6 | 0.8 | 1.2 | 7.4 | 5.0 | 8.6 |
| GLU | 22.0 | 54.2 | 9.8 | 54.2 | 80.3 | NA |
| GLY | 4.5 | 4.4 | 0.6 | 4.4 | 7.3 | NA |
| HIS | 3.3 | 7.9 | 1.2 | 7.0 | 4.9 | 5.3 |
| ILE | 5.7 | 12.8 | 3.4 | 5.9 | 8.6 | 9.9 |
| LEU | 9.5 | 24.7 | 6.0 | 20.5 | 16.3 | 15.8 |
| LYS | 8.9 | 21.3 | 6.1 | 20.1 | 4.8 | 10.4 |
| MET | 4.7 | 10.6 | 4.0 | 4.4 | 6.7 | 14.6 |
| PHE | 6.6 | 13.3 | 2.2 | 12.0 | 12.1 | 6.1 |
| PRO | 5.7 | 27.9 | 2.9 | 27.9 | 29.0 | 8.0 |
| SER | 5.9 | 13.1 | 2.5 | 13.1 | 10.8 | NA |
| THR | 4.6 | 10.4 | 3.5 | 11.3 | 5.6 | NA |
| TRP | 1.7 | 3.2 | 0.9 | 3.2 | 2.0 | 7.5 |
| TYR | 4.7 | 14.4 | 1.5 | 12.8 | 7.8 | 2.4 |
| VAL | 5.8 | 16.5 | 3.1 | 13.9 | 9.2 | 11.5 |

[1]Apparent ileal digestibility of amino acids (based on pigs) calculated with digestibility coefficients of amino acids in SBM, CAS DWP and WGM as listed by CVB (25). Gross amino acid composition is shown here for YMW-based diet.

[2]Diet: SBM is soybean meal, CAS is casein, DWP partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal, WGM is wheat gluten meal and YMW is yellow meal worm.

[3]NA: Not available

### *Feed intake and growth parameters*

All mice appeared to be healthy throughout the experimental period of 4 weeks. However, mean body weight was significantly lower ($P < 0.05$) for the DWP-fed compared to the SBM-fed mice throughout the experimental period (Supplementary Figure 2). Compared to SBM-fed mice, a significant increase of water consumption was observed in DWP- and WGM- fed mice. Water consumption was significantly lower for CAS-, SDPP- and YMW-fed mice compared to SBM-fed mice. In YMW-fed mice, we measured a significantly lower feed and water intake compared to the SBM-fed mice (Supplementary Figure 2), while the development of body weight over the experimental period did not differ compared to the SBM-fed mice.

***Clustering of metabolites according to diet***

Figure 1 shows the metabolic profile separation of the experimental treatments in two dimensional (2D) PLS-DA, considering score plots of the first two components of the PLS-DA. The PLS-DA score plots, based on the amine metabolites measured in serum on d 14 and 28, had a higher $Q^2$ value of predictability in relation to the protein source generating the metabolic profile (Figure 1A, C) than the metabolic profiles in urine samples collected on the same days (Figure 1B, D). The predictability performance of the PLS-DA model that is constructed with serum amine metabolites at d 28 to discriminate dietary treatments is high, based on the observed p-value in the permutation test (Supplementary Figure 3). Therefore we subsequently focussed on comparing serum amine metabolic profiles from different protein sources on d 28 (Figure 1C). The YMW group metabolite profile was clearly separated from those obtained from the other groups (Figure 1C). The DWP diet also showed a tendency to cluster separately, although an overlap with CAS and SBM was observed using a 95% confidence limit for the shaded plots. The serum amines responsible for the separation of YMW amine metabolic profiles from those to other protein sources, is explained by large amounts of 1-methylhistidine, glycine and 4-hydroxyproline (Supplementary figure 4). The serum amine metabolites responsible for the separation of the DWP in the PLS-DA plot 1C were (large amounts) of alpha-aminobutyric acid, s-methylcysteine, threonine, citrulline and alanine (Supplementary figure 5) and (low amounts) of leucine, valine, phenylalanine, glutathione, beta-alanine and putrescine (Supplementary Figure 6).

***Correlation between concentrations of dietary and plasma amino acids***

The calculated concentrations of apparent ileal digestible EAA in the experimental diets and the measured concentrations of EAA in serum are detailed in Supplementary Table 3. The correlation coefficients of the apparent dietary EAA and the serum concentrations of EAA is shown per diet in Figure 2. Based on the correlation values (r), diets prepared with different protein sources could be ranked as SDPP > SBM > DWP > CAS > YMW > WGM; r value ranging from 0.81 to -0.15 (strongest and positive correlation for SDPP, r = 0.72; weakest and negative correlation for WGM, r= -0.09; Figure 2).

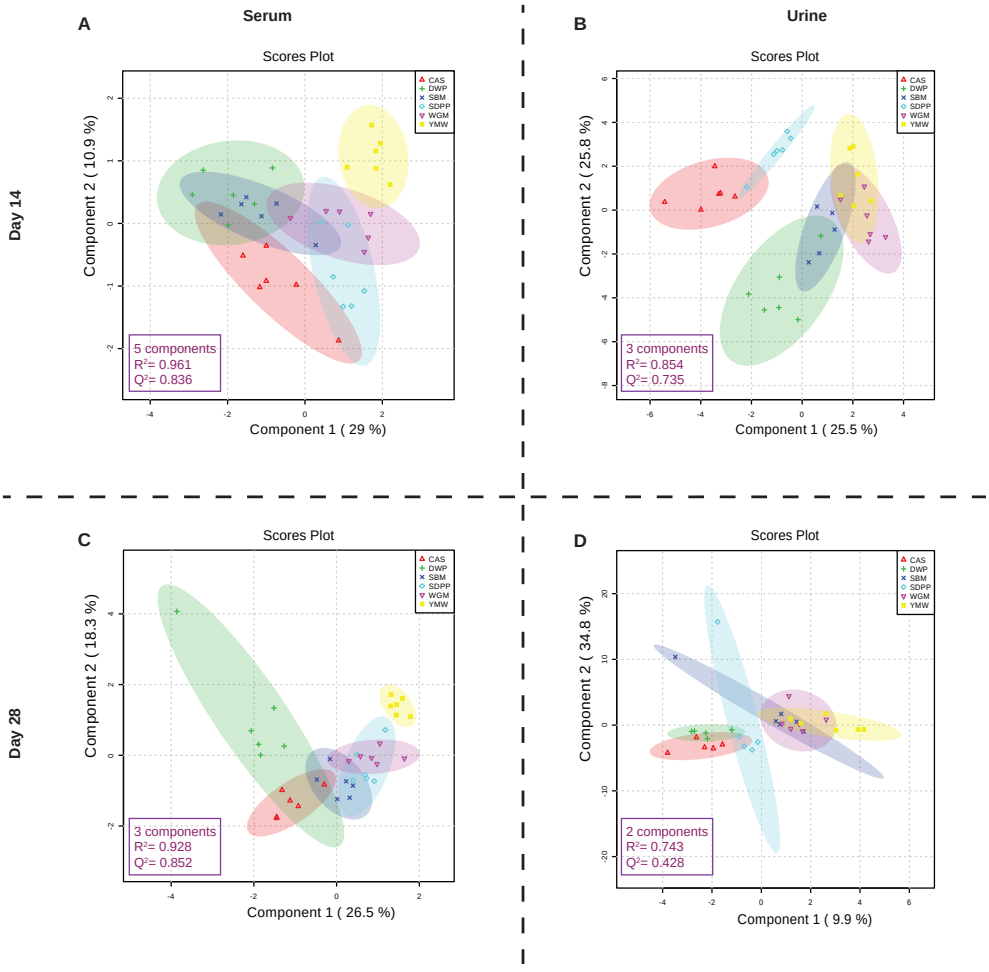



**FIGURE 1: PLS-DA score plots of amine profiles in serum and urine on d 14 and 28 of mice receiving diets with different protein sources.**

Panel A: PLS-DA model built using serum amine metabolites of mice at d 14 of the experimental period; B: PLS-DA model built using urine amine metabolites of mice at d 14 of the experimental period; C: PLS-DA model built using serum amine metabolites of mice at d 28 of the experimental period; D: PLS-DA model built using urine amine metabolites of mice at d 28 of the experimental period. The colored dots represent samples from different treatments. Coloured spherical areas displays 95% confidence region of respective experimental diets. Each symbol represents data from one mouse. Parameters within the magenta box represents the summary from cross validation of PLS-DA model. The 'components' describe the best number of components the model utilises to capture the discrimination within treatment group based on the highest $Q^2$ value in the cross validation of PLS-DA model. $R^2$ represents the quality and $Q^2$ represents the predictability of each PLS-DA model. Here, SBM is soybean meal, CAS is casein, DWP partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal, WGM is wheat gluten meal and YMW is yellow meal worm.

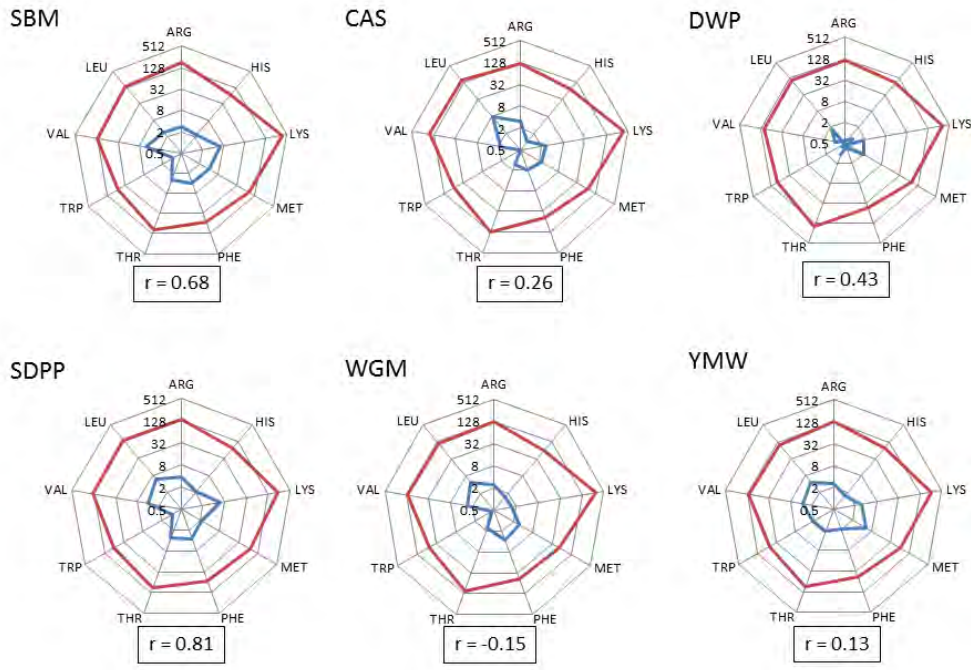



**FIGURE 2: Radial representation of log transformed calculated concentrations of apparent ileal digestible EAA in the experimental diets and the concentrations of free EAA in serum of mice fed the corresponding experimental diets (d 28).**

Number on circles refer to essential amino acid values in logarithmic scale base 2 ($log_2$) and the integers in the spider graph are the log transformed values. Blue line: Essential amino acid composition (g/kg) as determined in the experimental diets. Red line: Essential amino acid composition (μM/ml) as measured in serum on d 28. 'r' means correlation values between calculated EAA in the experimental diets and the absolute concentrations of EAA in serum of mice. Here, SBM is soybean meal, CAS is casein, DWP partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal, WGM is wheat gluten meal and YMW is yellow meal worm.

### *Diet associated amine metabolic-sets*

The clustered heat-map revealed seven (A-G) diets associated clusters of pathway-related amine metabolites in serum of mice at d 28 of the experimental period (Figure 3). All diets had 1 or two clusters of amine metabolites that were present in relatively higher or lower concentrations comparted to other diets. For example, Cluster A is comprises histidine, sarcosine and serine which were found at relatively high concentrations in the YMW-fed mice and cluster B is comprising threonine, alanine, citrulline, alpha-aminobutyric acid and s-methylcysteine were present at relatively high concentrations in DWP-fed mice.

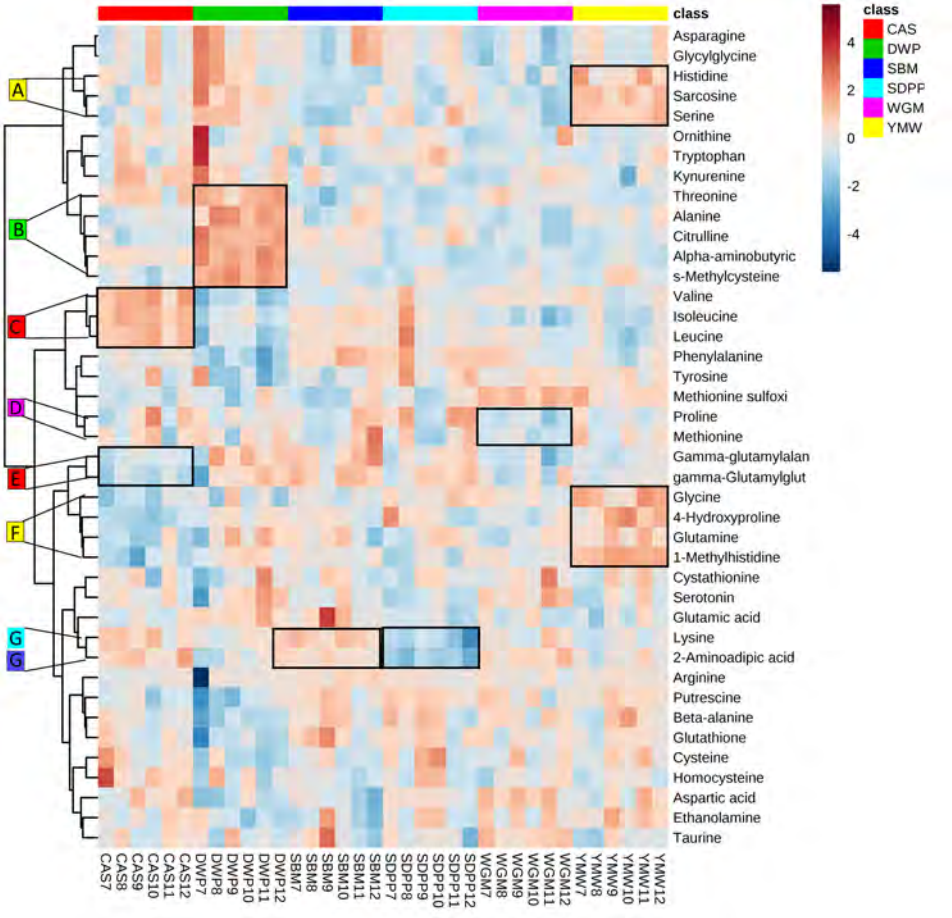



**FIGURE 3: Heat-map of clustering of serum amine metabolites in dependence of dietary treatment.**

CAS: Casein; DWP: partially delactosed whey powder; SBM: Soybean meal; SDPP: Spray dried plasma protein; WGM: Wheat gluten meal; YMW: Yellow meal worm. Values are the relative response ratios of each metabolite in individual mice per treatment group. Treatment groups comprise of six mice. Values are measured in serum of mice at d 28 of the experimental period. The heat map graphic distances were measured using Euclidean distances and the clustering algorithm using ward dendrogram. Each coloured cell on the map corresponds to a concentration value (normalized to the SBM treatment). The alphabets A-G represent the diets associated cluster formed by sets of metabolites (framed in black) having low (shades of blue) or high (shades of orange) concentration in the dietary treatments. Here, the coloured square corresponds to the respective experimental diets; i.e. red for CAS, green for DWP, navy blue for SBM, sky blue for SDPP, magenta for WGM and yellow for YMW. Here, SBM is soybean meal, CAS is casein, DWP partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal, WGM is wheat gluten meal and YMW is yellow meal worm. A to G indicate amine metabolite clusters referred to in the text.

***Functional analysis: Metabolic pathways***

Serum amine metabolites which differed in concentration between SBM and the other protein diets were mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways. This revealed seven metabolic pathways that were influenced by dietary protein source (Figure 4).

Comparing CAS vs SBM, the enriched KEGG pathways were: valine, leucine and isoleucine biosynthesis; arginine and proline metabolism and glutathione metabolism. For DWP vs SBM, the enriched KEGG pathways were: valine, leucine and isoleucine biosynthesis; glutathione metabolism; glycine, serine and threonine metabolism;

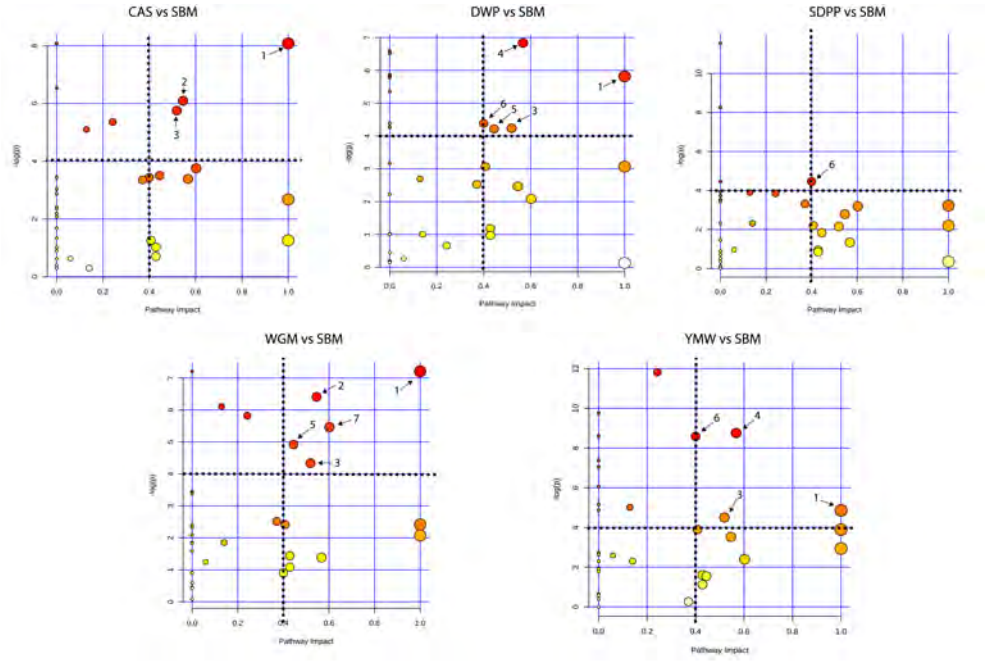



**FIGURE 4: Metabolic pathway analysis based on the plasma amine profiles in mice fed diets with different protein sources.**

Bubble graph representing the result of metabolic pathway analysis. Scores from enrichment analysis is represented on the 'y axis' and from topology analysis on the 'x axis'. The size and the colour (on gradient scale: white to red) of the nodes are all matched pathways according to P values from pathway enrichment analysis and pathway impact values from pathway topology analysis. The dotted black lines denotes the thresholds in both axis to identify the most significant matched pathways for all the dietary comparisons. The arrow indicates to the enriched pathways, where, 1 is Valine, leucine and isoleucine biosynthesis; 2 is Arginine and proline metabolism; 3 is Glutathione metabolism; 4 is Glycine, serine and threonine metabolism; 5 is Beta-alanine metabolism; 6 is Methane metabolism and 7 is Alanine, aspartate and glutamate metabolism. SBM is soybean meal, CAS is casein, DWP partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal, WGM is wheat gluten meal and YMW is yellow meal worm.

beta-alanine metabolism and methane metabolism. For SDPP vs SBM, the only enriched KEGG pathway was methane metabolism. For WGM vs SBM, the enriched KEGG pathways were: valine, leucine and isoleucine biosynthesis; arginine and proline metabolism; glutathione metabolism; beta-alanine metabolism and alanine, aspartate and glutamate metabolism. For YMW vs SBM, the enriched KEGG pathways were: valine, leucine and isoleucine biosynthesis; glutathione metabolism; glycine, serine and threonine metabolism and methane metabolism (Figure 4). The valine, leucine and isoleucine biosynthesis pathway was commonly enriched in all diets comparisons except in the SDPP vs SBM comparison. In the pathway, differences were observed in the concentration of the branched chain amino acids (BCAA, i.e. valine, leucine and isoleucine) between CAS, WGM, DWP and YMW (Supplementary Figure 7-10).

***Diet associated sets of amine metabolites***

PLS-DA plots were used to find the fundamental relations between different diets and amine metabolites in serum or urine (Figure 1). The importance of each variable i.e. amine metabolite in the projection represented for each diet in the PLS-DA plots is estimated by the variable importance in projection (VIP) score and metabolites having VIP scores greater than 2 can be considered as marker for a specific dietary protein source (26, 27).

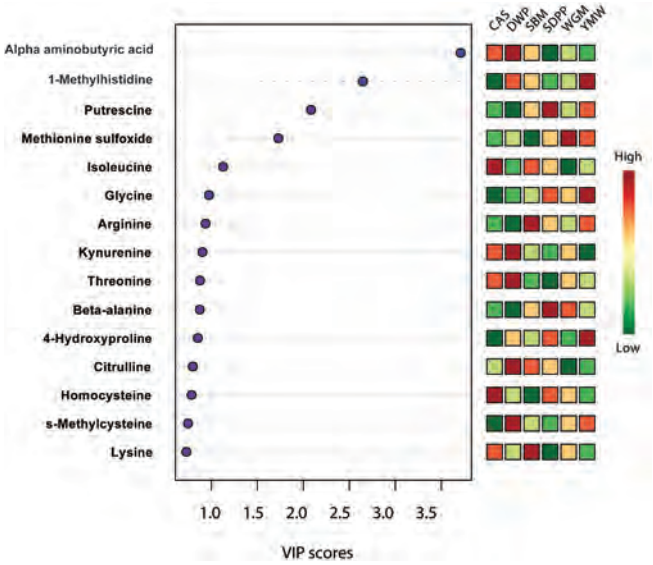



**FIGURE 5: Important features identified by PLS-DA model based on amines metabolites in serum at d 28 of experimental period.**

Metabolites with a high VIP score are predictive for certain diet e.g. alpha amino butyric acid is predictive for DWP and 1-methylhistidine for YMW. The colored boxes on the right indicate the relative concentration of the corresponding metabolites in each group under study. CAS: Casein; DWP: partially delactosed whey powder; SBM: Soybean meal; SDPP: Spray dried plasma protein; WGM: Wheat gluten meal; YMW: Yellow meal worm. Values are the mean of six replicates.

Three metabolites had VIP scores > 2 in the PLS-DA analysis (Figure 5); these were alpha-aminobutyric acid, 1-methylhistidine (1-MHis) and putrescine. A high concentration of alpha-aminobutyric acid and a low concentration of putrescine was observed for DWP-fed mice. Moreover, the dietary treatments can be discriminated based on differences in concentrations of the top seven metabolites. A high concentration of alpha-aminobutyric acid and a low concentration of putrescine was observed in the DWP-fed mice. Similar but reciprocal observations were noticed for alpha-aminobutyric acid and putrescine in the SDPP-fed mice. A high concentration of 1-MHis was observed in YMW-fed mice and a low concentration in the CAS-fed mice.

## Discussion

Here we show that the inclusion of different protein sources in the diets lead to the appearance of specific amine metabolite profiles in the serum and urine of mice. Differences in these metabolic endo-phenotypes were most pronounced for the YMW- and DWP-based diet. This might be due to the lower body weight gain for the DWP group and/or the lower feed and water intake in the YMW group and/or provision of amino acids in these diets.-

Amine-based profiles or endo-phenotypes originate from interactions of dietary protein components with the genotype of the animal, its associated microbiome and the production environment (Figure 6). The results described in this paper clearly demonstrate that amine metabolic profiles in blood and urine are affected by the source of dietary protein. These amine profiles are composed of essential and non-essential AAs, amine intermediate metabolites which are important for cellular protein synthesis but also biosynthesis of neurotransmitters (e.g. serotonin) and immunomodulatory metabolites (e.g. kynurenine). Additionally, amines generated by microbial metabolism (e.g. putrescine) and precursors for other nitrogen based biologically active motifs (e.g. glutathione) are also identified using this metabolic platform. Serum amine metabolic profiles were more reliable for discrimination of the dietary protein source than those from urine. This seems logical as amines entering the circulation from the GIT, can be further metabolized or stored in tissue depots of splanchnic organs or undergo further metabolic degradation elsewhere in the body,, whereas urine amine metabolites represent excreted amines which are either toxic or metabolically in excess and not required by the host (Figure 6).

The differences in nutrient composition of the diets used in the present study, is related to the fixed inclusion (300 g/kg) of different protein sources that contain

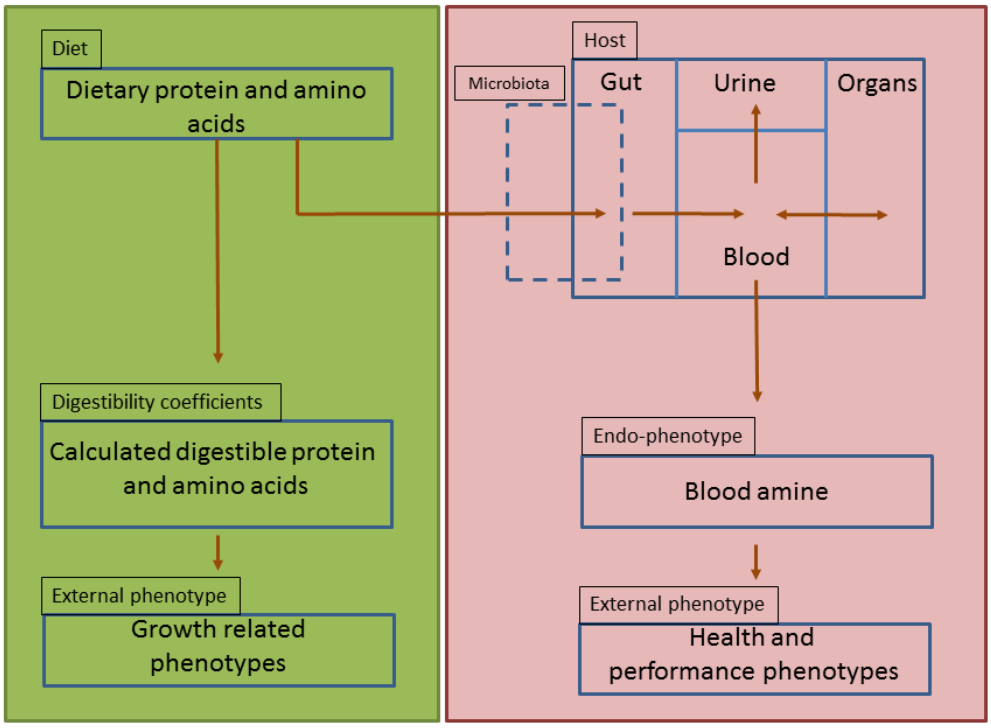



**FIGURE 6: Schematic representation of amine based endo-phenotype concept to evaluate dietary protein.**

Green panel represents the current dietary protein evaluation practice that is based on amino acids and their digestibility coefficients as function of only performance related phenotypes (strict-nutritional characteristics). Red panel represents the proposed use of amine based endo-phenotype in blood for dietary protein source evaluation.



protein as well as other constituents such as carbohydrates and fats (9). From a nutritional perspective, dietary protein has no nutritional value unless it is hydrolyzed by proteases and peptidases into free AAs, dipeptides, or tripeptides in the lumen of the small intestine (28) and subsequently absorbed and metabolized by the intestinal tissue or transmitted into the bloodstream. In this context, it was interesting to observe that the correlation between dietary and blood essential AAs differed between the used diets. Therefore, we conclude that EAA concentrations in blood are not only related to the quantitative provision of AAs from the diets, based on calculated values for apparent ileal digestible AAs, but also dependent on host, microbiota and/or environment related factors. Apparently, the correlation for the SDPP( r = 0.81) and SBM (r = 0.68) fit better with the resulting blood profile of the EAA than YMW (r = 0.13) and WGM (r = - 0.15) based diets. This indicates that the EAA profile of SDPP and SBM, better mimic the required AA profile by (farm) animals than YMW or WGM.

From the heat-map (Figure 3), we learned that the variation in amine profiles was due to variation in the concentration of metabolically-related clusters of metabolites. For example, a set of metabolites formed by valine, leucine and isoleucine, all showed a higher concentration in the CAS diet compared to other experimental diets. The differential abundance of metabolite clusters in animals fed specific diet suggests a dietary induced regulation of metabolic pathways rather than the regulation of single metabolic steps. Furthermore, we observed an enrichment of the valine, leucine and isoleucine biosynthesis pathway, based on the amine profile in the serum of mice fed with CAS when compared with SBM. Within the enriched metabolic pathway, the concentration level of valine, leucine and isoleucine (branched chain amino acids, BCAA) showed a high concentration in mice fed the CAS diet compared to the SBM diet. In particular, the valine level was significantly ($P < 0.05$) higher in the serum of mice fed the CAS based diet compared to the SBM based diet. Our results provide clear evidence that compared to SBM, casein is an excellent source for BCAA, based upon the combined results of the heat-map and the metabolic pathway analysis. Branched chain AAs have been reported to have health benefit effects because they spare lean body mass during weight loss (29), promote wound healing (30), promote muscle protein anabolism in muscle wasting with aging (31) and have beneficial effects in the setting of renal and liver disease (32). These observations are in agreement with previous findings where a casein based diet is capable of influencing amino acid metabolism mediated via BCAA biosynthesis pathway in humans and pigs, compared to soy proteins (33, 34).

As casein and whey are from milk derived protein source, one may expect that DWP would exhibit a similar metabolic profile to CAS when compared to SBM (35). However, we observed a contrasting result in serum of mice fed with DWP to CAS for AAs profile, in particularly to the BCAA profile (Supplementary Figure 8). This can be explained by the differences between whey and casein digestive properties or protein digestion kinetics (36). Studies have shown differences in the peak plasma levels of AAs following ingestion of whey and casein i.e., whey protein peaking earlier in time than casein (37, 38). Our results suggest that, due to the differences in the digestive properties or protein digestion kinetics of DWP (fast) and CAS (slow), these diets differentially alter host protein metabolism.

Based on the difference in the observed amine metabolite profile, we identified two metabolites (viz. alpha-aminobutyric acid for DWP and 1-MHis for YMW) that are highly discriminative for the various amine-based endo-phenotypes. Alpha-aminobutyric acid, is a key intermediate in the synthesis of a tripeptide analogue of glutathione (i.e. ophthalmic acid) with antioxidative properties. 1-methylhistidine (1-MHis) results from the metabolism of the dipeptide anserine found in meat sources (39, 40). The enzyme carnosinase, present in intestinal mucosal tissues, splits anserine into beta-alanine and

1-MHis. The latter has already been shown as a marker of meat consumption in humans (41). Higher level in YMW-based diet suggests a high dietary supply of anserine from this protein source. To date, there is no evidence to support that 1-methylhistidine itself has any detrimental effects on health (42), but some recent evidences have shown its association with health conditions (43, 44).

The systemic levels of amine-metabolites could have predictive value for external phenotypes of livestock animals such as regulation of growth performance, nutritional status and health characteristics (see Figure 6). On one hand, we observed that both the amine endo-phenotype as well as the external performance phenotypes varied as function of the used diet. This was most pronounced for the YMW and DWP, as the results of both the amine metabolites as well as the animal body weight gain measurements showed the highest variation among the different diets. On the other hand, we cannot conclude from this study whether amine-based endo-phenotypes, or (composite) biomarkers derived thereof, can be used to predict external phenotypes associated with specific sources of dietary protein. This will require longitudinal studies in healthy or disease cohorts with a higher number of animals fed diet containing different protein sources. Here we show that part of (protein) metabolism is influenced by the protein sources included in the diets. Such understanding of metabolism is require to move forward in nutritional science. In the future such knowledge may also be useful for predicting the effect of different protein sources on external phenotypes relevant to health and animal performance.

## References


1. van der Spiegel M, Noordam MY, van der Fels-Klerx HJ. Safety of Novel Protein Sources (Insects, Microalgae, Seaweed, Duckweed, and Rapeseed) and Legislative Aspects for Their Application in Food and Feed Production. Comprehensive Reviews in Food Science and Food Safety (2013) 12(6):662-78. doi: 10.1111/1541-4337.12032. PubMed PMID: WOS:000325631600005.
2. Boland MJ, Rae AN, Vereijken JM, Meuwissen MPM, Fischer ARH, van Boekel M, et al. The future supply of animal-derived protein for human consumption. Trends in Food Science & Technology (2013) 29(1):62-73. doi: 10.1016/j.tifs.2012.07.002. PubMed PMID: WOS:000314074500006.
3. Biesalski H-K, Dragsted LO, Elmadfa I, Grossklaus R, Muller M, Schrenk D, et al. Bioactive compounds: definition and assessment of activity. Nutrition (Burbank, Los Angeles County, Calif ) (2009) 25(11-12):1202-5. doi: 10.1016/j.nut.2009.04.023.
4. Jansman AJM. Health and functions of the gastrointestinal tract in pigs: Effects of functional ingredients and feed and ingredient processing. J Anim Sci (2016) 94:12-21. doi: 10.2527/jas2015-9886. PubMed PMID: WOS:000385006900002.
5. Saleh S, El-Demerdash E. Protective effects of L-arginine against cisplatin-induced renal oxidative stress and toxicity: role of nitric oxide. Basic Clin Pharmacol Toxicol (2005) 97(2):91-7. doi: PTOpto_114.
6. Mels CMC, Huisman HW, Smith W, Schutte R, Schwedhelm E, Atzler D, et al. The relationship of nitric oxide synthesis capacity, oxidative stress, and albumin-to-creatinine ratio in black and white men: the SABPA study. Age (Dordr) (2016) 38(1):9. doi: 10.1007/s11357-016-9873-6 [doi].
7. Langen M, Kas MJH, Staal WG, van Engeland H, Durston S. The neurobiology of repetitive behavior: of mice. Neurosci Biobehav Rev (2011) 35(3):345-55. doi: S0149-7634(10)00028-X.
8. Ransford CP. A role for amines in the antidepressant effect of exercise: a review. Med Sci Sports Exerc (1982) 14(1):1-10.
9. Kar SK, Jansman AJM, Boeren S, Kruijt L, Smits MA. Approximation of the amino acid composition and bio-functional properties of current and novel protein sources for pigs. J Anim Sci (2016) 94:30–9. doi: 10.2527/jas2015-9677.
10. Jahan-Mihan A, Luhovyy BL, Khoury DE, Anderson GH. Dietary Proteins as Determinants of Metabolic and Physiologic Functions of the Gastrointestinal Tract. Nutrients (2011) 3(5):574-603.
11. Kogut MH, Arsenault RJ. Editorial: Gut Health: The New Paradigm in Food Animal Production. Frontiers in Veterinary Science (2016) 3:71. doi: 10.3389/fvets.2016.00071. PubMed PMID: PMC5005397.
12. Chen TL, Xie GX, Wang XY, Fan J, Qiu YP, Zheng XJ, et al. Serum and Urine Metabolite Profiling Reveals Potential Biomarkers of Human Hepatocellular Carcinoma. Molecular

& Cellular Proteomics (2011) 10(7). doi: 10.1074/mcp.M110.004945. PubMed PMID: WOS:000292541500002.

13. Picard M, McManus MJ, Gray JD, Nasca C, Moffat C, Kopinski PK, et al. Mitochondrial functions modulate neuroendocrine, metabolic, inflammatory, and transcriptional responses to acute psychological stress. Proc Natl Acad Sci U S A (2015) 112(48):E6614-E23. doi: 10.1073/pnas.1515733112. PubMed PMID: WOS:000365988900007.
14. Wang Q, Wurtz P, Auro K, Morin-Papunen L, Kangas AJ, Soininen P, et al. Effects of hormonal contraception on systemic metabolism: cross-sectional and longitudinal evidence. International Journal of Epidemiology (2016) 45(5):1445-57. doi: 10.1093/ije/dyw147. PubMed PMID: WOS:000393184400019.
15. Noga MJ, Dane A, Shi SN, Attali A, van Aken H, Suidgeest E, et al. Metabolomics of cerebrospinal fluid reveals changes in the central nervous system metabolism in a rat model of multiple sclerosis. Metabolomics (2012) 8(2):253-63. doi: 10.1007/s11306-011-0306-3. PubMed PMID: WOS:000301041800008.
16. van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. Journal of Proteome Research (2009) 8(11):5132-41. doi: 10.1021/pr900499r. PubMed PMID: WOS:000276949600022.
17. Xia JG, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Research (2015) 43(W1):W251-W7. doi: 10.1093/nar/gkv380. PubMed PMID: WOS:000359772700039.
18. Heuzé V, Tran G, Kaushik S. Soybean meal. Feedipedia, a programme by INRA, CIRAD, AFZ and FAO. (2016 27.10.2016).
19. Gilbert R. World Animal Feed Industry. Protein sources for the animal feed industry. FAO,UN, Rome: Agriculture and consumer protection (2004). p. 1-8.
20. Barker M, Rayens W. Partial least squares for discrimination. Journal of Chemometrics (2003) 17(3):166-73. doi: 10.1002/cem.785. PubMed PMID: WOS:000182005800002.
21. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, et al. Assessment of PLSDA cross validation. Metabolomics (2008) 4(1):81-9. doi: 10.1007/s11306-007-0099-6. PubMed PMID: WOS:000252807000007.
22. Fujimura Y, Kurihara K, Ida M, Kosaka R, Miura D, Wariishi H, et al. Metabolomics-Driven Nutraceutical Evaluation of Diverse Green Tea Cultivars. PLoS One (2011) 6(8). doi: 10.1371/journal.pone.0023426. PubMed PMID: WOS:000295454200085.
23. Beckonert O, Bollard ME, Ebbels TMD, Keun HC, Antti H, Holmes E, et al. NMR-based metabonomic toxicity classification: hierarchical cluster analysis and k-nearest-neighbour approaches. Analytica Chimica Acta (2003) 490(1-2):3-15. doi: 10.1016/s0003-2670(03)00060-6. PubMed PMID: WOS:000185047900002.
24. Reeves PG, Rossow KL, Lindlauf J. Developement and testing of the AIN-93 purified diets for rodents- Results on growth, kidney calcification and bone mineralization in rats and mice. Journal of Nutrition (1993) 123(11):1923-31. PubMed PMID: WOS:A1993MF52600020.

25. CVB. Chemical composition and nutritional values of feedstuffs and feeding standards. October 2007 ed. The Hague, The Netherlands: Product Board Animal Feed (2007).
26. Palermo G, Piraino P, Zucht H-D. Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. Adv Appl Bioinforma Chem (2009) 2:57-70.
27. Perez-Enciso M, Tenenhaus M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. Hum Genet (2003) 112(5-6):581-92. doi: 10.1007/s00439-003-0921-9.
28. Wu GY. Dietary protein intake and human health. Food & Function (2016) 7(3):1251-65. doi: 10.1039/c5fo01530h. PubMed PMID: WOS:000372251500002.
29. Layman DK. The role of leucine in weight loss diets and glucose homeostasis. Journal of Nutrition (2003) 133(1):261S-7S. PubMed PMID: WOS:000180305900049.
30. Zhang XJ, Chinkes DL, Wolfe RR. Leucine supplementation has an anabolic effect on proteins in rabbit skin wound and muscle. Journal of Nutrition (2004) 134(12):3313-8. PubMed PMID: WOS:000225720500018.
31. Rieu I, Sornet C, Bayle G, Prugnaud J, Pouyet C, Balage M, et al. Leucine-supplemented meal feeding for ten days beneficially affects postprandial muscle protein synthesis in old rats. Journal of Nutrition (2003) 133(4):1198-205. PubMed PMID: WOS:000182036700037.
32. Marchesini G, Bianchi G, Merli M, Amodio P, Panella C, Loguercio C, et al. Nutritional supplementation with branched-chain amino acids in advanced cirrhosis: A double-blind, randomized trial. Gastroenterology (2003) 124(7):1792-801. doi: 10.1016/s0016-5085(03)00323-8. PubMed PMID: WOS:000183236200012.
33. Luiking YC, Deutz NEP, Jakel M, Soeters PB. Casein and soy protein meals differentially affect whole-body and splanchnic protein metabolism in healthy humans. Journal of Nutrition (2005) 135(5):1080-7. PubMed PMID: WOS:000228953600020.
34. Deutz NEP, Bruins MJ, Soeters PB. Infusion of soy and casein protein meals affects interorgan amino acid metabolism and urea kinetics differently in pigs. Journal of Nutrition (1998) 128(12):2435-45. PubMed PMID: WOS:000077552500021.
35. Otani H, Hata I. Inhibition of proliferative responses of mouse spleen lymphocytes and rabbit peyers patch cells by bovine-milk caseins and their digests. Journal of Dairy Research (1995) 62(2):339-48. PubMed PMID: WOS:A1995QW96100013.
36. Boirie Y, Dangin M, Gachon P, Vasson MP, Maubois JL, Beaufrere B. Slow and fast dietary proteins differently modulate postprandial protein accretion. Proc Natl Acad Sci U S A (1997) 94(26):14930-5.
37. Dangin M, Boirie Y, Garcia-Rodenas C, Gachon P, Fauquant J, Callier P, et al. The digestion rate of protein is an independent regulating factor of postprandial protein retention. Am J Physiol Endocrinol Metab (2001) 280(2):E340-8.
38. Bos C, Metges CC, Gaudichon C, Petzke KJ, Pueyo ME, Morens C, et al. Postprandial kinetics of dietary amino acids are the main determinant of their metabolism after soy or milk protein ingestion in humans. J Nutr (2003) 133(5):1308-15.

39. Block WD, Hubbard RW, Steele BF. Excretion of histidine and histidine derivatives by human subjects ingesting protein from different sources. J Nutr (1965) 85:419-25.
40. Sjolin J, Hjort G, Friman G, Hambraeus L. Urinary excretion of 1-methylhistidine: a qualitative indicator of exogenous 3-methylhistidine and intake of meats from various sources. Metabolism (1987) 36(12):1175-84.
41. Myint T, Fraser GE, Lindsted KD, Knutsen SF, Hubbard RW, Bennett HW. Urinary 1-methylhistidine is a marker of meat consumption in Black and in White California Seventh-day Adventists. Am J Epidemiol (2000) 152(8):752-5.
42. Owusu-Apenten R. Bioactive Peptides: Applications for Improving Nutrition and Health: CRC Press (2010). 414 Pages p.
43. Cross AJ, Major JM, Rothman N, Sinha R. Urinary 1-methylhistidine and 3-methylhistidine, meat intake, and colorectal adenoma risk. Eur J Cancer Prev (2014) 23(5):385-90. doi: 10.1097/CEJ.0000000000000027 [doi].
44. Rafikova O, Meadows ML, Kinchen JM, Mohney RP, Maltepe E, Desai AA, et al. Metabolic Changes Precede the Development of Pulmonary Hypertension in the Monocrotaline Exposed Rat Lung. PLoS One (2016) 11(3):e0150480. doi: 10.1371/journal.pone.0150480.

## Supplemental Information

**Supplementary Table 1. Ingredient and calculated or analysed nutrient composition of the experimental diets for mice, as fed basis**

| | Diets[1] | | | | | |
|---|---|---|---|---|---|---|
| **Item** | **SBM** | **CAS** | **DWP** | **SDPP** | **WGM** | **YMW** |
| **Ingredient, *g/kg*** | | | | | | |
| Maize | 297.5 | 297.5 | 297.5 | 297.5 | 297.5 | 297.5 |
| Dextrose | 132 | 132 | 132 | 132 | 132 | 132 |
| Sugar | 100 | 100 | 100 | 100 | 100 | 100 |
| Arbocell | 50 | 50 | 50 | 50 | 50 | 50 |
| Soybean Oil | 70 | 70 | 70 | 70 | 70 | 70 |
| AIN-93G MX | 35 | 35 | 35 | 35 | 35 | 35 |
| AIN-93-VX | 10 | 10 | 10 | 10 | 10 | 10 |
| Choline chloride | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
| DL-Methionine | 3 | 3 | 3 | 3 | 3 | 3 |
| Soybean meal | 300 | 0 | 0 | 0 | 0 | 0 |
| Casein | 0.0 | 300 | 0 | 0 | 0 | 0 |
| Delactosed whey powder | 0.0 | 0 | 300 | 0 | 0 | 0 |
| Spray dried plasma protein | 0.0 | 0 | 0 | 300 | 0 | 0 |
| Wheat gluten meal | 0.0 | 0 | 0 | 0 | 300 | 0 |
| Yellow meal worm | 0.0 | 0 | 0 | 0 | 0 | 300 |
| **Composition, *g/kg*** | | | | | | |
| Dry matter | 914 | 957 | 930 | 924 | 917 | 929 |
| Crude protein | 153 | 268 | 80 | 238 | 252 | 148 |
| Ash | 43 | 29 | 77 | 28 | 47 | 35 |
| Crude fat | 76 | 65 | 74 | 87 | 70 | 160 |
| Starch | 251 | 249 | 249 | 249 | 268 | 261 |
| Sugar | 295 | 263 | 403 | 263 | 271 | 263 |
| Non starch polysaccharides | 71 | 8 | 17 | 15 | 2 | 18 |
| Gross energy, *KJ/g* | 17 | 19 | 16 | 18 | 18 | 20 |
| Ca | 5.9 | 5.5 | 10.0 | 5.2 | 5.2 | 6.1 |
| P | 3.7 | 3.3 | 6.2 | 1.9 | 2.3 | 4.0 |
| K | 10.4 | 4.0 | 16.5 | 4.5 | 4.0 | 3.6 |
| Na | 1.1 | 1.2 | 5.9 | 8.2 | 1.3 | 1.0 |
| Cl | 1.7 | 2.1 | 10.4 | 12.7 | 1.9 | 1.6 |
| Linoleic acid | 38 | 36 | 36 | 36 | 36 | 36 |
| Electrolyte balance, *Meq/kg* | 266 | 94 | 388 | 115 | 106 | 92 |

[1]Diets: SBM is soybean meal, CAS is casein, DWP partially delactosed whey powder, SDPP is spray dried porcine plasma, SBM is soybean meal, WGM is wheat gluten meal and YMW is yellow meal worm.

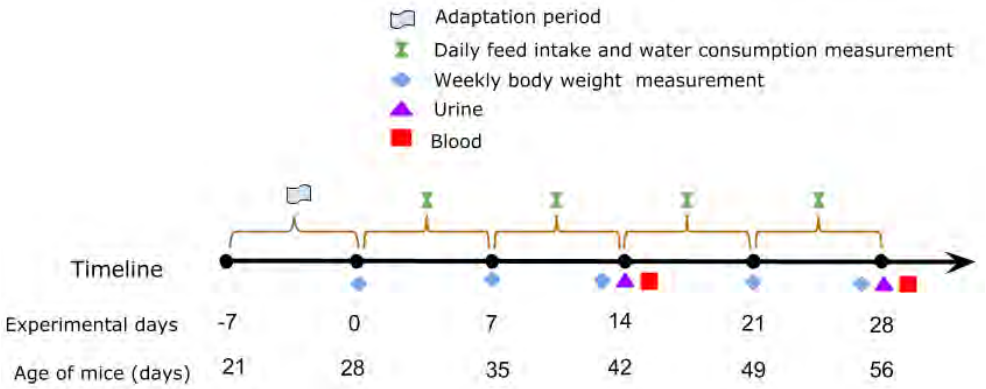
**Supplementary Table 2. Internal standards used in the amine profiling platform.**

| Internal Standards |
|---|
| 2-(4-hydroxy-3-methoxyphenyl) ethyl-1,1,2,2-d4-amine |
| Ala_C13N15 |
| Arg_C13N15 |
| Asn_C13N15 |
| Asp_C13N15 |
| Beta-alanine-2,2,3,3,-d4 |
| Gln_C13N15 |
| Glu_C13 N15 |
| Gly_C13N15 |
| Histamine-a,a,P,P-d4 2HCl |
| L-2-aminobutyric acid-d6 acid |
| L-3-(4-hydroxy-3-methoxy-d3-phenyl)-alanine |
| Leu_C13N15 |
| L-lle C13N15 |
| L-Methionine |
| L-NT-methyl-d3-L-histidine |
| L-ornithine-3,3,4, 4,5,5,-d6 |
| Lys_C13N15 |
| Phe_C13N15 |
| Pro_C13N15 |
| Ser_C13N15 |
| Thr_C13N 15 |
| Trp_C13N15 |
| Tyr_C13N15 |
| Val_C13N15 |

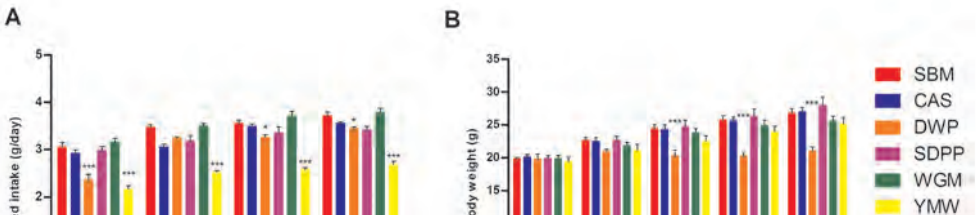
**Supplementary Table 3. Calculated concentrations of apparent ileal digestible essential amino acids (EAA) in the experimental diets and the measured concentrations of EAA in serum samples of d 28.**

| **Diets** | **SBM** | | **CAS** | | **DWP** | | **SDPP** | | **WGM** | | **YMW** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Amino Acids | Diet (g/kg) | Serum (uM/ml) | Diet (g/kg) | Serum (uM/ml) | Diet (g/kg) | Serum (uM/ml) | Diet (g/kg) | Serum (uM/ml) | Diet (g/kg) | Serum (uM/ml) | Diet (g/kg) | Serum (uM/ml) |
| ARG | 9.1 | 189.2 | 10.0 | 120 | 1.5 | 114.4 | 12.5 | 138.4 | 8.3 | 129.2 | 7.6 | 8.6 |
| HIS | 7.9 | 73.2 | 3.3 | 73.6 | 1.2 | 89.1 | 7.0 | 74.4 | 4.9 | 61.9 | 4.7 | 5.3 |
| LYS | 21.3 | 408.7 | 8.9 | 396.1 | 6.1 | 341.3 | 20.1 | 238 | 4.8 | 315.1 | 8.1 | 10.4 |
| MET[1] | 13.6 | 95.9 | 7.7 | 72.2 | 7.0 | 78.3 | 7.4 | 71.3 | 9.7 | 52.5 | 2.6 | 17.6 |
| PHE | 13.3 | 58.2 | 6.6 | 50 | 2.2 | 43.7 | 12.0 | 57.9 | 12.1 | 48.8 | 7.1 | 6.1 |
| THR | 10.4 | 95 | 4.6 | 125.9 | 3.5 | 158.1 | 11.3 | 92.6 | 5.6 | 106.2 | 6.6 | 7.5 |
| TRP | 3.2 | 56.4 | 1.7 | 61.8 | 0.9 | 78.8 | 3.2 | 65 | 2.0 | 54.1 | 2.1 | 7.5 |
| VAL | 16.5 | 126.8 | 5.8 | 172.2 | 3.1 | 104.1 | 13.9 | 134.2 | 9.2 | 125.1 | 10.5 | 11.5 |
| LEU | 9.5 | 142.1 | 24.7 | 163.6 | 6.0 | 109.3 | 20.5 | 146.5 | 16.3 | 115.1 | 15.9 | 15.8 |
| **Corel** | 0.68 | | 0.26 | | 0.43 | | 0.81 | | -0.15 | | 0.13 | |

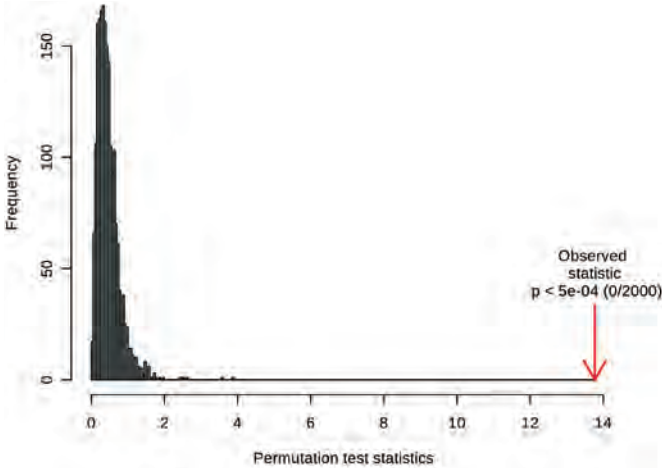
[1] Adjusted for 3 g of methionine used to prepare the experimental diets.



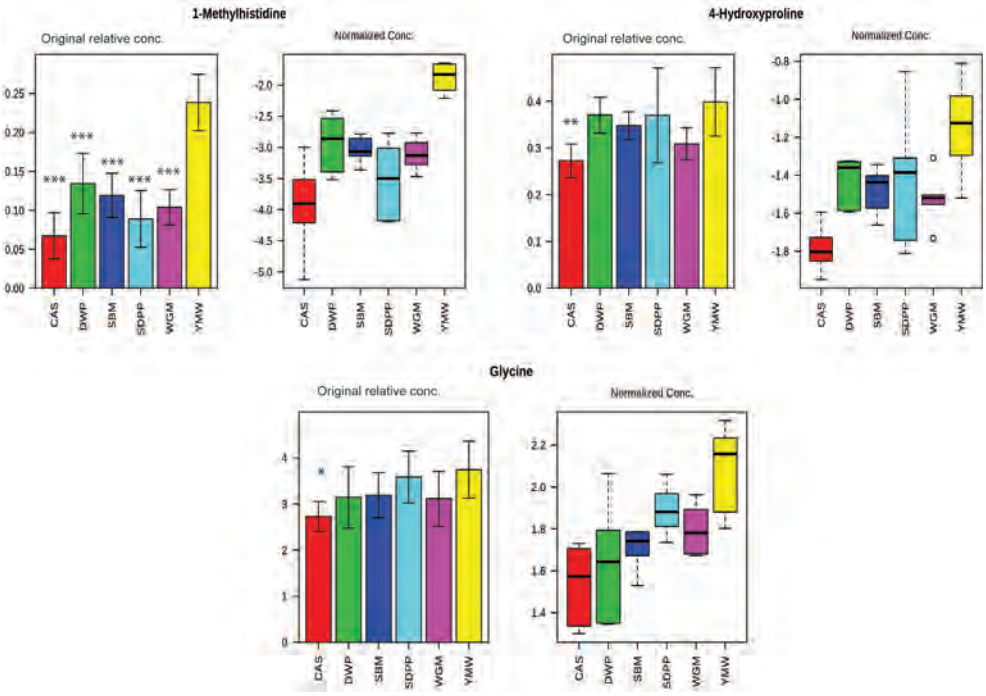

**Supplementary Figure 1.**
Design of the experiment. The solid black dot in the timeline represents the corresponding experimental days/ age of the mice (days). Blood and urine samples were used for amine metabolome analysis.




**Supplementary Figure 2.**
Animal performance parameters (feed intake, water intake bodyweight) measurements of mice fed with different experimental diets. Bars and whiskers represent mean values ± SEM (n = 6). *P < 0.05, ***P < 0.001 compared with SBM-fed mice fed. Here, SBM, soybean meal; CAS, casein preparation intended for animal use; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.

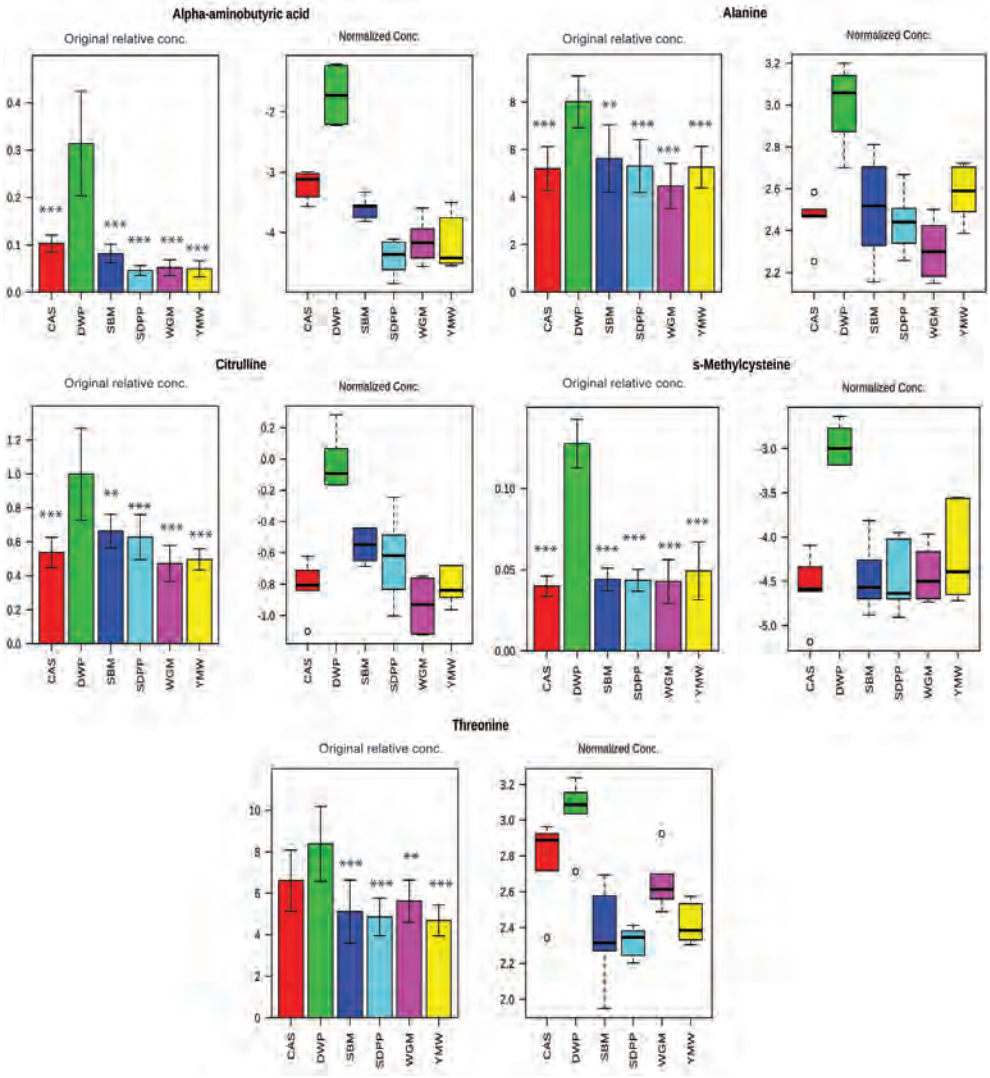



**Supplementary Figure 3: The result of permutation test summarised by a histogram for PLS-DA model built from serum amine of mice at d 28 of the experimental period.**
Red arrow indicates the empirical p value that is calculated by determining the number of times the permutated data yielded a better result than the one using the original labels.

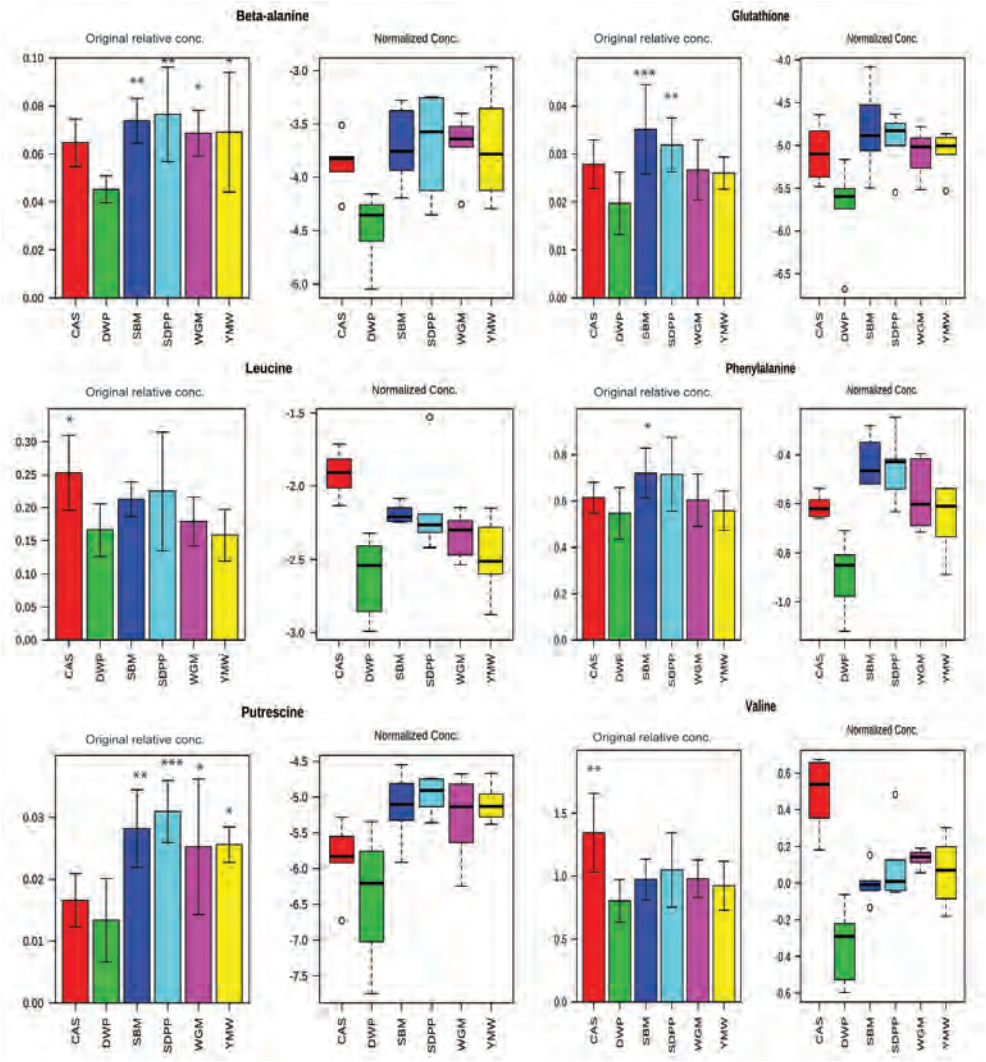



**Supplementary Figure 4: Underlying amines which are present relatively in high level in YMW compared to other (at least in one) treatments.**

According to loading plot, these featured amine are responsible for separation of YMW group in the PLS-DA model built with d 28 serum amine data. Bars are mean values of the original relative concentration and box are mean values of the normalized concentration in the treatment groups; whiskers are ± Standard Error Mean (n = 6). Statistical analysis was performed in the original concentration (i.e. data as relative response ratios) by one-way ANOVA followed by post hoc test (Dunnett test: compared all treatment vs. YMW group) using GraphPad prism version 5.03 for Windows Vista (GraphPad Software, San Diego, California, USA). Statistical significance was defined as $p < 0.05$. *P < 0.05, **P<0.01 and ***P < 0.001 compared with YMW group. SBM, soybean meal; CAS, casein preparation intended for animal use; DWP, partially delactosed whey powder; SDPP, spray dried porcine plasma; WGM, wheat gluten meal and YMW, yellow meal worm.

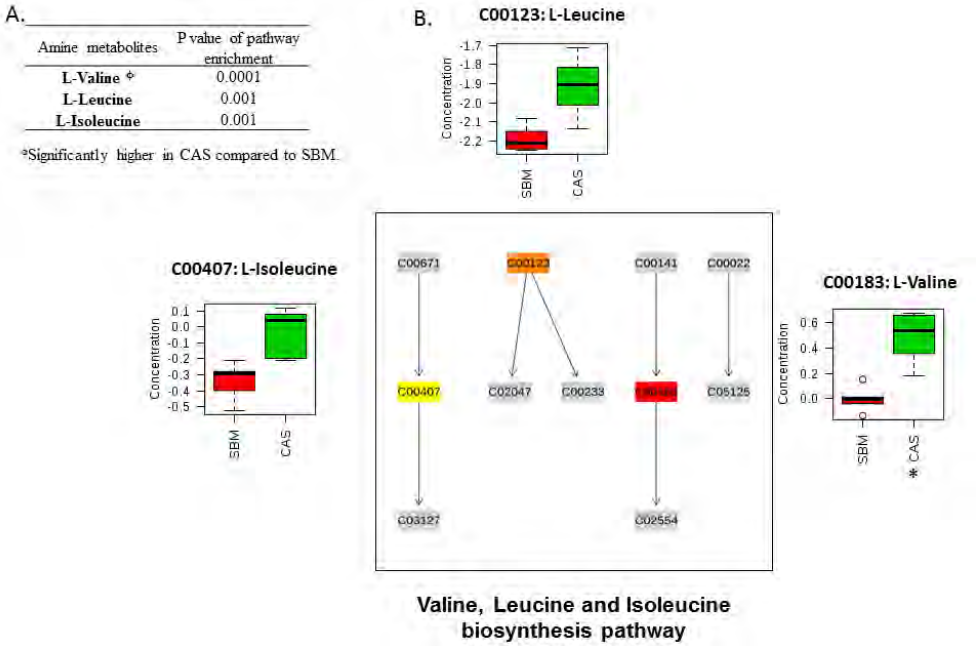



**Supplementary Figure 5: Underlying amines which are present relatively in high level in DWP compared to other (at least in one) treatments.**
According to loading plot, these featured amine are responsible for separation of DWP group in the PLS-DA model built with d 28 serum amine data. Bars are mean values of the original relative concentration and box are mean values of the normalized concentration in the treatment groups; whiskers are ± Standard Error Mean (n = 6). Statistical analysis (see supplementary figure 4.




**Supplementary Figure 6: Underlying amines which are present relatively in low level in DWP compared to other (atleast in one) treatments.**

According to loading plot, these featured amine are responsible for separation of DWP group in the PLS-DA model built with d 28 serum amine data. Bars are mean values of the original relative concentration and box are mean values of the normalized concentration in the treatment groups; whiskers are ± Standard Error Mean (n = 6). Statistical analysis (see supplementary figure 4).

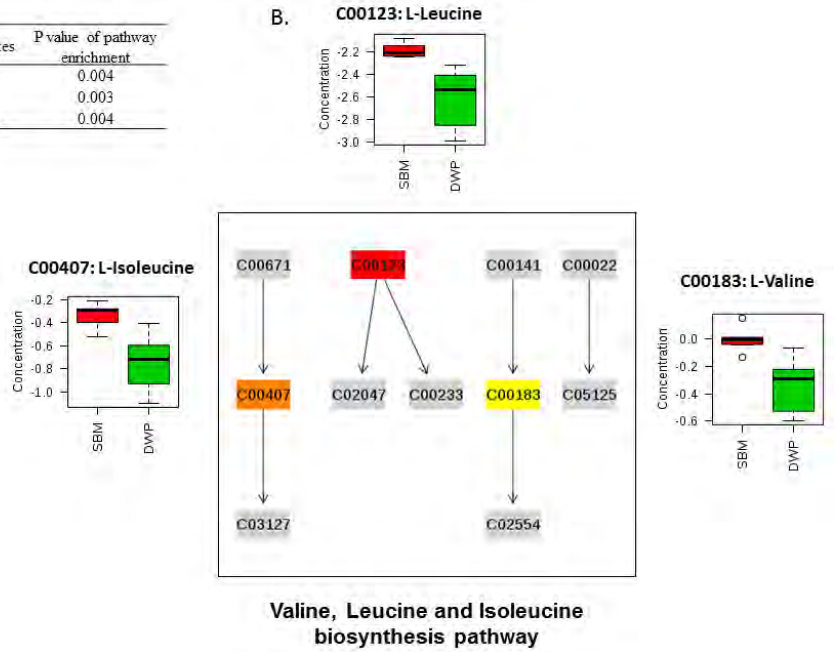



**Supplementary Figure 7: Valine, leucine and isoleucine biosynthesis pathway as represented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enriched in metabolic pathway analysis based on the amine profile in d 28 serum of mice fed CAS and SBM diet.**

A. In bold, amine metabolites that are significantly ($P < 0.05$) enriched in the pathway enrichment analysis. Valine is significantly ($P < 0.05$) higher in serum of CAS fed mice than SBM-fed mice. B. The valine, leucine and isoleucine biosynthesis pathway along with it significant enriched amine metabolites. "Alpha-numeric" representations are the KEGG identifiers. The colour (on gradient scale: cream to red, where towards red denotes lower P value) of the rectangular-nodes in the KEGG pathway are all matched amines according to P values from pathway enrichment analysis. The grey rectangular nodes are the un-matched amines in the KEGG pathway. Box plots are mean value of the normalized concentration in the treatment groups; whiskers are ± Standard Error Mean (n = 6 mice per dietary group).

A.

| Amine metabolites | P value of pathway enrichment |
|---|---|
| **L-Valine** | 0.004 |
| **L-Leucine** | 0.003 |
| **L-Isoleucine** | 0.004 |



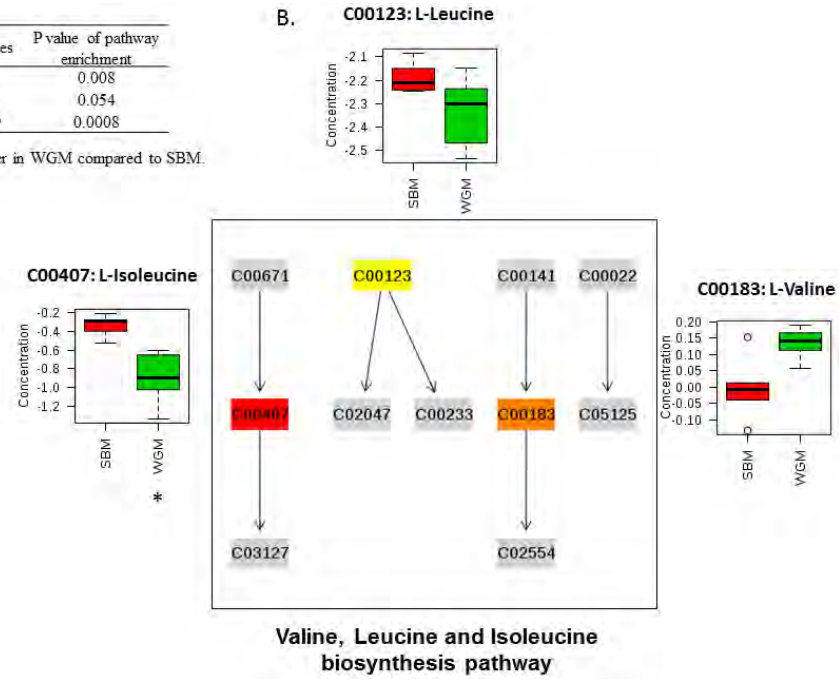

**Supplementary Figure 8: Valine, leucine and isoleucine biosynthesis pathway as represented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enriched in metabolic pathway analysis based on the amine profile in d 28 serum of mice fed DWP and SBM diet.**

A. In bold, amine metabolites that are significantly ($P < 0.05$) enriched in the pathway enrichment analysis. None of the enriched metabolites significantly ($P < 0.05$) differ in DWP from SBM. B. The valine, leucine and isoleucine biosynthesis pathway along with it significant enriched amine metabolites. "Alpha-numeric" representations are the KEGG identifiers. The colour (on gradient scale: cream to red, where towards red denotes lower P value) of the rectangular-nodes in the KEGG pathway are all matched amines according to P values from pathway enrichment analysis. The grey rectangular nodes are the un-matched amines in the KEGG pathway. Box plots are mean value of the normalized concentration in the treatment groups; whiskers are ± Standard Error Mean (n = 6 mice per dietary group).

A.

| Amine metabolites | P value of pathway enrichment |
|---|---|
| **L-Valine** | 0.008 |
| L-Leucine | 0.054 |
| **L-Isoleucine** * | 0.0008 |

*Significantly lower in WGM compared to SBM.



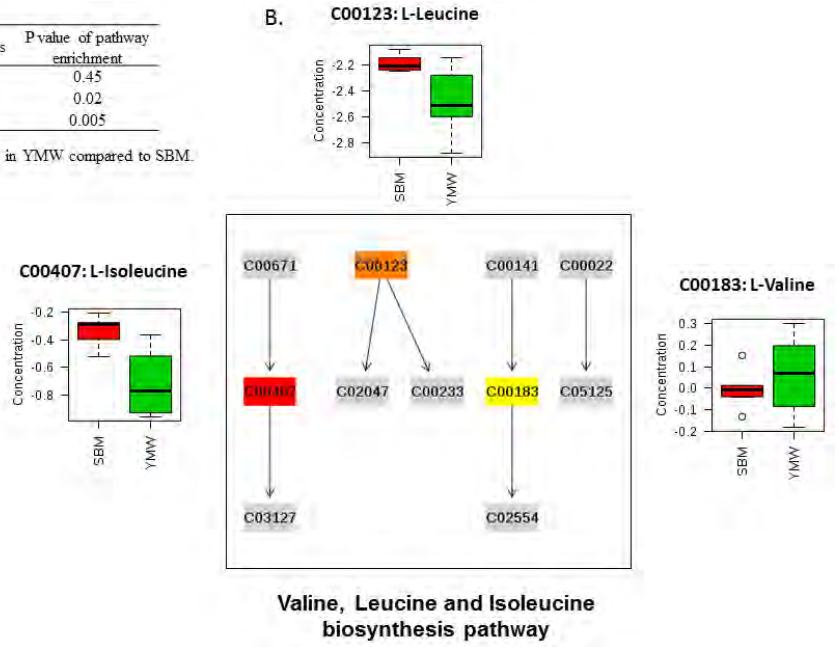

**Supplementary Figure 9: Valine, leucine and isoleucine biosynthesis pathway as represented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enriched in metabolic pathway analysis based on the amine profile in d 28 serum of mice fed WGM and SBM diet.**

A. In bold, amine metabolites that are significantly ($P < 0.05$) enriched in the pathway enrichment analysis. Isoleucine is significantly ($P < 0.05$) lower in serum of WGM fed mice than SBM-fed mice. B. The valine, leucine and isoleucine biosynthesis pathway along with it significant enriched amine metabolites. "Alpha-numeric" representations are the KEGG identifiers. The colour (on gradient scale: cream to red, where towards red denotes lower P value) of the rectangular-nodes in the KEGG pathway are all matched amines according to P values from pathway enrichment analysis. The grey rectangular nodes are the un-matched amines in the KEGG pathway. Box plots are mean value of the normalized concentration in the treatment groups; whiskers are ± Standard Error Mean (n = 6 mice per dietary group).

A.

| Amine metabolites | P value of pathway enrichment |
|---|---|
| L-Valine | 0.45 |
| **L-Leucine** | 0.02 |
| **L-Isoleucine** | 0.005 |

*Significantly lower in YMW compared to SBM.




**Supplementary Figure 10: Valine, leucine and isoleucine biosynthesis pathway as represented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enriched in metabolic pathway analysis based on the amine profile in d 28 serum of mice fed YMW and SBM diet.**

A. In bold, amine metabolites that are significantly ($P < 0.05$) enriched in the pathway enrichment analysis. None of the enriched metabolites significantly ($P < 0.05$) differ in YMW from SBM. B. The valine, leucine and isoleucine biosynthesis pathway along with it significant enriched amine metabolites. "Alpha-numeric" representations are the KEGG identifiers. The colour (on gradient scale: cream to red, where towards red denotes lower P value) of the rectangular-nodes in the KEGG pathway are all matched amines according to P values from pathway enrichment analysis. The grey rectangular nodes are the un-matched amines in the KEGG pathway. Box plots are mean value of the normalized concentration in the treatment groups; whiskers are ± Standard Error Mean (n = 6 mice per dietary group).

# 6

# CHAPTER 6

## Multi-level integration of environmentally perturbed internal phenotypes reveals key points of connectivity between them

*Nirupama Benis*[*1,2], *Soumya Kanti Kar*[*1,2], *Vitor A.P. Martins dos Santos*[3,4], *Mari A. Smits*[1,2,5], *Dirkjan Schokker*[2], *Maria Suarez-Diez*[3].

[1]Host Microbe Interactomics Group,Wageningen University & Research, Wageningen, the Netherlands
[2]Wageningen University & ResearchAnimal Breeding and Genomics, Wageningen, The Netherlands
[3]Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands
[4]Lifeglimmer GmbH, Berlin, Germany
[5]Wageningen Bioveterinary Research, Wageningen University & Research, Wageningen, the Netherlands
*These authors contributed equally to the work

*Frontiers in Physiology, 2017. 8:388. doi: 10.3389/fphys.2017.00388*

## Abstract

The genotype and external phenotype of organisms are linked by so-called internal phenotypes which are influenced by environmental conditions. In this study, we used five existing –omics datasets representing five different layers of internal phenotypes, which were simultaneously measured in dietary perturbed mice. We performed ten pair-wise correlation analyses verified with a null model built from randomized data. Subsequently, the inferred networks were merged and literature mined for co-occurrences of identified linked nodes.

Densely connected internal phenotypes emerged. 45 nodes have links with all other data-types and we denote them 'connectivity hubs'. In literature, we found proof of 6% of the 577 connections, suggesting a biological meaning for the observed correlations. The observed connectivity's between metabolite and cytokines hubs showed higher numbers of literature hits as compared to the number of literature hits on the connectivity's between the microbiota and gene expression internal phenotypes. We conclude that multi-level integrated networks may help to generate hypotheses and to design experiments aiming to further close the gap between genotype and phenotype. We describe and/or hypothesize on the biological relevance of four identified multi-level connectivity hubs.

**Keywords: Data integration, Internal phenotype, Transcriptomics, Proteomics, Metabolomics, Microbiota, Gastrointestinal tract, Systems Biology.**

## Introduction

The information encoded in the genome (genotype) and the external quantitative traits or characteristics (phenotype) of an organism are linked to each other by several layers of so-called, intermediate (Fontanesi, 2016; Leuchter et al., 2014) or internal (Houle et al., 2010) phenotypes. Several of these internal phenotypic layers are shown in Figure 1 that visualizes the conceptual relationship between the external phenotype (P), the genotype (G), the environment (E) and the G&E interactions. The epigenome is tightly associated with the genome and represents the programming of gene expression which is not dependent on the DNA code itself. The transcriptome layer represents direct effects of the environment on the gene expression of the (epi-)genome. Translation of the transcriptome into proteins represents the next internal phenotype. The subsequent layer is represented by complex metabolite profiles. The organism-associated microbiota, especially those in the gut, can be regarded as a separate internal phenotypic layer, because it is not only dependent on the host genome but also heavily influenced by its environment, particularly by nutrition (Montiel-Castro et al., 2013; Schwartz et al., 2012).

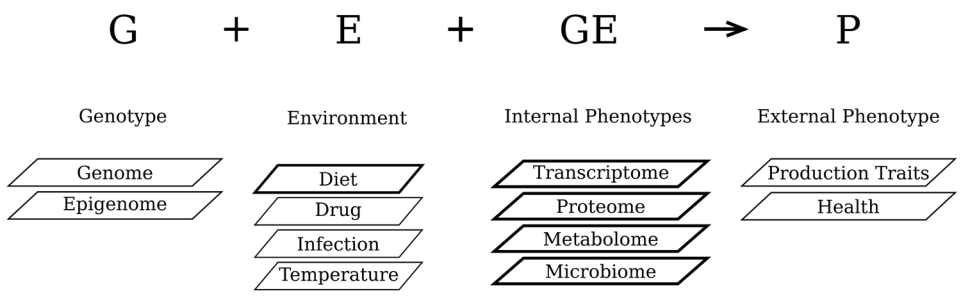



**Figure 1: Relationship between the external phenotype (P), the genotype (G), the environment (E) and the G&E interactions.**

The internal phenotypic layers and the environmental factor with a darker outline are included in the present study.

Although for several traits the quantitative effects of the environment on the external phenotypes are known (Cani et al., 2008; de Wit et al., 2011; Gentry et al., 2004), the specific effects of the environment on the internal phenotypes are largely unknown. Furthermore, it is obvious to assume that the various layers of internal phenotypes are connected to each other and that their joint profiles ultimately determine the external phenotype (Fontanesi, 2016; Leuchter et al., 2014). Unfortunately, most of these assumptions are not based on solid evidence and at best represent oversimplifications of the dynamic nature of processes involved in determining external phenotypes. It, furthermore, partly explains the knowledge gap that exists between the genotype and the external phenotype.

Therefore the objective of this study was to develop methodologies to identify components in the internal phenotypic layers that are connected to components in other internal phenotypic layers. To this end, we integrated multi-scale quantitative (-omics) data using a regression approach. The used data sets were derived from a single experiment with inbred mice which were exposed to five different dietary interventions as a means to perturb the different internal phenotypes. With a data-driven approach we were able to identify a large number of potential connections between the various intermediate phenotypes and for several we found proof of causal relationships in literature. The results of this study provide a basis to understand how various internal phenotypic layers are connected to each other. The identified connections may be crucial for the identification of causal relationships (Civelek and Lusis, 2014) between various biological scales and to uncover mechanisms involved in determining external phenotypes.

## Materials and Methods

### *Origin of data*

We used data from an experiment with six-week old inbred mice that were fed for four weeks with six different semi-synthetic diets (Kar et al. Submitted). In brief: thirty-six 21-day-old C57BL/6J mice (Harlan Laboratories, Horst, the Netherlands) were divided into 6 groups and housed in pairs with ad libitum access to diet and water. After adaptation for one week to a standard diet, the mice were fed semi-synthetic diets containing 300 g/kg (as fed basis) of one of the alternative protein sources for 28 days: soybean meal; casein; partially delactosed whey powder; spray dried plasma protein; wheat gluten meal and yellow meal worm. At the end of the experiment, mice were sacrificed to collect ileal tissue to acquire gene expression data, ileal digesta to study changes in microbiota, blood serum to profile cytokines and chemokines and blood and urine to profile amine metabolites. All procedures were approved by the animal experimentation board at Wageningen University & Research Center (accession number 2012062.c) and carried out according to the guidelines of the European Council Directive 86/609/EEC dated November, 1986. Multi-omics data were obtained with regards to: whole genome gene expression profiles of ileal tissue as measured with Affymetrix GeneChip mouse gene 1.1 ST microarrays (Affymetrix, Santa Clara, CA, USA); community scale microbiota composition of ileal digesta by targeted-amplicon DNA sequencing of the bacterial 16S rDNA V3 region on an Illumina Mi-Seq sequencer; 23 serum cytokine and chemokine concentrations (pg/ml) using a Bio-Rad Mouse 23-plex kit (Bio-Rad, Hercules, CA, USA); and amine metabolic profiles of serum and urine using an ACQUITY UPLC system coupled online with a Xevo Tandem quadrupole mass spectrometer (Waters) operated using QuanLynx data acquisition software (version 4.1; Waters) (Chapter 4 and 5 of this

thesis). The data from the ileum reflects the local effects of the dietary interventions, the other three data assess the systemic effects.

***Pre-processing and selection of data***

An overview of the five types of data and their specifics are given in Table 1. Each dataset was pre-processed in a similar way using the R package limma (Smyth, 2005) to find the differentially significant data-points. The data is first log transformed and then this data is fitted to a linear model using the function lmFit (Phipson et al., 2016) which will give back information on the differences between the genes in different arrays and subsequently different comparisons of control vs treatment. Then we used the function eBayes (Phipson et al., 2016) which applies an empirical Bayes method to compute p-values for a t-statistic under the assumption that only 1% of the genes are differentially regulated among all the genes in the arrays. This p-value is then subjected to a Benjamini-Hochberg (Benjamini and Hochberg, 1995) multiple testing, also known as a False Discovery Rate (FDR).

This analysis was done by comparing the data of each dietary group against the data of the dietary group that received soy bean meal as protein source, which is the most common source of protein in animal feed. The FDR value of the data, is used to gauge significance and data-points that were significant in at least one of the five comparisons of the diets were included in the integration analysis. Except for the Cytokine and Metabolomics Serum (using the amine measurement), all the data-types had some samples not included due to quality control. Two types of metabolomics measurements were done on the sampled urine; Amine and Acyl-carnitine. The amine dataset did not have sufficient statistically significant data-points so was discarded. We only work with the Acyl-carnitine measurement in urine.



**Table 1: Pre-processing and specificities of each data-type.**

| Properties | Transcriptomics | Microbiota | Cytokine | Metabolomics Serum | Metabolomics Urine |
|---|---|---|---|---|---|
| Sampling | Ileum | Ileum | Serum | Serum | Urine |
| Before pre-processing | 16,410 * 33 | 148 * 33 | 23 * 36 | 41 * 36 | 16 * 28 |
| After pre-processing | 52 * 33 | 22 * 33 | 13 * 36 | 26 * 36 | 16 * 28 |

Details of the site of sampling and data dimensions before and after pre-processing are indicated. The first number indicates the number of variables in the data and the second number denotes the number of samples.

***Data integration, network generation and network assessment.***

All significantly different data-points were used in the integration which was initially performed with two datasets at a time, so that from the 5 datasets 10 integrated networks were generated. The integration was performed using the function sPLS (sparse Partial Least Squares) in regression mode with ncomp = 5, from the R package mixOmics (Dejean et al., 2011; González et al., 2012; Lê Cao et al., 2009). The regression mode is used to model causal relationship between variables in both datasets by identifying combinations of variables between both datasets. Weight vectors used in the regression modelling are termed loading vectors. sPLS is used to perform simultaneous variable selection in the two datasets to be integrated and employs LASSO (Least Absolute Shrinkage and Selection Operator) penalization (Tibshirani, 2011) on the loading vectors. This approach requires one data set, $\boldsymbol{X}$ with $n_x$ elements, to be designated the predictor and the other, $\boldsymbol{Y}$ with $n_y$ elements, the response. As an output, the approach produces a matrix $Ma(\boldsymbol{X},\boldsymbol{Y})$ of size $n_x$ x $n_y$ representing the relevant correlations between both datasets, so that:

$$ma_{ij} = \begin{cases} 0, if\ Y_j\ independent\ of\ X_i \\ cor(X_i, Y_j), \quad if\ Y_j\ dependent\ on\ X_i \end{cases}, with\ i\ \in \{1, \dots, n_x\}\ and\ j\ \in \{1, \dots, n_y\}$$

where $cor(X_i, Y_j)$ is Pearson's correlation between elements *i* and *j* from datasets **X** and **Y** respectively. The correlation is computed across all available samples (here corresponding to dietary exposures).

Since it is not trivial to determine the predictor and response with biological data, we swapped the two types of data to compute $Mb(\boldsymbol{Y}, \boldsymbol{X})$, a matrix of size $n_y$ x $n_x$ where the roles of **X** and **Y** have been interchanged. Both matrices, *Ma* and *Mb* where combined into a final matrix $M(\boldsymbol{X},\boldsymbol{Y})$ size $n_x$ x $n_y$ using

$$M(\boldsymbol{X},\boldsymbol{Y}) = Ma(\boldsymbol{X},\boldsymbol{Y}) + t(Mb(\boldsymbol{Y},\boldsymbol{X})$$

where *t* represents matrix transposition. Thus, non null elements of the matrix $M(\boldsymbol{X},\boldsymbol{Y})$ represent correlations between data types that have been deemed associated. This matrix can be seen as a weighted adjacency matrix representing a network where two nodes $X_i$ and $Y_j$ are connected via an edge if a non-null weight can be assigned to the edge. This weight is represented by the matrix value $m_{ij}$.

To further prune the network of (possibly) spurious interaction two additional thresholds $(th_l < 0$ ; and $th_h > 0)$ were imposed to obtain an unweighted adjacency matrix $A(X, Y)$ of size $n_x$ x $n_y$

$$A_{ij} = \begin{cases} 1 \; if \; m_{ij} \geq th_h \; or \; m_{ij} \leq th_l \\ 0 \; if \; |m_{ij}| < |th_l| \; and \; m_{ij} < th_h \end{cases}$$

where |x| represents the absolute value. $th_l$ and $th_h$ were selected for each network so that only top 5% of the highest (positive) and lowest (negative) weights were kept for building the networks.

Networks represented by these adjacency matrix were transformed into the edge-list format, a two column table of the connected nodes in a network where each row represents an edge and visualized in Cytoscape (Ono et al., 2015; Shannon et al., 2003). For each pair of integrated datasets a null model of the association networks was constructed using a strategy based on random permutations of measured values (Saccenti et al., 2015). Measured data-points were randomly permuted over samples before data integration to obtain randomized datasets that still retained the same value distribution for each variable. The randomized datasets were then used for data integration following the afore mentioned approach thereby generating randomized associations networks. The process was iterated $N_{it}$ = 1,000 times for each pair of datasets; For each iteration, *k*, the values of the dynamic cut-offs ($th_{lk}$ and $th_{hk}$) (5% of the highest and lowest correlation) were recorded. For the ten pairwise combinations of datasets, the values obtained for $th_l$ and $th_h$ obtained using the unpermuted dataset, were compared with the distribution of values of $th_{lk}$ and $th_{hk}$ with k = {1,...,$N_{it}$} to get networks from the random data to compare to the networks from the biological data.

***Network merging and topological analysis***

The ten networks arising from pair-wise data integration of the 5 data sets were merged in a combined network including all the nodes and edges of the 10 networks. This network is then restricted by only including nodes present in at least two of the separate networks. We used the igraph R package (Csardi and Nepusz, 2006) to further analyse the network, which was treated as non-directed, since no particular directionality was assigned to the edges. We obtained values for the following topological properties of the merged network (Barabasi and Oltvai, 2004; Csardi and Nepusz, 2006; Zhu et al., 2007): Degree: number of neighbors of a given node, that is the number of nodes connected to it. Clustering coefficient of a node is the ratio of the number of connections between the neighbors of a node and the total number of possible connections between said neighbors. Characteristic path length: median of the average distance between a node and all the rest. Network density: ratio between the total number of existing edges and the total number of possible edges (given the number of nodes in the network). Connected components maximal subgraphs in a network such that each node is connected to all the rest by means of network paths. For node level metrics, such as

degree or clustering coefficient average values were computed over all nodes. Cytoscape was used for network visualization.

***Literature mining***

To investigate the co-occurrence of the names of the connected nodes in the correlation network, we used the R package rentrez (Winter, 2016). This package searches for selected keywords in PubMed abstracts while making use of the MeSH (Medical Subject Headings) thesaurus to maximize results via the API from NCBI. The search was not restricted to a specific tissue type or organism. These results were examined, although not exhaustively, to find literature evidence of established relationships between nodes connected through identified edges; these were then considered as true positive search results.

The script used to generate all these results will be made available on request. All the above mentioned operations were performed using existing functions from R packages.

## Results

***Analysis of the individual datasets***

A dietary intervention was performed on mice where the protein content was changed and multi-omics data were obtained with regard to whole genome gene expression profiles of ileal tissue (Transcriptomics), community scale microbiota composition of ileal digesta (Microbiota), 24 different cytokine levels in blood serum (Cytokine) and protein-associated metabolic profiles of serum (Metabolomics Serum) and urine (Metabolomics Urine). These data were pre-processed and analysed separately by fitting a linear model on the data-points and looking for differentially expressed readouts in each treatment versus the control. Each dataset had its own p-value (corrected for multiple testing with the Benjamini-Hochberg method) threshold, ranging from 0.001 to 0.1 for difference between the tested and reference diet. The highest number of statistically significant entities was found in Transcriptomics. Furthermore, all the measured variables in Metabolomics Urine were found to be significantly different in at least one comparison.

***Pairwise data association and network generation***

We performed the integration by linking two data-types at a time and in such a way that after the pairwise analysis all the observed correlation data could be combined to build a multi- level interaction network. Therefore, each data-type was integrated with the other four types of data, resulting in ten correlation networks. The topological characteristics of all these ten networks are given in Table 1 and Figure 2 and the network graphs are available in Supplementary Figure S1 as an image.

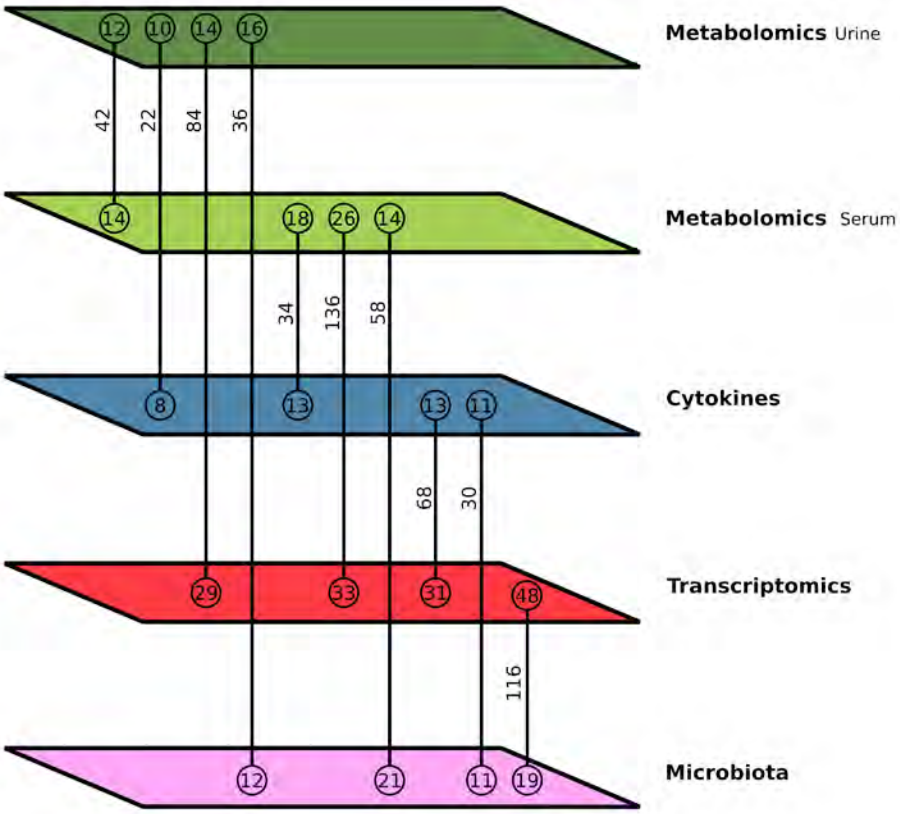



**Figure 2: Multilevel integration.**

This schematic image shows the number of connections between each internal phenotypic level with the other levels in a merged network. The colours of the parallelograms denote the internal phenotypic level to which the data-types belong. Green is Metabolomics (light green – Metabolite from Serum and dark green – Metabolite from Urine), blue is Cytokines, red is Transcriptomics and pink is Microbiota. Each line connects two levels and the vertical number above the line indicates the number of edges in the correlation network between those two phenotypic levels. The number of connected nodes in each level is given in circles above and below the connecting lines.

Supplementary File 1 (available on request) has the networks in a format that can be uploaded into Cytoscape in order to further explore the connectivity's of these networks by simply clicking on these nodes. Table 1 shows the higher and lower thresholds that were used separately for the correlation network. Connections between pairs of data points with correlation values between the threshold values, i.e. and the Edges corresponding to Low Threshold (negative threshold) and High Threshold (positive threshold) as indicated in Table 2 were discarded from the final network. There were two disconnected sub-graphs in five of the networks while the other five have only a single, fully connected graph.

**Table 2: The 10 individual correlation networks.**

| Network Names (data A & data B) | Low Threshold | High Threshold | # of nodes (A) | # of nodes (B) | Connected components |
|---|---|---|---|---|---|
| Metabolomics Urine & Cytokine | -0.32 | 0.55 | 10 | 8 | 2 |
| Metabolomics Urine & Transcriptomics | -0.55 | 0.54 | 14 | 29 | 1 |
| Metabolomics Urine & Microbiota | -0.28 | 0.42 | 16 | 12 | 1 |
| Metabolomics Serum & Metabolomics Urine | -0.51 | 0.6 | 14 | 12 | 2 |
| Metabolomics Serum & Cytokine | -0.33 | 0.5 | 18 | 13 | 2 |
| Metabolomics Serum & Transcriptomics | -0.31 | 0.35 | 26 | 33 | 2 |
| Metabolomics Serum & Microbiota | -0.38 | 0.3 | 14 | 21 | 1 |
| Transcriptomics & Cytokine | -0.27 | 0.34 | 31 | 13 | 2 |
| Microbiota & Cytokine | -0.38 | 0.35 | 11 | 11 | 1 |
| Microbiota & Transcriptomics | -0.28 | 0.27 | 19 | 48 | 1 |

Each row represents one of the 10 correlation networks. Low Threshold and High Threshold represent the thresholds used for the correlation values.

The largest network, in terms of nodes, is the Microbiota & Transcriptomics network. This seems logical as it represents the most comprehensive datasets and spacial interactions between the two data-types are known to occur. Overall, networks involving Transcriptomics data had higher number of nodes than other networks. The smallest network with 18 nodes and 22 edges was the Metabolomics Urine & Cytokine network.

***Technical validation of pairwise integration networks by random permutation***

We performed the same method of integration on the five different data-types after randomly permuting the measured data, this process was iterated a 1000 times. In this way, the networks obtained from random permutations are considered a null model with no biological information and used to assess the significance of the results obtained with the non-permuted data. Figure 3 shows the spread of correlation values for the integration of Metabolomics Serum and Transcriptomics. The thresholds for network reconstruction were selected so that only the 5% highest and lowest correlations were kept. The separation between the values obtained for the integrated data and the randomly permutated datasets indicates the high significance of the edges in the integration networks. In this way, selection of the 5% highest and lowest correlations and significant limits the number of spurious correlations that could be due to chance alone while retaining maximum information in the networks.

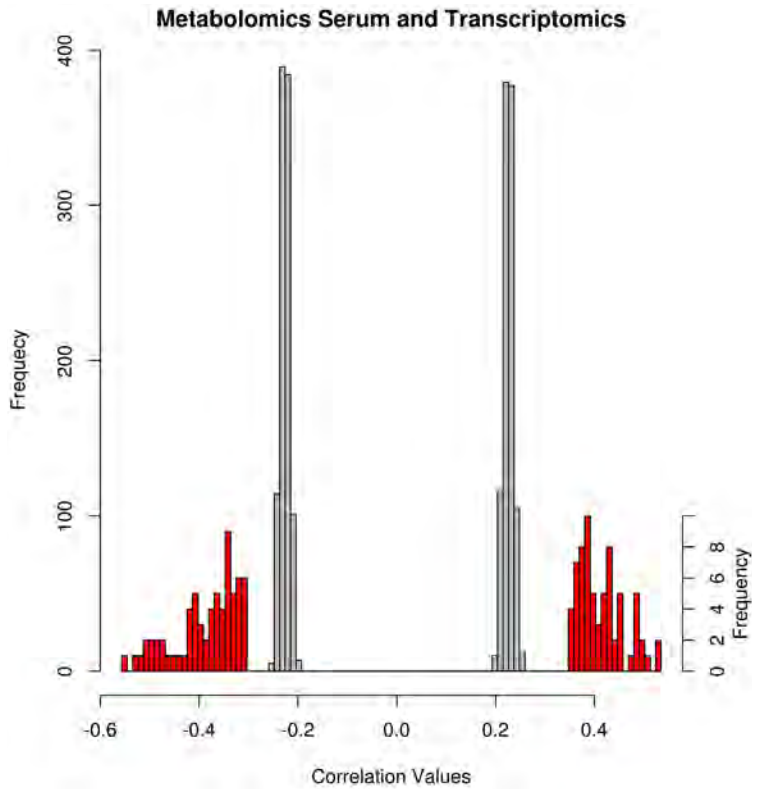



**Figure 3: Distribution of network correlations and random network cut-offs of the Metabolomics Urine and Transcriptomics networks.**

The x-axis depicts the range of correlation values and the y- axis shows its frequency. The grey bars denote the distribution of the thresholds of the 1,000 random correlation networks with frequency on the left y-axis. The red bars are distributions of the correlation values of the inferred network with frequency on the right y-axis.

Similar results were obtained for most of the integration networks (Supplementary Figure 2). In three of the networks, there is an overlap between the correlation values from the inferred network and the values arising from the randomly generated networks. The overlaps are in the networks Metabolomics Urine & Microbiota, Metabolomics Urine & Cytokine and Transcriptomics & Cytokine network. The highest overlap appears in the first two and mostly affects edges with negative correlations.

***Merged Network***

All the 10 integration networks (Supplementary File 1) were merged and only nodes linked with nodes of at least two other data-types were kept (see Table 3). The gene expression data has the highest number of nodes in the merged network. However, nodes with the highest degree (number of connecting edges) arise from the microbiota data, with S24-7 having 57 neighbors and Bifidobacterium having 47 neighbors. The merged network encompasses 45 nodes that are connected to all the other types of data. For that reason we denote them 'Connectivity hubs' and they are included in Table 3 and Supplementary Table 1 (available on request).

**Table 3: Characteristics of the merged network.**

| Network Statistics | |
|---|---|
| Total number of nodes | 112 (45) |
| Total number of edges | 577 |
| Number of Metabolomics Urine nodes | 15 (8) |
| Number of Metabolomics Serum nodes | 24 (11) |
| Number of Cytokine nodes | 13 (7) |
| Number of Transcriptomics nodes | 43 (12) |
| Number of Microbiota nodes | 17 (7) |
| Degree range | 2 to 57 |
| Average number of neighbours | 10.35 |
| Clustering coefficient | 0.20 |
| Characteristic path length | 2.31 |
| Network density | 0.09 |
| Connected components | 1 |

Characteristics of the merged correlation network. The number of nodes from each data-type are given in rows four to eight. Between brackets the number of connectivity hubs is indicated.

***Functional validation of merged network by text mining***

A PubMed literature search for co-occurrence of linked nodes gave results for 6% of the links corresponding to 37 edges. We further investigated reported causality effects between the nodes in question. Most of the retrieved results are related to metabolites and cytokines measurements whereas a few results confirming causal relationships were found involving gene nodes. We were able to find literature confirmation pertaining to associations for six out of the ten pair-wise connections between phenotypes, as summarised in Table 4 and Supplementary Table 2 (available on request). Among the nodes with literature results, four are from Microbiota, two from Transcriptomics, 15 from Metabolomics Serum, three from Metabolomics Urine and six from Cytokines. The node with the highest number of hits in literature is Tnf $\alpha$ which co-occurs 8,563 times with nine metabolites from the Metabolomics Serum data and one bacterial group *(Bifidobacterium)*.

**Table 4: Overview of text mining results.**

| Data Connections | PubMed Ids | Distinct edges |
|---|---|---|
| Cytokines & Metabolomics Serum | 9,554 | 16 |
| Metabolomics Serum & Metabolomics Urine | 906 | 6 |
| Microbiota & Metabolomics Serum | 254 | 7 |
| Microbiota & Cytokines | 250 | 5 |
| Transcriptomics & Microbiota | 83 | 3 |
| Metabolomics Serum & Transcriptomics | 59 | 2 |

The first column shows the types of data that are connected by the edges that were found in the PubMed literature search.

Of the 30 data-points from all the types of data that have literature results, 15 are connectivity hubs. One such connectivity hub is Glutathione (GSH) which has 21 direct neighbors from four data-types as shown in Figure 4. This hub is especially interesting because six of the connected nodes (Carnitine, Tnf $\alpha$ , Il-1b, Il17c, Bifidobacterium and Dapk2) have textual co-occurrences found by the text mining algorithm. The terms GSH and Tnf $\alpha$ were found 2,231 times in the abstracts of Pubmed indexed articles. Full text inspection shows that some of the connections are causal relationships as one of the connected nodes activates or inhibits the other.

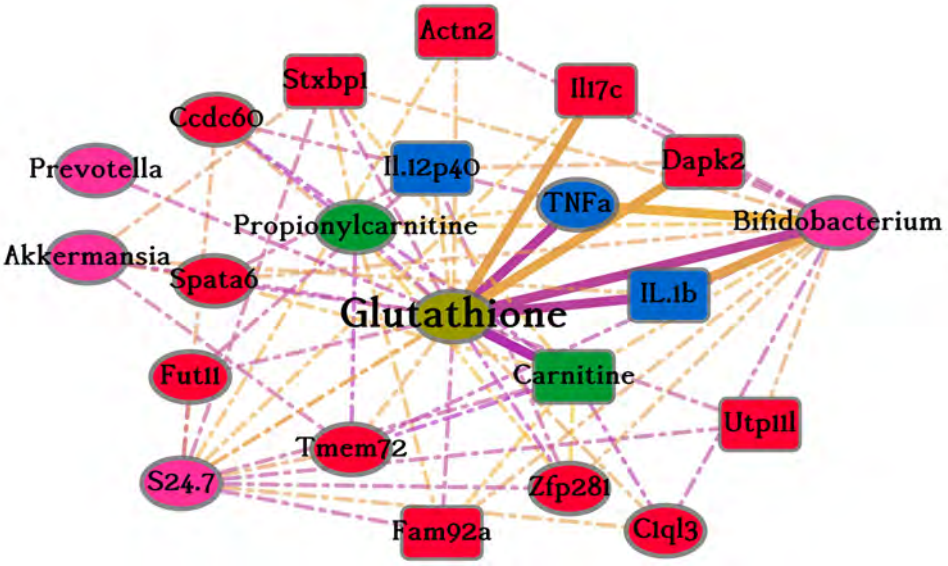



**Figure 4: Glutathione sub-network.**

This figure shows the 21 connections of the node Glutathione in the merged network. The different colours of nodes indicate the data-type of internal phenotypic level of that node, pink is Microbiota, red is Transcriptomics, blue is Cytokines and green is Metabolites (light green – Metabolites from Serum and dark green – Metabolites from Urine). Oval nodes are connectivity hubs. Dotted lines show un-validated edges and continuous, thicker edges show connections also present in the results retrieved from scientific literature. Edge colour, yellow and purple, indicates positive and negative correlations, respectively.

## Discussion

In this study we developed and used a set of computational methods to identify components in internal phenotypic layers that are connected to components in other internal phenotypic layers of an organism. We successfully integrated multi-scale quantitative (-omics) data, derived from a single experiment with inbred mice and which were exposed to five different diets. Since the data originated from an animal experiment that was not designed for the detection of genetically and/or dietary induced differences in external phenotypes, we only focused on the connectivity between 5 intermediate phenotypic levels. Some studies have reported pairwise data integration of two (Benis et al., 2015; Lu et al., 2014; Rajasundaram et al., 2014) or three data sets (Adourian et al., 2008). But this is, to the best of our knowledge, the first time that an integration of such heterogeneous data-types from different tissues, arising from a single experiment, has been reported. The approach as described here could, in principle, be applied on any number and type of datasets, as long as they are from the same experiment, from samples at the same time-point and have comparable dimensions of differentially regulated data.

***Internal phenotypic data and pairwise data integration***

Each used data-type represents a different internal phenotype and a different layer of the system that (co-) drives the manifestation of external phenotypes. We subjected each data-type to a separate analysis in order to correlate only those changes induced by the dietary intervention. Nodes with significantly different values could easily be identified in each of the sampled tissues and fluids (ileum, blood and urine) thereby representing the local and systemic effects of the interventions and the need of a multi-scale approach.

In order to investigate connections between the five data-types we used sPLS, an integration method that can be applied to several types of data, two at a time. This method can also handle the dimensionality problem of biological datasets where the number of variables is usually higher than the number of samples. sPLS has been previously used for integration of microbiota with gene expression data (Benis et al., 2015; Steegenga et al., 2016) and measurements on cell wall polysaccharides of fibers with phenotypic characterizations of fibers in cotton balls (Rajasundaram et al., 2014).

We performed pairwise integration of the datasets, resulting in ten networks with varying spreads of correlation values. Deciding on a threshold to distinguish genuine from spurious correlations is a major bottleneck for the definition of correlation networks. While a 0.8 threshold (absolute value) has been suggested for gene expression data (Schäfer and Strimmer, 2005), other authors suggested smaller values (0.6) in metabolomics data sets (Camacho et al., 2005). The correlation values greatly depend on the biological dataset under study and its dimensionality. There are several methods to choose a threshold based on the data: use assigned p-values as threshold; use network characteristics of the correlations; or use a percentage of the correlation distribution. Borate and colleagues (2009) concluded that threshold selection methods based on network properties such as the clustering coefficient are best for gene co-expression networks. This would not work here because the generated networks always induce connections between data points of different type and as a result they have a zero clustering coefficient for every node. While integrating two types of metabolomics datasets with gene expression of the tissues in which they were measured Adourian and colleagues (2008) assigned p-values to the correlation values and then set a threshold. In this study, we used the top 5% of the correlation values because this dynamic threshold (separate for the positive and negative values) eliminates bias towards the size of the datasets.

We further validated the observed correlations by comparing them with a null model obtained by randomly permuting the data along the samples (Eguíluz et al., 2005; Saccenti et al., 2015). In two of the networks, Metabolomics Urine & Microbiota and Metabolomics Urine & Cytokine (the smallest network), the significance of the negative

correlation values could not be established as we observed a considerable overlap between the negative correlation values of this network and the negative thresholds of the random networks. This calls for caution when biologically interpreting these networks. For five of the networks we observed a very clear separation of the random thresholds and the start of the correlation values in the network (Supplementary Figure 2). The other networks showed slight overlaps between the random threshold distribution and the network correlation distribution. This extra validation step reassured us that the observed correlations are rooted in biological phenomena. To our knowledge this technical validation step is not common in current studies of this type.

The edges of the inferred networks, indicate significant computationally-determined correlations between values of connected nodes. Our approach does not require a mechanistic model on how the associations are established and in each network these associations may be caused through entirely different mechanisms. In some cases the associations would be due to causal relationships between the connected nodes, such as increased expression levels of a cytokine gene linked to increased cytokine levels. However, in many cases, the associations could be indirect, mediated by elements that have not been measured in the experimental set up. In a formal mathematical model, they are considered hidden variables. Such would be the case of, for example, the changes in the metabolite levels of urine. These changes might have been caused by the colonic microbiota, in turn affected by the ileal microbiota. Since we only used the ileal microbiota data, we observe correlations between the ileal microbial populations and the urine metabolite levels which could be in reality, indirect relationships mediated by the colonic microbiota.

***Network of connected internal phenotypes***

The pair-wise integration method allowed us to merge the ten individual networks into a single network. Correlations within a dataset were deliberately excluded from this study because we only wanted to focus on connections between different internal phenotypes, where little work has been done. Thus in the ten networks, all detected connections are between two different data types and every node has a zero clustering coefficient. However, in the merged network, a non-zero clustering coefficient emerges as a result of nodes connecting to multiple data types (Table 3). This emphasizes the biological relevance of this method because the ten networks were built without any information on cross-linking. Thus, we identified individual nodes that directly or indirectly participate in processes of the other four individual networks. Because they seem to connect different internal phenotypes, we denoted them 'Connectivity Hubs'. Starting the procedure as developed and applied here with networks with non-zero clustering coefficients (correlating within a dataset) would, however, not alter the connections between internal phenotypes.

***Functional validations of phenotype connections***

Results of the text-mining were used to validate some of the identified links. This revealed insights into the mechanistic relationships between the variables predicted to be linked to each other. 37 of the 577 (6%) computational inferred links have already been described in literature as detected by our text-mining approach, which was not exhaustive because it focused only on text in journal abstracts. This indicates that our method identifies currently known biological interactions. The rest of the predicted links have not been discovered and investigated yet, have not been mentioned in abstracts, or do not exist in the biological system. Furthermore, by inspecting some of the retrieved abstracts and corresponding articles, we were even able to find causal relationships between some of the computational identified nodes where one of the nodes was used as an experimental perturbation and the other node was measured as a response parameter. Several indirect associations were also validated through reports on experiments where nodes, found to be connected in this study, were measured in response to another perturbation. During text-mining, in order to retrieve as many results as possible, search terms were matched against the MeSH thesaurus, irrespective of the organism and all the synonyms were included in the search. The downside to this approach is the inclusion of several false textual associations. The most striking case is that of the identified association between Glutathione and Il17c. In the literature results, the reported association is between Glutathione and Il17a and not Il17c. Through the thesaurus, Il17c was mapped to Il17 and subsequently to Il17a thereby giving rise to that falsely identified association in literature.

In order to increase the precision and recall of text mining searches and overcome problems associated to the use of a thesaurus, one needs to move from mining text, to mining the knowledge embedded in the text and the use of data hidden in public databases. Such an approach requires the use of knowledge management tools and representations that can be automatically accessed (Antezana et al., 2009). Semantic web technologies represent a new class of tools that include natural language processing, ontologies, machine learning algorithms and much more to facilitate integration knowledge from heterogeneous sources. The expansion of the use of semantic technologies in the life sciences domain will allow associating concepts such that inferences on causality, regulation, organism or tissue can be made using high-throughput methods and automated reasoning.

Among the interactions retrieved from the automated literature search, a high prevalence of associations involving cytokines and/or metabolites was observed. In fact, such type of interactions represent 97% of the retrieved results. This probably highlights the extraordinary amount of work that has been done in these types of data in the past. On the opposite extreme, only 8% of the retrieved interactions involved associations between the expression of genes, reflecting the fact that most of the available gene

expression data originates from genome-wide techniques. In such type of experiments, papers, especially abstracts, usually report on systems behaviors and pathways and less frequently on the individual behavior or role of individual genes and connected response nodes.

***Validated connectivity hubs***

Even though we only performed integrations of two datasets at a time, we find data-points (metabolites, cytokines, genes or microbial groups) that correlate with different types of data. We identified 45 connectivity hubs in the merged network that seem to have associations with all four types of data. More than 30% of them are involved in links that were retrieved in literature. To further support the biological relevance of identified multi-level connectivity's we discuss the implications of two of the 15 biologically validated connectivity hubs as examples. The two connectivity hubs were chosen because of the large amount of literature results for these hubs. The first hub, Tnf $\alpha$ has the highest number of literature results among all the nodes in the network and the other hub, Glutathione, has literature validations to the most number of data-types. Tnf $\alpha$ is a connectivity hub in the merged network, with links to several neighbors belonging to the four other types of data. The position of this cytokine in our merged network shows that it plays a role in processes of the other internal phenotypes. The literature validated links are between Tnf $\alpha$ and two other types of data (Metabolomics Serum, Microbiota). Many of the validated links represent causal relationships. With regards to immune responses and as a drug target, Tnf $\alpha$ has been studied in great detail (Cicha and Urschel, 2015). The un-validated edges show that Tnf $\alpha$ could be a regulator of other internal phenotypes as well, than currently known.

The metabolite Glutathione (GSH) was measured in the serum and in the merged network is a connectivity hub proving that it is vital part of the system that connects several internal phenotypes. Among the 15 connectivity hubs with functionally validated links, GSH is the only one that has validated links to all other data-types based on our literature mining. These results support our claim of GSH being a connectivity hub, a biological component influencing several internal phenotypes. Several PubMed results for GSH are from in-vivo studies where GSH was administered to alleviate symptoms of a disease. Our literature results show that GSH has been studied in relation to all different types of data. Of the six validated links in our merged network, five represent proven causal relationships (see Figure 4 and discussion of the functional validation). These neighboring nodes in the merged network are mostly related to immune and homeostatic mechanisms. GSH is a tripeptide, ubiquitously distributed in living cells and plays an important role in the intracellular defense mechanism against oxidative stress (Couto et al., 2016; Diaz-Vivancos et al., 2015). It is known that GSH metabolism is very important for the antioxidant and detoxifying action of the intestine. It is also essential for

the maintenance of the luminal thiol-disulfide ratio involved in regulation mechanisms of the protein activity of epithelial cells (Iantomasi et al., 1997) which could be important since the intervention is changed in protein. Our results also demonstrate the manifold and central role of GSH when it comes to proteins, peptides and amino acids in nutrition. These observations indicate that the presented merged network represents, at least in part, associations of biological phenomena.

***Potential relevance of selected connectivity hubs***

There are 30 connectivity hubs in the merged network that do not co-occur with their connected nodes in our literature search. However, the prominence of these nodes in our merged network indicates that they could represent potential relevant interactions with components of the other internal phenotypes. In order to demonstrate how the results of this study may be used to hypothesize on functional relationships between different molecular components, we here describe the potential biological relevance of two highly linked connectivity hubs, Tmem72 and *S24-7*. Both hubs are not yet described in literature abstracts in conjunction with other data-types.

The high number of connectivity hubs in the Transcriptomics layer suggest that the expression of several intestinal genes is involved in many more interactions than currently known. None of the observed Transcriptomics connectivity hubs popped-up in our literature mining results. The most highly connected Transcriptomics node, Tmem72 (Transmembrane Protein 72), has only been studied in the kidney so far (Habuka et al., 2014) and not much information is available on it. But in the merged network this node has 27 links to other data-types (is visualized in Supplementary File 1), mostly to metabolites from both the metabolomics datasets. Based on this, we hypothesize that Tmem72 is not specific to the kidney and that it has some sort of communication function in intestinal mucosa as well. The fact that Tmem72 is a transmembrane protein is supportive for this. Given its observed links with different microbiota, metabolites and cytokines, it might be involved in diverse interactions with other internal phenotypes. Based on such an hypothesis, targeted experimental designs may be developed in order to investigate the hypothesized "communication" function of Tmem72 in intestinal mucosal tissue.

The most highly linked node of the merged network is the bacterial family classification, S24-7, suggesting an important role for this species in gut functionality. In some of the inferred individual correlation networks we already found it to be linked to a high number of nodes. Unfortunately, this node is not represented in literature abstracts together with the here observed neighbors. However, there is compelling literature that shows this microbial classification to be a significant part of the gut microbial community structure (Harris et al., 2014; Jakobsson et al., 2015). This family classification does not

have a good functional definition, yet several studies show that it could be an important player in the functionality of the gut (Evans et al., 2014; Harris et al., 2014; Rooks et al., 2014). The latter claims are in line with the high number of neighbors that S24-7 has in our merged network. The current technical inability to cultivate S24-7 is most certainly due to the absence of knowledge on S24-7 interactions. However, a recent in-silico study (Ormerod et al., 2016) shows that S24-7 species have the ability to survive on different types of carbohydrate sources, similar to the genus *Bifidobacteria*. In the merged network, the connectivity hubs S24-7 and *Bifidobacteria*, share the highest number of neighbors (directly linked nodes). Among them are 16 genes and neither S24-7 nor *Bifidobacteria* have literature results with any of these genes. An enrichment analysis on these shared network gene neighbors shows that they are involved in functions related to linoleic and linolenic acid metabolism (data not shown). It is known that these fatty acids are produced by *Bifidobacteria* (Teran et al., 2015) and that they are involved in the maintenance of the epidermal barrier function (Muñoz-Garcia et al., 2014). The observation that in our network these genes are shared between S24-7 and Bifidobacteria underscores the here hypothesized importance of S24-7 and indicates that these two bacterial groups are indeed closely related in function as hypothesized before (Ormerod et al., 2016).

From the results described in this paper, we conclude that we successfully developed methodologies to identify components in internal phenotypic layers that are connected to components in other internal phenotypic layers. By integrating multi-scale quantitative (~omics) data using a regression approach, we were able to provide provisional insight into potential ways internal phenotypic layers are connected to each other, including those between local and systemic layers. By technical and functional validations, we underscored the relevance of our findings. Based on data generated by this type of integrated approaches, hypothesis driven and targeted research may be developed to identify causal relationships between various biological scales in order to diminish our knowledge gap between genotype and external phenotype. In addition, by expanding comparable approaches by incorporating data on genetic diversity and/or variation in external phenotypes, this knowledge gap may be even further closed down. The analysis pipeline that we developed is very general and can easily be applied to any other type or number of data sets.

## Availability of data and materials

Transcriptomics data has been uploaded into GEO with the accession number GSE84442.

The microbiota data, the two metabolomics datasets and the cytokine data are available on request.

The R scripts using functions from existing R packages are also available on request.

## Author Contributions

NB performed the data analysis and prepared the manuscript. SK performed the animal experiment and contributed significantly to the biological interpretation of the results. VMS contributed to the direction of the analysis and the manuscript. MS was involved in the animal experiment and helped with the direction and critical revision of the manuscript. DS was involved in the data analysis and biological interpretation. MSD helped with the direction of the data analysis and the manuscript. All authors have read and approved of the final manuscript.

## Competing Interests

The authors declare that they have no competing interests.



## Acknowledgment

This work has been financially supported by the Systems Biology Investment Programme of Wageningen University, KB-17-003.02-022.

## References

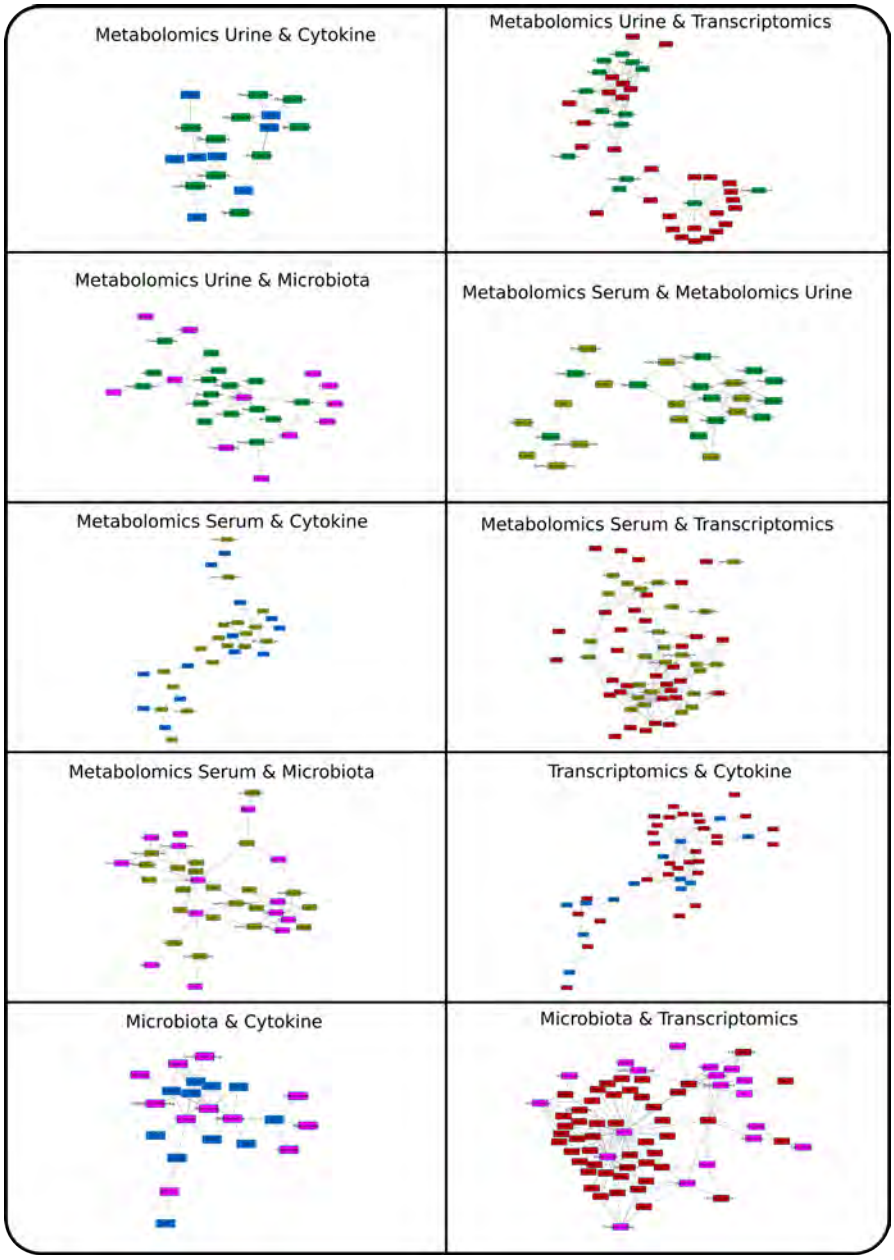

1. Adourian, A., Jennings, E., Balasubramanian, R., Hines, W. M., Damian, D., Plasterer, T. N., et al. (2008). Correlation network analysis for data integration and biomarker selection. *Mol. Biosyst*. 4, 249–259. doi:10.1039/B708489G.
2. Antezana, E., Kuiper, M., and Mironov, V. (2009). Biological knowledge management: the emerging role of the Semantic Web technologies. *Brief. Bioinform.* 10, 392–407. doi:10.1093/bib/bbp024.
3. Barabasi, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5, 101–113. Available at: http://dx.doi.org/10.1038/nrg1272.
4. Benis, N., Schokker, D., Suarez-Diez, M., Martins Dos Santos, V. A. P., Smidt, H., and Smits, M. A. (2015). Network analysis of temporal functionalities of the gut induced by perturbations in new-born piglets. *BMC Genomics* 16, 556. doi:10.1186/s12864-015-1733-8.
5. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B 57*, 289–300. Available at: http://www.jstor.org/stable/2346101.
6. Borate, B. R., Chesler, E. J., Langston, M. A., Saxton, A. M., and Voy, B. H. (2009). Comparison of threshold selection methods for microarray gene co-expression matrices. *BMC Res. Notes 2, 240.*
7. Camacho, D., De La Fuente, A., and Mendes, P. (2005). The origin of correlations in metabolomics data. *Metabolomics 1, 53–63.*
8. Cani, P. D., Bibiloni, R., Knauf, C., Waget, A., Neyrinck, A. M., Delzenne, N. M., et al. (2008). Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Diet–Induced Obesity and Diabetes in Mice. *Diabetes 57*, 1470–1481. Available at: http://diabetes.diabetesjournals.org/content/57/6/1470.abstract.
9. Cicha, I., and Urschel, K. (2015). TNF-$\alpha$ in the cardiovascular system: from physiology to therapy. *Int. J. Interf. Cytokine Mediat. Res.* 7, 9. doi:10.2147/IJICMR.S64894.
10. Civelek, M., and Lusis, A. J. (2014). Systems genetics approaches to understand complex traits. *Nat. Rev. Genet. 15*, 34–48. doi:10.1038/nrg3575.
11. Couto, N., Wood, J., and Barber, J. (2016). The role of glutathione reductase and related enzymes on cellular redox homeostasis network. *Free Radic. Biol. Med. 95*, 27–42. doi:10.1016/j.freeradbiomed.2016.02.028.
12. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Syst*. 1695, 1–9.
13. de Wit, N. J. W., Boekschoten, M. V, Bachmair, E.-M., Hooiveld, G. J. E. J., de Groot, P. J., Rubio-Aliaga, I., et al. (2011). Dose-Dependent Effects of Dietary Fat on Development of Obesity in Relation to Intestinal Differential Gene Expression in C57BL/6J Mice. *PLoS One* 6, e19145. Available at: http://dx.doi.org/10.1371%252Fjournal.pone.0019145.

14. Dejean, S., Gonzalez, I., Lê Cao, K.-A., and Monget, P. (2011). mixOmics: Omics data integration project. *R Packag. version,* 2–9.
15. Diaz-Vivancos, P., de Simone, A., Kiddle, G., and Foyer, C. H. (2015). Glutathione - linking cell proliferation to oxidative stress. *Free Radic. Biol. Med.* 89, 1154–64. doi:10.1016/j.freeradbiomed.2015.09.023.
16. Eguíluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., and Apkarian, A. V. (2005). Scale-Free Brain Functional Networks. *Phys. Rev. Lett.* 94, 18102. Available at: http://link.aps.org/doi/10.1103/PhysRevLett.94.018102.
17. Evans, C. C., LePard, K. J., Kwak, J. W., Stancukas, M. C., Laskowski, S., Dougherty, J., et al. (2014). Exercise Prevents Weight Gain and Alters the Gut Microbiota in a Mouse Model of High Fat Diet-Induced Obesity. *PLoS One* 9, e92193. Available at: http://dx.doi.org/10.1371%252Fjournal.pone.0092193.
18. Fontanesi, L. (2016). Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. *Anim. Front.* 6. doi:10.2527/af.2016-0011.
19. Gentry, J. G., McGlone, J. J., Miller, M. F., and Blanton, J. R. (2004). Environmental effects on pig performance, meat quality, and muscle characteristics1. *J. Anim. Sci.* 82, 209–217. doi:/2004.821209x.
20. González, I., Cao, K.-A. L., Davis, M. J., and Déjean, S. (2012). Visualising associations between paired "omics" data sets. *BioData Min.* 5, 19. doi:10.1186/1756-0381-5-19.
21. Habuka, M., Fagerberg, L., Hallström, B. M., Kampf, C., Edlund, K., Sivertsson, Å., et al. (2014). The Kidney Transcriptome and Proteome Defined by Transcriptomics and Antibody-Based Profiling. *PLoS One* 9, e116125. doi:10.1371/journal.pone.0116125.
22. Harris, J. K., El Kasmi, K. C., Anderson, A. L., Devereaux, M. W., Fillon, S. A., Robertson, C. E., et al. (2014). Specific Microbiome Changes in a Mouse Model of Parenteral Nutrition Associated Liver Injury and Intestinal Inflammation. *PLoS One* 9, e110396. doi:10.1371/journal.pone.0110396.
23. Houle, D., Govindaraju, D. R., and Omholt, S. (2010). Phenomics: the next challenge. *Nat Rev Genet* 11, 855–866. Available at: http://dx.doi.org/10.1038/nrg2897.
24. Iantomasi, T., Favilli, F., Marraccini, P., Magaldi, T., Bruni, P., and Vincenzini, M. T. (1997). Glutathione transport system in human small intestine epithelial cells. *Biochim. Biophys.* Acta 1330, 274–283.
25. Jakobsson, H. E., Rodríguez-Piñeiro, A. M., Schütte, A., Ermund, A., Boysen, P., Bemark, M., et al. (2015). The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep.* 16, 164–177. doi:10.15252/embr.201439263.
26. Lê Cao, K.-A., González, I., and Déjean, S. (2009). integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* 25, 2855–2856.
27. Leuchter, A. F., Hunter, A. M., Krantz, D. E., and Cook, I. A. (2014). Intermediate phenotypes and biomarkers of treatment outcome in major depressive disorder. *Dialogues Clin. Neurosci.* 16, 525–537. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4336921/.

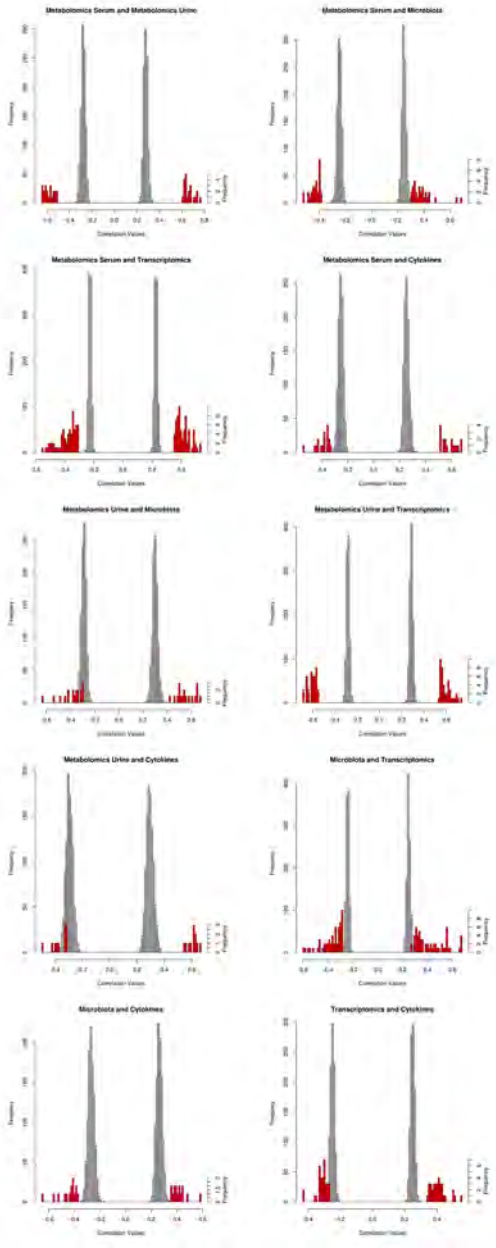
28. Lu, K., Abo, R. P., Schlieper, K. A., Graffam, M. E., Levine, S., Wishnok, J. S., et al. (2014). Arsenic exposure perturbs the gut microbiome and its metabolic profile in mice: An integrated metagenomics and metabolomics analysis. *Environ. Health Perspect.* 122, 284–291. doi:10.1289/ehp.1307429.
29. Montiel-Castro, A. J., González-Cervantes, R. M., Bravo-Ruiseco, G., and Pacheco-López, G. (2013). The microbiota-gut-brain axis: neurobehavioral correlates, health and sociality. *Front. Integr. Neurosci.* 7. doi:10.3389/fnint.2013.00070.
30. Muñoz-Garcia, A., Thomas, C. P., Keeney, D. S., Zheng, Y., and Brash, A. R. (2014). The importance of the lipoxygenase-hepoxilin pathway in the mammalian epidermal barrier. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1841, 401–408. doi:10.1016/j.bbalip.2013.08.020.
31. Ono, K., Muetze, T., Kolishovski, G., Shannon, P., and Demchak, B. (2015). CyREST: Turbocharging Cytoscape Access for External Tools via a RESTful API [version 1; referees: 2 approved]. *F1000Research* 4. doi:10.12688/*f1000research.6767*.1.
32. Ormerod, K. L., Wood, D. L. A., Lachner, N., Gellatly, S. L., Daly, J. N., Parsons, J. D., et al. (2016). Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of homeothermic animals. *Microbiome* 4, 1–17. doi:10.1186/s40168-016-0181-2.
33. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S., and Smyth, G. K. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* 10, 946–963. doi:10.1214/16-AOAS920.
34. Rajasundaram, D., Runavot, J.-L., Guo, X., Willats, W. G. T., Meulewaeter, F., and Selbig, J. (2014). Understanding the Relationship between Cotton Fiber Properties and Non-Cellulosic Cell Wall Polysaccharides. *PLoS One* 9, e112168. Available at: http://dx.doi.org/10.1371%252Fjournal.pone.0112168.
35. Rooks, M. G., Veiga, P., Wardwell-Scott, L. H., Tickle, T., Segata, N., Michaud, M., et al. (2014). Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *ISME J* 8, 1403–1417. Available at: http://dx.doi.org/10.1038/ismej.2014.3.
36. Saccenti, E., Suarez-Diez, M., Luchinat, C., Santucci, C., and Tenori, L. (2015). Probabilistic Networks of Blood Metabolites in Healthy Subjects As Indicators of Latent Cardiovascular Risk. *J. Proteome Res.* 14, 1101–1111. doi:10.1021/pr501075r.
37. Schäfer, J., and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4.
38. Schwartz, S., Friedberg, I., Ivanov, I. V, Davidson, L. A., Goldsby, J. S., Dahl, D. B., et al. (2012). A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. Genome Biol. 13, 1–16. doi:10.1186/gb-2012-13-4-r32.
39. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303.

40. Smyth, G. K. (2005). "limma: Linear Models for Microarray Data," *in Bioinformatics and Computational Biology Solutions Using R and Bioconductor SE - 23* Statistics for Biology and Health., eds. R. Gentleman, V. Carey, W. Huber, R. Irizarry, and S. Dudoit (Springer New York), 397–420. doi:10.1007/0-387-29362-0_23.
41. Steegenga, W. T., Mischke, M., Lute, C., Boekschoten, M. V, Lendvai, A., Pruis, M. G. M., et al. (2016). Maternal exposure to a Western-style diet causes differences in intestinal microbiota composition and gene expression of suckling mouse pups. *Mol. Nutr. Food Res.* doi:10.1002/mnfr.201600141.
42. Teran, V., Luna Pizarro, P., Zacarias, M. F., Vinderola, G., Medina, R., and Van Nieuwenhove, C. (2015). Production of conjugated dienoic and trienoic fatty acids by lactic acid bacteria and bifidobacteria. *J. Funct. Foods* 19, 417–425. doi:10.1016/j.jff.2015.09.046.
43. Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* 73, 273–282.
44. Winter, D. (2016). rentrez: Entrez in R. Available at: https://cran.r-project.org/web/packages/rentrez/rentrez.pdf.
45. Zhu, X., Gerstein, M., and Snyder, M. (2007). Getting connected: analysis and principles of biological networks. *Genes Dev.* 21, 1010–1024. doi:10.1101/gad.1528707.

## Supplemental Information

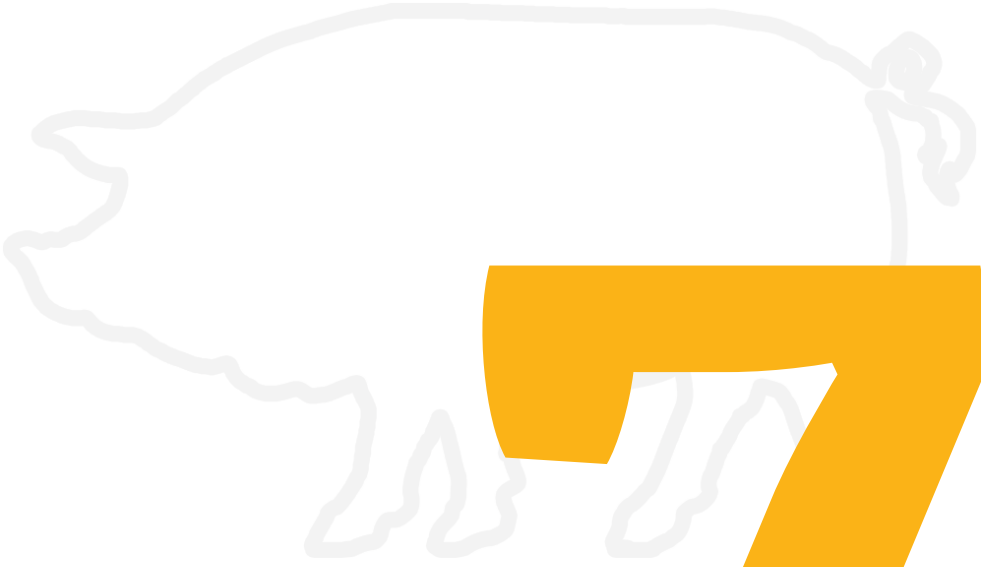



**Supplementary Figure S1. 10 correlation networks.**
The images show the 10 networks from pairwise integration of 5 datasets. The colours indicate different types of data, pink is Microbiota, red is Transcriptomics, blue is Cytokines, and green is Metabolites (light green – Metabolite Serum and dark green – Metabolite Urine).




**Supplementary Figure 2: Distribution of network correlations and random network cut-offs of all 10 networks.**

The x-axis of all these graphs has the range of correlation values, the y- axis the frequency of the correlation values. The grey bars denote the distribution of the cut-offs of the thousand random correlation networks. The red bars are distributions of the correlation values of the real networks. The y-axis on the left has the frequency of the random network cut-offs and the y-axis on the right has the frequency of the real network correlations.



7

# CHAPTER 7

## Functional properties of dietary protein sources towards the gut and systemic metabolism in pigs

*Soumya K. Kar[1,2], Alfons J. M. Jansman[3], Amy C. Harms[4,5,] Leo Kruijt[2,3,] Dirkjan Schokker[2], Mari A. Smits[1,2,6].*

[1]Host Microbe Interactomics Group, Wageningen University & Research, Wageningen, the Netherlands
[2]Wageningen University & Research Animal Breeding and Genomics, Wageningen, The Netherlands
[3]Wageningen Livestock Research, Animal Nutrition, Wageningen University & Research, Wageningen, The Netherlands
[4]Netherlands Metabolomics Centre, Leiden University, Leiden, the Netherlands
[5]Department of Analytical Biosciences, Leiden University, Leiden, the Netherlands
[6]Wageningen Bioveterinary Research, Wageningen University & Research, Wageningen, the Netherlands

*Manuscript submitted for publication*

## Abstract

Changes in various local and systemic physiological parameters in pigs were evaluated as induced by diets containing common and new protein sources. Systemic effects were measured by the concentrations of blood cytokines and chemokines along with amine metabolites. The community structure of the small intestinal microbiome and the transcriptional response of intestinal mucosal tissue (jejunum and ileum) served as parameters for the local response. Growing pigs (BW 35 kg) were fed with experimental diets containing a single, common or new protein sources viz. soybean meal (SBM), black soldier fly larvae (BSF), spray dried blood plasma (SDPP), rapeseed meal (RSM), and wheat gluten meal (WGM) over a period of 4 weeks. The SBM treatment served as reference in the evaluation of the results. Among the nine cytokines measured in serum, no significant differences in their concentrations were observed between treatments. Amine profiles in plasma, however, were distinct for the animals receiving the BSF and SDPP-based diet, compared to other treatments. Microbiota composition in both jejunal and ileal digesta were affected by dietary treatment. Several gene sets were differentially expressed in the transcriptional response of both jejunal and ileal mucosal tissues among some of the dietary treatments, relative to the SBM treatment. We have shown that a "multiomics" approach enables to get detailed information on potential functional properties of dietary proteins sources in pigs in relation to gut health and function.

**Key words: amine metabolites, gut health, metabolomics, microbiota, pigs, transcriptomics.**

## Introduction

Due to global protein shortage arising from the increasing demand for protein sources, the livestock sector faces challenges to provide sufficient food of animal origin to the growing world population while simultaneously reducing its environmental footprint (Boland et al., 2013). Thus, there is a strong need to improve the utilization of current and new protein sources for animal feed (van der Spiegel et al., 2013). Unfortunately, information on the functional properties of (new) protein sources towards their consumer is scarce.

In general practice, pig feeds are formulated based on the provision of ileal or faecal digestible nutrients and derived net energy by feed ingredients. The nutrients relate to proteins/amino acids (AA), starch and sugars, fat, fermentable non-starch-polysaccharides, minerals and vitamins. The former can be characterised as the "strict-nutritional" value of feed ingredients. However, apart from the "strict-nutritional" value, diet and their ingredients and constituents have other "non-strict-nutritional" functional properties in relation to e.g. feed intake (satiety), passage rate through the gastro intestinal tract (GIT), pro- and antimicrobial properties, antioxidative and oxidative effects, immune signalling and metabolic effects (Biesalski et al., 2009; Jansman, 2016). From the context of protein ingredients, we have previously identified potential functional properties of the protein fraction in some common and new protein sources for pigs, using an *in silico* approach (Kar et al., 2016). In Chapters 4 and 5 of this thesis, we have evaluated the effects of several of these protein sources on mice. The current practice of pig diet formulation does generally not take into account such functional properties.

Soybean meal (SBM), which is most frequently used as a protein source for pig feeds, contains a range of proteins, carbohydrates, such as non-starch polysaccharides (NSP) and phytochemicals such as isoflavones, that can influence the activity of the immune system (Karr-Lilienthal et al., 2005). Black soldier fly larvae (BSF), which have been proposed as an alternative source for dietary protein, contain relatively high concentrations of protein and fat (Veldkamp et al., 2012). As a component of a complete diet BSF have been found to support growth performance of pigs (Newton et al., 1977). At present there is lack of knowledge towards possible functional properties of BSF. Spray dried plasma protein (SDPP) is recognized as a high-quality feed ingredient for farm animals, including pigs and has been recommended in animal diet as a source of immunological support due to its high level of globulin proteins, including immunoglobulins (Gao et al., 2011). Rapeseed meal (RSM) contains a well-balanced AAs profile when compared to SBM (Newkirk et al., 2003). Compared to SBM, RSM is a richer source of vitamins such as biotin, niacin, choline, thiamine, vitamin B6 and niacin (NRC, 2012). However, there is scarce information towards possible functional properties of RSM. Wheat gluten meal

(WGM) is also considered as a good source of dietary proteins for many mammals, as it is highly digestible and regarded as an excellent source of glutamine and glutamic acid, amino acids which are known to play a role in the modulation of gut immunity (Ruth and Field, 2013).

In the animal nutrition domain, gut health is an important component of animal health. Gut functions comprise a number of physiological and functional features. It includes, nutrient digestion and absorption, post-absorptive metabolism. Energy generation by volatile fatty acids formed during the intestinal fermentation, maintenance of a stable microbiome, mucus layer development, barrier function and mucosal immune responses are some of the important physiological and functional features (Kogut and Arsenault, 2016). A healthy gut has the capacity to exert all these functional properties in a homeostatic manner allowing the animal to achieve its potential productive performance under a variety of environmental conditions (Jansman, 2016).

It is known that diet and their ingredients and constituents can modulate the composition and diversity of the intestinal microbiome and influence host metabolism and physiology at the systemic as well as the local intestinal level. Systemic effects of diets can be captured in the blood whereas local effects can be recorded within the gut mucosa. Several of the functions that comprise gut health are known to be effected by hydrolysis and fermentation products formed during the passage of dietary proteins through the gut as shown in Figure 1.

When using (new) protein sources in pig feed, it would be beneficial to not only evaluate their strict-nutritional value, but also to consider their functional properties and effects towards the consuming animal. Unfortunately, this type of information is largely missing from literature and partly explains the knowledge gap that exist for evaluation of functional properties of proteins sources. Moreover, there are currently no assessment methods designed to evaluate the variety of functional properties of proteins sources beyond their capacity to provide (essential) AA and other nutrients (Jahan-Mihan et al., 2011).

In the present study, we evaluate changes in local and systemic physiological parameters as induced in pigs by diets containing common and new protein sources. We focussed on effects potentially related to the functional properties of dietary protein. The systemic effects were measured in the blood by the concentrations of cytokines and chemokines along with amine metabolites. The community structure of the small intestinal microbiome and the transcriptional response of mucosal tissue of the small intestine were measured to evaluate the local response. Growing pigs were fed with experimental diets containing a single, common or new protein sources viz. SBM, BSF,

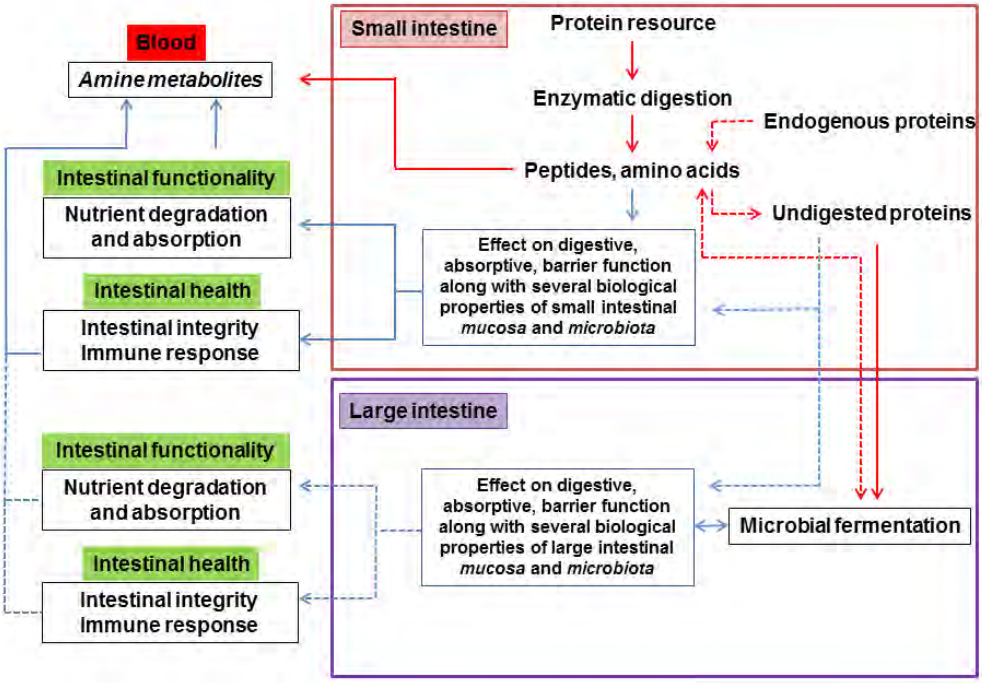



**Figure 1: Overview of bio-processing and bio-functional effects of digested dietary protein in monogastrics.**

Italicised words describe the molecular level parameters. Red arrows are bio-processing steps during protein digestion and blue arrows are the bio-functional effect of the degraded products formed during protein digestion. Continuous arrows are major effect and dotted are minor effects.

SDPP, RSM, WGM. SBM-based diets served as reference to make comparisons with other experimental diets for all analysis, as it is currently the most widely used protein source in pig feed (Gilbert, 2004; Heuzé et al., 2016).

## Materials and Methods

### *Experimental design and housing*

This study was approved by the Animal Care and Use Committee of Wageningen University and Research Center (Wageningen, the Netherlands 2014099.b). A summary of the experimental design is shown in supplementary Figure 1. A total of 40 growing pigs (boars) (Topigs 20 × Tempo from Van Beek, Lelystad, the Netherlands) aged 10 weeks with an average initial body weight of 34.9 ± 3.4 kg on the day (d) of arrival (d 0) were used. Pigs were blocked on litter. Within a block, pigs were randomly allocated to one of the five experimental diets, with eight pigs per experimental diets. Pigs were housed individually in metabolic cages (1.3 x 1.3 m or 2.0 x 1.0 m) with a tender foot floor. The ambient temperature was kept at 24 °C on d 1 and 2, at 23 °C on d 3 and constant at 22 °C from d 4 and onwards. During d 1 to 27, the lights were turned on between 5.30 h and 19.00 h. During d 28 to 29, the lights were turned on between 2.30 h and 19.00 h.

### *Protein sources and experimental diets*

The protein sources evaluated were soybean meal (SBM), wheat gluten meal (WGM), rapeseed meal (RSM) (all commodity batches obtained via Research Diet Services, Wijk bij Duurstede, the Netherlands), spray dried porcine plasma protein (SDPP) (obtained from Darling Ingredients Inc., Irving, TX, USA) and black soldier fly larvae meal (BSF) (obtained from Protix, Dongen, the Netherlands). All five experimental diets were formulated to be iso-proteinaceous (CP, 160 g/kg as-fed basis) and included the respective protein containing ingredients as the only source of protein. Free L-lysine, DL-methionine, L-threonine and L-tryptophan were only included to assure that the dietary concentrations were at a level of at least 70% of the assumed requirement values for these amino acids in growing pigs (CVB, 2007). Titanium dioxide ($TiO_2$) was included in all the diets as an indigestible marker at 2.5 g/kg feed (as-fed basis). The ingredient and calculated nutrient composition of the five experimental diets is presented in supplementary Table 1.

### *Feeding*

During d 1 to 6, pigs were fed a commercial diet and gradually adapted to the experimental diets. From d 7 and onwards, pigs were fed only the experimental diets. The experimental diets were provided in a mash form and mixed with water at a ratio of 1 : 2. Water supply was restricted and only 0.3 l of extra water was provided per animal after each feeding. The feeding level was 2.5 times the maintenance requirement for energy (293 kJ NE/kg $BW^{0.75}$). During d 7 to 26, the daily feed allowance was divided into two equal amounts, fed at 8.00 h and 16.00 h. During d 27 to dissection days i.e. (d 28-29), the feed allowance was divided into six equal portions, fed starting at 5.30 h at intervals of 3 hours.

***Sample collection and dissection procedure***

Blood samples were collected via an ear-vein catheter for both serum and plasma, at d 7 and dissection days (d 28-29) after the morning meal ingestion. For serum collection, blood samples were collected in sterile Vacuette tubes containing Z-serum separator clot activator (Greiner Bio-One B.V., Alphen aan den Rijn, the Netherlands). Tubes were gently inverted and allowed to clot for at least 30 min. All tubes were centrifuged at 2,200 g for 15 min at 20°C and serum was extracted. For plasma, blood samples were collected in sterile Vacuette tubes containing lithium-heparin and immediately centrifuged at 3,000 g for 10 min at 4°C and plasma was extracted. Both serum and plasma were stored at -80°C for further analysis on systemic cytokines and amine metabolites, respectively.

At the dissection days, pigs were anesthetised by injecting pentobarbitone in the ear vein and sacrificed to collect samples. The small intestine was separated from the stomach and the large intestine. Each of jejunum and ileum were divided into three equal segments; proximal, middle and distal. From the middle segments in each tissue, two sub-segment (6-7 cm) were sampled for collection of the intestinal mucosa and its residing microbiota. For collection of the intestinal mucosa, one of the sub-segments was cut open longitudinally along the lumen and washed with sterile normal saline solution. With sterile glass slides mucosal layer was collected, snap frozen in liquid nitrogen and stored at -80ºC for further analysis of genome-wide gene expression profiling. Another sub-segment was also cut open longitudinally along the lumen and luminal digesta was collected using a sterile spatula to perform a community-scale analysis of gut microbiota.

***Chemical analysis of diets***

All chemical analyses were performed according to standard laboratory methods. The experimental feed were analysed for dry matter (DM) (ISO:6496, 1999), ash (ISO:5984, 2002), acid-hydrolysed ether extract (ISO:6492, 1999) and nitrogen by Kjeldahl method (ISO:5983-1, 2005).

***Cytokine and chemokine profiling***

Serum cytokine and chemokine concentrations (pg/ml) were measured in the serum samples collected on d 7 and dissection days using a ProcartaPlex Porcine kit (Affymetrix, eBIOscience, Vienna, Austria) . Calibration curves from recombinant cytokine and chemokine standards were prepared for the 8-point standard dilution set with 4-fold dilution steps in sterile PBS. The samples were measured using a Bio-Plex MagPix Multiplex Reader (Bio-Rad Laboratories Inc. by the Luminex Corporation, The Netherlands). The Bio-Plex Manager software's five-parameter logistic curve fitting (5PL) method was used for raw data analysis and calculation of cytokine concentrations. Nine cytokines and chemokines (IL-1beta, IL-4, IL-6, IL-8, IL-10, IL-12p40, IFN-alpha, IFN-gamma, TNF-alpha) were tested. Concentration level of only cytokines/chemokines over the minimal

detectable limit with at-least six pigs per treatment in each time point are reported. Cytokine concentrations are presented as mean ± standard error of the mean (SEM). Statistical analysis was performed by one-way Analysis of Variance (ANOVA) followed by post hoc testing (Dunnett test: compared each treatment group vs. SBM group as control) using GraphPad prism version 5.03 for Windows Vista (GraphPad Software, San Diego, California, USA). P values < 0.05 were considered statistically significant.

***Amine metabolite profiling and functional analysis***

Blood plasma samples collected on the dissection days from pigs fed diet containing proteins from different sources, were analysed for amine metabolic profiles. Details of biological replicate per treatment are shown in supplementary table 2. The amine profiling was performed as described previously (Noga et al., 2012). Except for one pig belonging to SDPP group, plasma samples were collected from all pigs. Briefly, 5 µL of each sample (unable to procure plasma sample from one pig) was spiked with an internal standard solution (supplementary material 1, available on request). Thiol amines are released from proteins and converted to reduced form using Tris-(2-Carboxyethyl) phosphine (TCEP). Subsequently proteins are precipitated by the addition of MeOH. The supernatant was transferred to an Eppendorf tube (Eppendorf, Germany) and dried in a speedvac (Eppendorf, Germany). The residue was reconstituted in borate buffer (pH 8.5) with AQC reagent (Waters, Etten-Leur, The Netherlands). After reaction, the vials were transferred to an autosampler tray (Waters, Etten-Leur, The Netherlands) and cooled to 10°C prior to injection. For amine metabolite analysis, 1 µL of the reaction mixture was injected into the UPLC-MS/MS system using an Accq-Tag Ultra column (Waters, Etten-Leur, The Netherlands).

We employed an ACQUITY UPLC system with autosampler (Waters, Etten-Leur, The Netherlands). It was coupled with a Xevo Tandem Quadrupole (TQ) mass spectrometer (Waters, Etten-Leur, The Netherlands) which was operated using QuanLynx data acquisition software (version 4.1; Waters, Etten-Leur, The Netherlands). The Xevo TQ was used in the positive-ion electrospray mode and all analytes were monitored in multiple reaction monitoring (MRM) using nominal mass resolution.

Acquired data were evaluated using TargetLynx software (Waters, Etten-Leur, The Netherlands), by integration of assigned MRM peaks and normalization using proper internal standards. For analysis of amino acids, their $^{13}C^{15}N$-labeled analogues were used. For other amines, the closest-eluting internal standard was employed (supplementary material 1). Blank samples were used to correct for background and in-house developed algorithms were applied using the pooled quality check (QC) samples to compensate for shifts in the sensitivity of the mass spectrometer over the batch analysis. Out of 76 targeted amine metabolites, we could detect 58 amines that comply with the acceptance

criteria of QC corrections (van der Kloet et al., 2009). The data are represented as relative response ratios (amine target area/area of internal standard; unit free) of these metabolites (after QC) and are available in the supplementary material 1.

To analyse the amine metabolomics data, the MetaboAnalyst 3.0 suite was employed (Xia and Wishart, 2016). Briefly, the amine metabolic profiles of individual pigs were normalized to a pooled sample of the SBM group and subsequently log transformation was carried out. With the log transformed data exploratory (multivariate) statistical analysis and functional (pathway) analysis were performed. Within the exploratory statistical analysis module of MetaboAnalyst 3.0, principal component analysis (PCA) was performed to get more insight into the variability of the data. The pathway analysis module of MetaboAnalyst 3.0 was employed to determine the amine metabolism that were affected by the treatments compared to the SBM diet. We used the amine metabolic profile of individual pigs, to prepare the data matrix for comparing each treatment with SBM. All the compound names of the metabolites were matched with the human metabolome database (Wishart et al., 2013). Normalization was performed as described for exploratory statistical analysis. Thereafter, a human pathway library (Xia and Wishart, 2016) was selected and all compounds in the selected pathways were used as a reference metabolome because a pig pathway library is not available. The analysis includes pathway enrichment analysis and topological analysis. The impact-value threshold calculated from topology analysis was set at 0.4 and from pathway enrichment the –log (p) value calculated and was set to 5, in order to identify the most related metabolic pathway.

***Microbiota profiling***

Luminal digesta from jejunum and ileum were collected. Details of biological replicates per treatment is shown in supplementary table 2. The biological replicates range from 5 to 8 animals per treatment group as enough luminal digesta were not available for all samplings. From individual samples, microbial DNA was extracted by using the protocol described previously (Schokker et al., 2015). Briefly, luminal digesta of each small intestinal location were mixed 1:1 with PBS and vortexed, spun for 5 min (300 g) at 4 °C. DNA was isolated from snap frozen luminal digesta and the bacterial 16S rDNA V3 region was sequenced by targeted-amplicon 16S sequencing on a Illumina Mi-Seq. Sequenced data were analysed for taxonomy profile per sample with clustering by profile using QIIME (Caporaso et al., 2010).

PCA was performed with relative abundances of microbiota at the genus level in order to get more insight into the microbiota variability. Furthermore, statistical significance testing for "taxon-treatment" association, based on relative abundance of microbiota, was calculated for the genus level by performing the following statistics. Of each sample the

relative abundance of microbiota at genus level was compared to SBM for all treatments. Following order of significance testing was employed. Levene's test is performed to assess the equality of variances among the compared groups. If Levene's test is of equal variance, then the ANOVA is allowed and when ANOVA was significant, Tukey test was used. If Levene's test is not of equal variance, the Kruskal test was followed by the Wilcox test or Welch-based testing was used. The $P < 0.05$ is considered significant.

***Transcriptome profiling***

From individual tissue samples total RNA was extracted. Labelling, hybridization, scanning and feature extraction as well as QC by statistical analysis, was performed as described previously (Schokker et al., 2014). Briefly, total RNA from individual samples from each intestinal tissue were extracted using TRizol reagent (Life Technologies) as recommended by the manufacturer. Homogenised tissue samples were dissolved in 5ml of TRizol reagent. After centrifugation the supernatant was transferred to a fresh tube. Subsequently a phase separation with chloroform was performed as described by the manufacturer. The RNA was precipitated and dissolved and quantified by absorbance measurements at 260 nm. Quality check of the RNA samples was performed with the Agilent Bioanalyzer (Agilent Technologies Netherlands B.V., Amstelveen, The Netherlands). Details of biological replicates per treatment are shown in the supplementary table 2. A few RNA samples from animals belonging to SDPP and BSF group failed to pass the QC performed with the Agilent Bioanalyzer. Labelling, hybridization and washing was done as recommended by Agilent Technologies using the One-Color Microarray Based Gene Expression Analysis Low input Quick Amp Labelling. The input for labelling was 10 ng of total RNA and 600 ng of labelled cRNA was used for hybridization on an 8 pack array. Hybridization was done in the G2545A hybridization oven (Agilent Technologies Netherlands B.V., Amstelveen, The Netherlands) at 65°C with rotation speed 10 rpm for 17 hours, after which the arrays were washed. The arrays where scanned using the DNA microarray scanner with Surescan high resolution Technology (Agilent Technologies Netherlands B.V., Amstelveen, The Netherlands), with resolution of 5µm, 16 bits and PMT of 100%. Feature extraction was performed using Agilent protocol 10.7.3.1 (v10.7) for one colour gene expression.

Gene Set Enrichment Analysis (GSEA; Subramanian et al., 2005) was performed separately for jejunum and ileum samples. We loaded the normalized intensity values of all annotated genes per treatment and the gene expression data of each treatment was compared to the data of SBM. The following settings were different from the default settings: permutations were performed on the gene set and the chip platform was set to gene symbol. Gene Ontology related gene sets of biological processes along with KEGG pathway related gene sets databases (v5.1) were loaded for analysis. Significant GSEA results (FDR < 0.05) were visualized with InteractiVenn (Heberle et al., 2015).

## Results

All pigs remained healthy throughout the whole experimental period. The body weights did not differ significantly ($P < 0.05$) on replacing SBM with BSF, SDPP, RSM and WGM in pig diets (supplementary Figure 2).

### *Dietary nutrient composition*

The WGM-, SDPP- and BSF-based diets contained higher contents of starch than the SBM- and RSM-based diet. The RSM- and BSF-based diets recorded higher amounts of fat than the SBM-, SDPP- and WGM-based diet (Table 1).

**Table 1: Analysed nutrient composition of the experimental diets for pigs, as fed basis**

| | Diets | | | | |
|---|---|---|---|---|---|
| **Composition, g/kg** | **BSF** | **SDPP** | **RSM** | **WGM** | **SBM** |
| Dry matter | 919 | 912 | 911 | 904 | 904 |
| Crude protein | 158 | 158 | 163 | 156 | 166 |
| Sugar | 150 | 162 | 199 | 152 | 175 |
| Starch | 385 | 431 | 236 | 453 | 296 |
| Fat | 45 | 20 | 48 | 19 | 31 |
| Ash | 48 | 57 | 52 | 45 | 53 |

### *Cytokine and chemokine analysis*

Out of nine tested cytokines and chemokines, three (i.e. INF-$\alpha$, IL-12p40 and IL-6) were consistently detected over the minimal detectable limits in serum of six pigs per treatment at each time point. No significant ($P < 0.05$) differences in the concentration levels of INF-$\alpha$, IL-12p40 and IL-6 were observed between the experimental diets as compared to the SBM-based diet (supplementary Figure 3).

### *Amine metabolomic analysis*

To evaluate the metabolic response to SBM substitution with BSF, SDPP, RSM and WGM in the pig diet, a targeted approach was employed (supplementary material 1). Principal Component Analysis (PCA) was performed to get more insight into the variability of the metabolite data. To this end, only the first and second principal components were taken into account. The variance explained by the first two axes was 43.3%. Data clustering occurred for similar treatments (Figure 2). On comparing the amine profiles of the plasma samples collected at the dissection day, the profiles induced by the BSF- and SDPP-based diets formed separate clusters in the PCA without any overlap at the 95% confidence regions. The SBM-, RSM- and WGM-based diets showed a tendency to form separate clusters, but an overlap with the 95% confidence regions among these three plant source-based diets was observed (for detail see Figure 2).

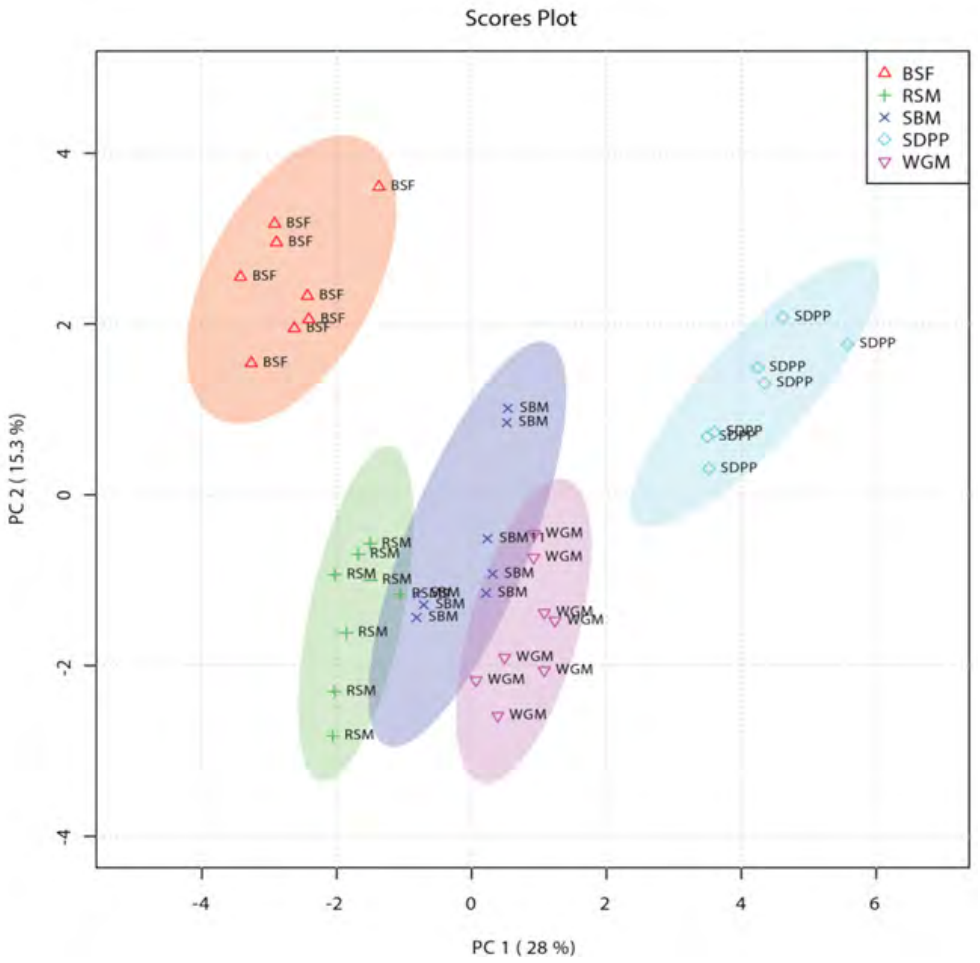



**Figure 2: PCA score plot based on the amine profile in the plasma samples collected from the dissection days from pigs fed diets containing proteins from different source.**

Colored spherical area displays 95% confidence region of respective experimental diets. Each dot represent an animal.

By employing MetaboAnalyst 3.0, metabolic pathway analysis was performed to find out whether whole metabolic pathways were regulated, instead of single metabolic conversions due to the varying concentrations of identified amines in plasma (Figure 3). Four comparisons were made: i.e. BSF vs SBM, SDPP vs SBM, RSM vs SBM and WGM vs SBM to investigate the effect of the protein ingredients from different source on amine metabolic pathways. For BSF vs SBM and SDPP vs SBM three metabolic pathways were affected: 1) glycine, serine and threonine metabolism; 2) arginine and proline metabolism; and 3) aspartate and glutamate metabolism. For WGM vs SBM the pathways 1 and 2 were affected and for RSM vs SBM only pathway 1. The glycine, serine and threonine metabolic pathway was the only pathway that was affected in all

four comparisons. As much as nine of the forty-eight detected amine metabolites form part of the serine and threonine metabolic pathway (supplementary Figures 4A-D). The nine enriched metabolites are: L-aspartic acid, L-homoserine, L-threonine, glycine, L-serine, L-tryptophan, L-cystathionine, cysteine and sarcosine. When comparing BSF and SDPP with the SBM diet, the same amine metabolites were mapped with varying levels of concentration. However, sarcosine was significantly (P <0.0001) high in BSF and low in SDPP (supplementary Figures 4A, B). On comparing RSM with the SBM diet, L-tryptophan was significantly (P <0.0001) lower in RSM (supplementary Figure 4C). For the WGM vs SBM comparison, the nine same amine metabolites were mapped with varying levels of concentration, however, none of the affected metabolites significantly differ in WGM from SBM (supplementary Figure 4D).

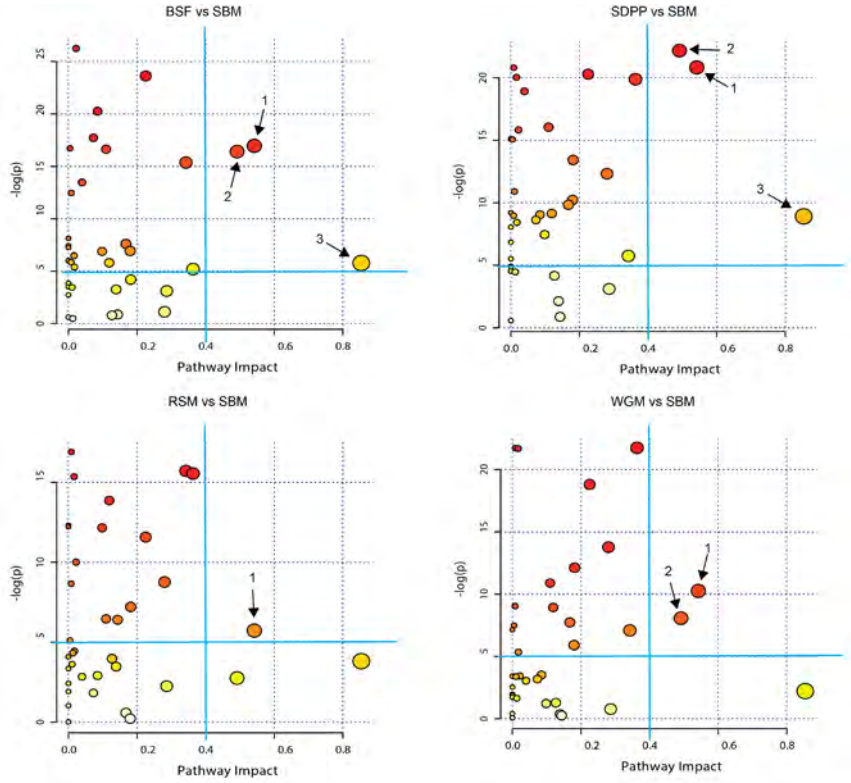



**Figure 3: Metabolic pathway analysis based on the amine profile in the plasma samples collected from the dissection days from pigs fed diets containing proteins from different sources.**

Scores from enrichment analysis is represented on the 'y axis' and from topology analysis on the 'x axis'. The size and the color (on gradient scale: white to dark orange) of the nodes are all matched pathways according to P values from pathway enrichment analysis and pathway impact values from pathway topology analysis. The sky blue lines denotes the thresholds in both axis to identify the most significant matched pathways for all the dietary comparisons. Here, 1 is Glycine, serine and threonine metabolism; 2 is Arginine and proline metabolism; 3 is Alanine, aspartate and glutamate metabolism.

### *Microbiota analysis*

The bacterial 16S rRNA V3 region was sequenced by Illumina Mi-Seq from digesta samples of jejunum and ileum to evaluate the microbiota response to SBM substitution with BSF, SDPP, RSM and WGM in pig diet (supplementary material 2; available on request). Similar to the metabolic analysis, PCA was performed to get more insight into the variability in the microbiota data. To this end, only the first and second principal components were taken into account because they were found to explain the largest part of the variation in the data. The variance explained by the first two axes was 38.9% for jejunum and 48.9% for ileum samples. Data clustering occurred for similar treatments (Figure 4). Compared to the other experimental diets, the BSF diet formed clear separate clusters for both the jejunal and ileal samples without any overlap at the 95% confidence regions (Figure 4).

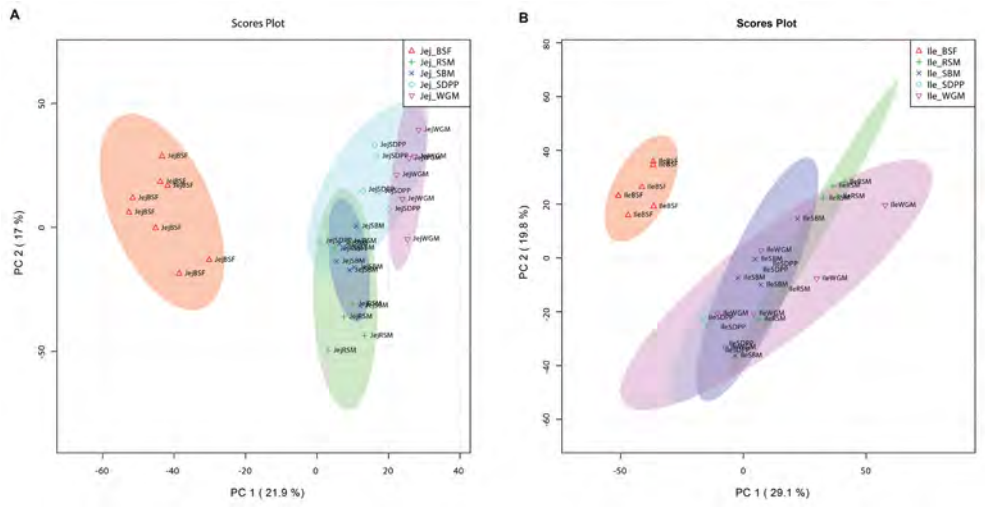



**Figure 4: PCA score plot based on microbiota composition of genus/species level in A) jejunum and B) ileum of pigs fed diets containing protein from different sources.**

Colored spherical area displays 95% confidence regions of respective experimental diets. Each dot represent an animal.

In jejunum, the 95% confidence clusters formed by the RSM diet completely overlapped with the cluster formed by the SBM diet and partially with the SDPP diet. However, the clusters formed by both the SBM and RSM were completely separated from the BSF and WGM diet. Among the two sample location, overall microbial diversity varied mostly among the dietary groups in the ileum, as shown by the results of the various diversity analyses (supplementary Figure 5). The patterns of the Shannon, phylogenic diversity (PD) whole tree and Chao indexes also agreed with the diversity shown by the OTUs in alpha diversity. Chao1, Shannon and PD whole tree indexes results showed that within the ileum, BSF induces significantly higher and SDPP induces significantly lower microbial diversity compared to SBM. The Shannon index recorded only a significant difference in

the microbial diversity in the comparisons for BSF vs SBM within each location. As shown in Figure 5, different phyla were found in the intestinal content samples. *Firmicutes, Proteobacteria,* unassigned group and *Actinobacteria* were the top four most abundant bacterial groups. The separated PCA clustering for both jejunum and ileum samples is due to the difference in the composition and relative abundance of microbes belonging to the above mentioned top four most abundant bacterial group. (Figure 5). In order to further investigate the changes in microbial groups between the experimental diets, genus-level phylogenetic groups were identified for which the relative abundance significantly differed between the dietary treatments. For this, a univariate analysis was used, focusing on all experimental diets vs SBM in both jejunal and ileal samples (Table 2). This analysis revealed that 11 microbial groups differed significantly when comparing BSF to SBM samples in jejunum. In WGM and in SDPP as well as RSM three and two microbial groups differed significantly, respectively, when comparing samples of these diets with samples of SBM In jejunum compared to SBM group, three microbial groups differed significantly in WGM and SDPP groups. Further in jejunum, two microbial groups differed significantly in RSM group when compared to SBM. In ileum, 12 microbial groups differed significantly when BSF group was compared to SBM group (for detail see supplementary Table 3).

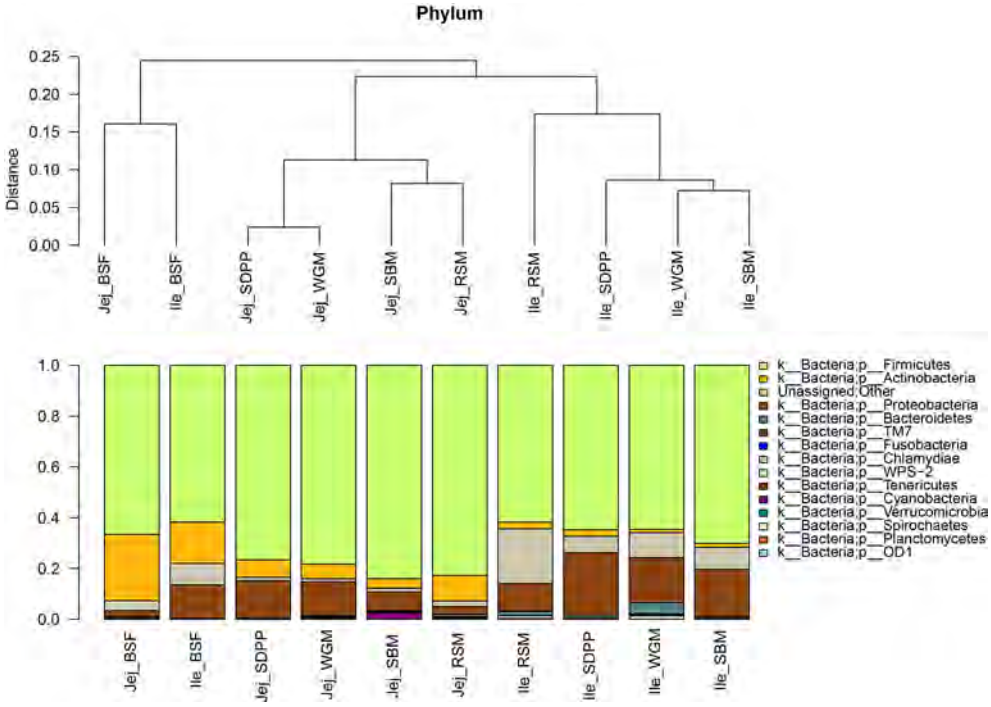



**Figure 5: Hierarchical clustering based on microbiota composition at phylum level in jejunum and ileum of pigs fed diets containing protein from different sources.**

**Table 2: Number of differential enriched gene-sets (FDR <0.1) and core enriched genes expression within the significant enriched gene-sets in the jejunal and ileal mucosa of pigs fed diets with different protein sources compared to a diet with SBM.**

| Tissue | Diet with different protein sources | Number of Gene-sets up-regulated | Number of Gene-sets down-regulated | Number of core enriched genes up-regulated* | Number of core enriched genes down-regulated |
|---|---|---|---|---|---|
| Jejunum | BSF | 1 | 11 | 6 | 118 |
| | SDPP | 0 | 36 | 0 | 340 |
| | RSM | 3 | 0 | 12 | 0 |
| | WGM | 8 | 0 | 99 | 0 |
| Jejunum | BSF | 2 | 8 | 31 | 55 |
| | SDPP | 0 | 0 | 0 | 0 |
| | RSM | 0 | 0 | 0 | 0 |
| | WGM | 7 | 22 | 50 | 152 |

***Transcriptome analysis***

Similar to amine metabolic and microbiota analysis, PCA was performed to get more insight into the variability in the gene expression data. Only the first and second principal components were taken into account for mucosal samples taken from jejunum and ileum. The variance explained by the first two axes was 39.7% for jejunum and 31.7% for ileum. Data clustering occurred for similar treatments, however, an overlap at the 95% confidence regions of each clusters with all experimental diets was observed for both jejunal and ileal tissue (Figure 6). The data is available in the Gene Expression Omnibus from NCBI with the accession number GSE98261.

In-depth analysis was carried out by employing Gene Set Enrichment Analysis (GSEA) using all 'expressed' and annotated genes. In this approach, no pre-filtering of genes is performed. All probes/genes are used as input to GSEA. Four comparisons, i.e. BSF vs SBM, SDPP vs SBM, RSM vs SBM and WGM vs SBM, were made to investigate the effect of the dietary protein sources on jejunum and ileum mucosal gene expression. It resulted in the identification of multiple significant (FDR <0.1) differential enriched gene-sets. Further analysis revealed multiple core-enriched genes (Subramanian et al., 2005) ranging from 6 to 340 which are involved in the significant differential enriched gene-sets (Table 2). Details of the biological processes and pathways of these enriched gene sets are summarised in supplementary Table 4 and 5.

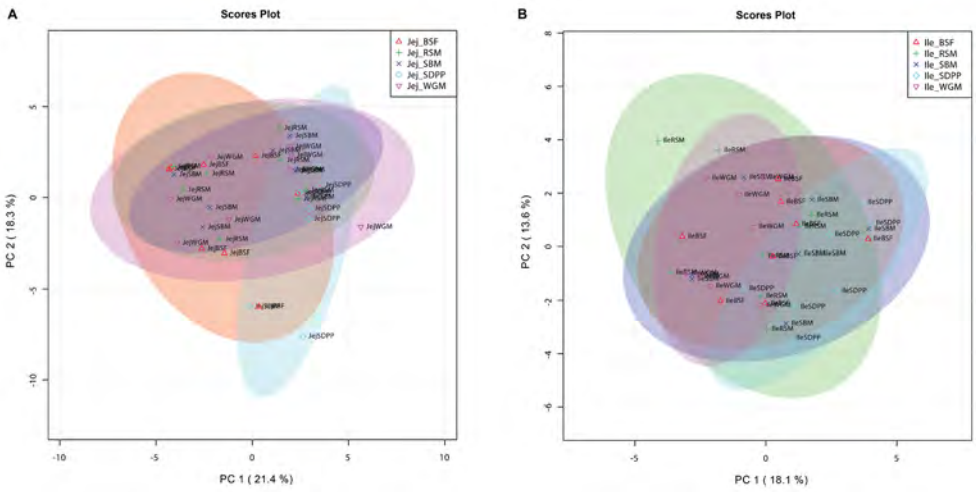



**Figure 6: PCA score plot based on genome-wide transcriptomic response measured by microarray.A) jejunum and B) ileum in pigs fed diets containing protein from different sources.** Colored spherical area displays 95% confidence region of respective experimental diets. Each dot represents an animal.

## Discussion

In this study we have shown that diets prepared with different protein sources, differ in their ability to modulate the physiology of pigs at the local as well as the systemic level. We have shown that different protein sources elicit specific responses in blood amine metabolite profiles and that they have an effect on the composition and diversity of the intestinal microbiota as well as on the intestinal mucosal gene expression.

The experiment was designed so that all five dietary groups were exposed to the same environmental conditions, except for the protein source in the diets. The observed variability in the nutrient composition of the diets is related to the differences in nutrient composition of the protein sources, as they were included at different levels in order to make the experimental diets iso-proteinaceous (CP, 160 g/kg as-fed basis; supplementary Table 1). To investigate the possible antigenic properties of the dietary protein sources, we measured the concentration of nine systemic cytokines and chemokines in serum. Cytokines are short-lived and their synthesis represents an accurate "snapshot" of immune activity. Since no significant differences in cytokine levels were measured between the experimental diets we conclude that in pigs, BSF, SDPP-, RSM- and WGM-based diets evoke (possible) an immunogenic response which is not different from the response towards an SBM-based diet, fed for equal time periods.

Based on the current knowledge on the process of protein digestion in monogastrics, Figure 1 visualizes the factual relationships and interactions of multiple biological events that occur during protein hydrolysis, absorption and metabolism in the GIT. The total fraction of blood amine metabolites contains products from nitrogen metabolism in organs and tissues, including the tissue of the digestive tract, but also a large fraction of AAs derived of the dietary protein source. In addition, metabolism of endogenous protein in the gut (e.g. from desquamated mucosal cells and in the form of digestive enzymes), lysed microbial cells and metabolites originating from microbial metabolism in the gut contribute to the systemic amine metabolite pool. Therefore, measuring amine metabolites in serum or urine is considered to be a useful approach to assess the host and microbiota metabolism of dietary proteins. A profile of blood amine metabolites, thus, on one hand could explain the efficiency of nutrient utilization from the diet and on the other hand explains the absorptive capacity and functioning of the digestive tract in livestock, which also relates to gut-health (Kogut and Arsenault, 2016). The observed plasma amine metabolite profiles were specific for each of the experimental diets based on the various protein sources. We ascribe this result to the differences in AA composition of the protein ingredients fed to the pigs (Hoffman and Falvo, 2004; Kar et al., 2016).

The amine metabolic pathway analysis represents an accurate "snapshot" of the amine metabolic status of the pigs. Result showed that the glycine, serine and threonine metabolic pathway was induced by BSF, SDPP, RSM and WGM-based diets in comparisons to the SBM-based diet. Nine metabolites of this pathway reflect the diet-specific functioning of the epithelial AA transport system present in both the apical and basolateral part of the intestinal epithelium (Broer, 2008). Among the metabolites, sarcosine, an AA derivative that is primarily found in muscle, including insect tissue, was found in significantly higher serum concentrations after feeding the BSF-based diet compared to the SBM-based diet. Further, sarcosine is found naturally as an intermediate in the metabolism of choline to glycine (ChEBI-Database, 2016). BSF is reported to be high in choline (Finke, 2013) that might contribute to the high serum levels of sarcosine as this is formed from dietary intake of choline. SDPP (Henn et al., 2013) has almost five times less choline than SBM (NRC, 2012) and this might have resulted in significantly low levels of sarcosine in SDPP fed pigs compared to SBM fed pigs. As such, sarcosine has no known toxicity (ChEBI-Database, 2016; Pubchem-Database, 2017). Analysis of the amine profiles in the plasma of pigs fed diets with different protein sources could not detect any potential toxic amine that could compromise health status of the pig.

We observed a clear discrepancy in the effect of insect as protein sources on the small intestinal microbiome in pigs compared to the microbial community structure induced by plant and animal based protein sources. Such changes in microbial community structure may have consequences not only for protein metabolism, but also for homeostasis of

intestinal bacteria and in balancing inflammatory and tolerance immune mechanisms of gut mucosal tissues (van Baarlen et al., 2013). In our study, the change in diversity indexes of Shannon, PD whole tree and Chao indicated that the richness and evenness of microbiota changes with the change in dietary protein source (supplementary Figure 5). This is consistent with previous findings which report that dietary protein sources induce significant disparity regarding the bacterial communities of the small intestine in pigs (Yuan et al.; Zhao et al., 2015). Compared to colonic microbiota, the small intestine is colonized by a simpler microbial community structure, both in number and diversity (van Baarlen et al., 2013; Hugenholtz, 2015). As shown in Figure 5, *Firmicutes, Proteobacteria and Actinobacteria* dominated the jejunum and ileum which is in agreement with other studies (Ley et al., 2008; Zhao et al., 2015; Cao et al., 2016). The proportion of each phylum was fluctuant between each dietary group in both small intestinal segments. Members of the phyla *Actinobacteria* and *Firmicutes* were highly abundant in the BSF group compared to the SBM group and both are indicative of a normal intestinal microbiome in pigs (Ley et al., 2008; Zhao et al., 2015). Moreover, the BSF-based diet clearly resulted in a greater microbial diversity in comparison to all other diets. A high microbial diversity in the lumen of the gut is, in general, regarded to affect gut health in a positive manner (Ley et al., 2006; Turnbaugh et al., 2008; Lozupone et al., 2012; van Baarlen et al., 2013). Several human clinical studies have shown that diseased subjects have a lower microbial diversity than their healthy controls (Giongo et al., 2011). The microbial diversity analysis further revealed that SDPP, RSM and WGM maintained a bacterial diversity that is comparable to the bacterial diversity in pigs fed the SBM-based diet. Although we observed some differences in composition and diversity of the gut microbiota between dietary treatments, our results suggest that replacing SBM as protein source in pig diet with BSF, SDPP, RSM and WGM results in the development of an intestinal microbiota community structure that is regarded as "normal" and which is not associated with higher risks for disease.

Unlike the results from the amine and microbiota analysis, we could not see clear effects of the dietary protein sources on the small intestinal mucosal gene expression response in pigs as examined for differential expression of individual genes analyses by LIMMA approaches (Ritchie et al., 2015). This suggests that only subtle effects on the intestinal mucosal gene expression are induced in pigs by diets containing protein from different sources. This is in agreement with a number of other transcriptomic studies on nutritional effects where extreme dietary manipulations were avoided (Oster et al., 2012; Óvilo et al., 2014; Benítez et al., 2016). However, differences in biological mechanisms cannot be explained by the activity of single genes alone, they are often attributed to groups of genes (Mootha et al., 2003) based on their annotation. These gene-sets (Ashburner et al., 2000) or pathways (Kanehisa and Goto, 2000) are representations of basic coherent mechanisms in cells, tissues, or the organism that are used to achieve a particular physiological response. Examination of the gene expression data using the

GSEA approach indicated that the intestinal mucosal tissue responded in a diet-specific manner. The biological processes and pathways that differed between the diets are related to gut function and health attributed by cellular differentiation, development, maintaining homeostatic functions, metabolism and immunity (supplementary Tables 4 and 5). Compared to SBM diet, these physiological responses were upregulated in pigs fed diet with BSF, SDPP, RSM, or WGM. In a previous study in mice (Chapter 4), a SBM-based diet also deviated strongly from diets, based on other protein sources in affecting mucosal gene expression. In mice, however, the SBM-based diet deviated strongly from the other diets in the expression of mTOR signalling and downstream response genes. Such responses could not be detected here in pigs and we cannot provide a good explanation for this.

Due to the mammalian digestive physiology, it is obvious that the presence of both strict-nutritional- and non-strict-nutritional components derived from the diet during the digestion process will be in abundance in jejunum compared to ileum. However, we observed that the ileal mucosa still showed gene expression responses for several of the experimental diets. This suggests that dietary protein source may influence the timing of the release and absorption of dietary components by the GIT. This is in agreement with previous findings (Liu and Selle, 2015). If true, these observations point towards a differential kinetics of protein digestion of diets formulated with different protein sources.

In this study we employed state-of-the-art ~omics based technologies to evaluate the impact of SBM substitution with BSF, SDPP, RSM and WGM in pig diets on molecular parameters in relation to systemic and gut-related parameters. We have shown that a "multi-omics" approach enable to get detailed information on the evaluation of the functional properties of proteins source. Further knowledge development in this field may enable the formulation of animal diet with protein sources taking into account their functional properties.

## Acknowledgments

We thank all animal care taker in Experimental animal facility, Wageningen University and Research, Lelystad, The Netherlands for helping us with successful execution of this animal experiment. Additionally we want to thank Frank Harders from Wageningen Bioveterinary Research, Wageningen University & Research, Wageningen, the Netherlands for expertise and help with the Illumina Mi-Seq sequencing. The authors acknowledge the financial support from the Wageningen UR "IPOP Customized Nutrition" programme financed by Wageningen UR, the Dutch Ministry of Economic Affairs, Agriculture & Innovation, the graduate school (WIAS) and industrial partners Nutreco, The Netherlands and Darling Ingredients Internationals, The Netherlands.

## Reference


1. Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, and C. Gene Ontology. 2000. Gene Ontology: tool for the unification of biology. Nature Genetics 25: 25-29.
2. Benítez, R., Y. Núñez, A. Fernández, B. Isabel, C. Rodríguez, A. Daza, C. López-Bote, L. Silió, and C. Óvilo. 2016. Adipose tissue transcriptional response of lipid metabolism genes in growing Iberian pigs fed oleic acid v. carbohydrate enriched diet. Animal : an international journal of animal bioscience 10: 939-946. doi:10.1017/s1751731115003055
3. Biesalski, H.-K., L. O. Dragsted, I. Elmadfa, R. Grossklaus, M. Muller, D. Schrenk, P. Walter, and P. Weber. 2009. Bioactive compounds: definition and assessment of activity. Nutrition (Burbank, Los Angeles County, Calif ) 25: 1202-1205. doi:S0899-9007(09)00225-1
4. Boland, M. J., A. N. Rae, J. M. Vereijken, M. P. M. Meuwissen, A. R. H. Fischer, M. van Boekel, S. M. Rutherfurd, H. Gruppen, P. J. Moughan, and W. H. Hendriks. 2013. The future supply of animal-derived protein for human consumption. Trends in Food Science & Technology 29: 62-73. doi:10.1016/j.tifs.2012.07.002
5. Broer, S. 2008. Amino acid transport across mammalian intestinal and renal epithelia. Physiological Reviews 88: 249-286. doi:10.1152/physrev.00018.2006
6. Cao, K. F., H. H. Zhang, H. H. Han, Y. Song, X. L. Bai, and H. Sun. 2016. Effect of dietary protein sources on the small intestine microbiome of weaned piglets based on high-throughput sequencing. Letters in Applied Microbiology 62: 392-398. doi:10.1111/lam.12559
7. Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Tumbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7: 335-336. doi:10.1038/nmeth.f.303
8. ChEBI-Database. 2016. CHEBI:15611 - sarcosine. The European Bioinformatics Institute, United Kingdom.
9. CVB. 2007. Chemical composition and nutritional values of feedstuffs and feeding standards No. CVB series nr. 36. Product Board Animal Feed, The Hague, The Netherlands.
10. Finke, M. D. 2013. Complete Nutrient Content of Four Species of Feeder Insects. Zoo Biology 32: 27-36. doi:10.1002/zoo.21012
11. Gao, Y. Y., Z. Y. Jiang, Y. C. Lin, C. T. Zheng, G. L. Zhou, and F. Chen. 2011. Effects of spray-dried animal plasma on serous and intestinal redox status and cytokines of neonatal piglets. J Anim Sci 89: 150-157. doi:10.2527/jas.2010-2967
12. Gilbert, R. 2004. World Animal Feed Industry Protein sources for the animal feed industry. p 1-8. Agriculture and consumer protection, FAO,UN, Rome.

13. Giongo, A., K. A. Gano, D. B. Crabb, N. Mukherjee, L. L. Novelo, G. Casella, J. C. Drew, J. Ilonen, M. Knip, H. Hyoty, R. Veijola, T. Simell, O. Simell, J. Neu, C. H. Wasserfall, D. Schatz, M. A. Atkinson, and E. W. Triplett. 2011. Toward defining the autoimmune microbiome for type 1 diabetes. The ISME journal 5: 82-91. doi:ismej201092
14. Heberle, H., G. V. Meirelles, F. R. da Silva, G. P. Telles, and R. Minghim. 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. Bmc Bioinformatics 16.
15. Henn, J. D., L. Bockor, M. S. Vieira, A. M. L. Ribeiro, A. M. Kessler, L. Albino, H. Rostagno, J. D. Crenshaw, J. M. Campbell, and L. F. S. Rangel. 2013. Inclusion of porcine spray-dried plasma in broiler diet. Journal of Applied Poultry Research 22: 229-237. doi:10.3382/japr.2012-00613
16. Heuzé, V., G. Tran, and S. Kaushik. 2016. Soybean meal. Feedipedia, a programme by INRA, CIRAD, AFZ and FAO. Accessed Date Accessed.| doi:DOI|
17. Hoffman, J. R., and M. J. Falvo. 2004. Protein – Which is Best? Journal of Sports Science & Medicine 3: 118-130.
18. Hugenholtz, F. 2015. Mouse gut microbiomics of short chain fatty acid metabolism and mucosal responses, Wageningen University and Research, Wageningen, NL.
19. ISO:5983-1. 2005. Animal feeding stuffs—Determination of N content and calculation of crude protein content—Part 1: Kjeldahl method. International Organization for Standardization, Geneva, Switzerland.
20. ISO:5984. 2002. Animal feeding stuffs—Determination of ash content. International Organization for Standardization, Geneva, Switzerland.
21. ISO:6492. 1999. Animal feeding stuffs—Determination of fat content. International Organization for Standardization, Geneva, Switzerland.
22. ISO:6496. 1999. Animal feeding stuffs—Determination of moisture and other volatile matter content. International Organization for Standardization, Geneva, Switzerland.
23. Jahan-Mihan, A., B. L. Luhovyy, D. E. Khoury, and G. H. Anderson. 2011. Dietary Proteins as Determinants of Metabolic and Physiologic Functions of the Gastrointestinal Tract. Nutrients 3: 574-603.
24. Jansman, A. J. M. 2016. Health and functions of the gastrointestinal tract in pigs: Effects of functional ingredients and feed and ingredient processing. J Anim Sci 94: 12-21. doi:10.2527/jas2015-9886
25. Kanehisa, M., and S. Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28: 27-30. doi:gkd027
26. Kar, S. K., A. J. M. Jansman, S. Boeren, L. Kruijt, and M. A. Smits. 2016. Approximation of the amino acid composition and bio-functional properties of current and novel protein sources for pigs. J Anim Sci 94: 30–39. doi:10.2527/jas2015-9677
27. Karr-Lilienthal, L. K., C. T. Kadzere, C. M. Grieshop, and G. C. Fahey. 2005. Chemical and nutritional properties of soybean carbohydrates as related to nonruminants: A review. Livestock Production Science 97: 1-12. doi:10.1016/j.livprodsci.2005.01.015
28. Kogut, M. H., and R. J. Arsenault. 2016. Editorial: Gut Health: The New Paradigm in Food Animal Production. Frontiers in Veterinary Science 3: 71. doi:10.3389/fvets.2016.00071

29. Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, and J. I. Gordon. 2008. Evolution of mammals and their gut microbes. Science 320: 1647-1651. doi:1155725
30. Ley, R. E., P. J. Turnbaugh, S. Klein, and J. I. Gordon. 2006. Microbial ecology: human gut microbes associated with obesity. Nature 444: 1022-1023.
31. Liu, S. Y., and P. H. Selle. 2015. A consideration of starch and protein digestive dynamics in chicken-meat production. Worlds Poultry Science Journal 71: 297-309. doi:10.1017/s0043933915000306
32. Lozupone, C. A., J. I. Stombaugh, J. I. Gordon, J. K. Jansson, and R. Knight. 2012. Diversity, stability and resilience of the human gut microbiota. Nature 489: 220-230.
33. Mootha, V. K., C. M. Lindgren, K. F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstrale, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, and L. C. Groop. 2003. PGC-1 alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics 34: 267-273. doi:10.1038/ng1180
34. Newkirk, R. W., H. L. Classen, and M. J. Edney. 2003. Effects of prepress-solvent extraction on the nutritional value of canola meal for broiler chickens. Animal Feed Science and Technology 104: 111-119. doi:10.1016/s0377-8401(02)00331-0
35. Newton, G. L., C. V. Booram, R. W. Barker, and O. M. Hale. 1977. Dried hermetia-illucens larvae meal as a supplement for swine. J Anim Sci 44: 395-400.
36. Noga, M. J., A. Dane, S. N. Shi, A. Attali, H. van Aken, E. Suidgeest, T. Tuinstra, B. Muilwijk, L. Coulier, T. Luider, T. H. Reijmers, R. J. Vreeken, and T. Hankemeier. 2012. Metabolomics of cerebrospinal fluid reveals changes in the central nervous system metabolism in a rat model of multiple sclerosis. Metabolomics 8: 253-263. doi:10.1007/s11306-011-0306-3
37. NRC. 2012. Nutrient Requirements of Swine. National Academic Press, Washington, D.C. 20418 USA.
38. Oster, M., E. Murani, C. C. Metges, S. Ponsuksili, and K. Wimmers. 2012. A low protein diet during pregnancy provokes a lasting shift of hepatic expression of genes related to cell cycle throughout ontogenesis in a porcine model. BMC Genomics 13: 93. doi:10.1186/1471-2164-13-93
39. Óvilo, C., R. Benítez, A. Fernández, B. Isabel, Y. Núñez, A. I. Fernández, C. Rodríguez, A. Daza, L. Silió, and C. López-Bote. 2014. Dietary energy source largely affects tissue fatty acid composition but has minor influence on gene transcription in Iberian pigs1. J Anim Sci 92: 939-954. doi:10.2527/jas.2013-6988
40. Pubchem-Database. 2017. Compound Summary for CID 1088. National Center for Biotechnology Information, USA.
41. Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43: e47. doi:gkv007

42. Ruth, M. R., and C. J. Field. 2013. The immune modifying effects of amino acids on gut-associated lymphoid tissue. Journal of Animal Science and Biotechnology 4. doi:10.1186/2049-1891-4-27
43. Schokker, D., G. Veninga, S. A. Vastenhouw, A. Bossers, F. M. de Bree, L. M. T. E. Kaal-Lansbergen, J. M. J. Rebel, and M. A. Smits. 2015. Early life microbial colonization of the gut and intestinal development differ between genetically divergent broiler lines. BMC Genomics. doi:DOI 10.1186/s12864-015-1646-6
44. Schokker, D., J. Zhang, L. L. Zhang, S. A. Vastenhouw, H. Heilig, H. Smidt, J. M. J. Rebel, and M. A. Smits. 2014. Early-Life Environmental Variation Affects Intestinal Microbiota and Immune Development in New-Born Piglets. PLoS One 9. doi:10.1371/journal.pone.0100040
45. Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545-15550.
46. Turnbaugh, P. J., F. Backhed, L. Fulton, and J. I. Gordon. 2008. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe 3: 213-223. doi:S1931-3128(08)00089-9
47. van Baarlen, P., J. M. Wells, and M. Kleerebezem. 2013. Regulation of intestinal homeostasis and immunity with probiotic lactobacilli. Trends Immunol 34: 208-215. doi:S1471-4906(13)00014-8
48. van der Kloet, F. M., I. Bobeldijk, E. R. Verheij, and R. H. Jellema. 2009. Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. Journal of Proteome Research 8: 5132-5141. doi:10.1021/pr900499r
49. van der Spiegel, M., M. Y. Noordam, and H. J. van der Fels-Klerx. 2013. Safety of Novel Protein Sources (Insects, Microalgae, Seaweed, Duckweed, and Rapeseed) and Legislative Aspects for Their Application in Food and Feed Production. Comprehensive Reviews in Food Science and Food Safety 12: 662-678. doi:10.1111/1541-4337.12032
50. Veldkamp, T., G. v. Duinkerken, A. v. Huis, C. M. M. lakemond, E. Ottevanger, G. Bosch, and M. A. J. S. v. Boekel. 2012. Insects as a sustainable feed ingredient in pig and poultry diet- a feasibility study. 638, Wageningen University & Research, Wageningen, The Netherlands.
51. Wishart, D. S., T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, and A. Scalbert. 2013. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic acids research 41: D801-807.
52. Xia, J., and D. S. Wishart. 2016. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. Curr Protoc Bioinformatics 55: 14 10 11-14 10 91. doi:10.1002/cpbi.11
53. Zhao, W., Y. Wang, S. Liu, J. Huang, Z. Zhai, C. He, J. Ding, J. Wang, H. Wang, W. Fan, J. Zhao, and H. Meng. 2015. The Dynamic Distribution of Porcine Microbiota across Different Ages and Gastrointestinal Tract Segments. PLoS One 10: e0117441. doi:10.1371/journal.pone.0117441

## Supplemental Information

**Supplementary Table 1: Ingredient and estimated nutrient composition of the experimental diet for pigs, as fed basis**

| Items | Diets | | | | |
|---|---|---|---|---|---|
| | SBM | BSF | SDPP | RSM | WGM |
| Ingredients, g/kg | | | | | |
| Maize starch | 376.2 | 451.0 | 521.4 | 258 | 527.9 |
| Sugar | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Dextrose | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 |
| Cellulose | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 |
| Soybean oil | 43.3 | 6.4 | 32.6 | 30.0 | 17.6 |
| Chalk | 14.4 | 0.0 | 15.7 | 5.1 | 15.9 |
| Mono sodium phosphate | 10.1 | 11.7 | 14.2 | 4.6 | 13.6 |
| NaCl | 4.1 | 0.0 | 0.0 | 4.0 | 2.0 |
| Sodium bicarbonate | 1.4 | 6.7 | 0.0 | 1.2 | 3.9 |
| Calcium carbonate | 0.0 | 5.0 | 11.2 | 0.0 | 10.6 |
| Calcium chloride | 0.0 | 4.8 | 0.0 | 0.0 | 0.0 |
| Premix | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 |
| Titanium di-oxide | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
| L-Lysine HCl | 0.0 | 0.0 | 0.0 | 0.0 | 5.8 |
| DL-Methionine | 0.3 | 1.3 | 1.4 | 0.0 | 0.0 |
| L-Threonine | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 |
| L-Tryptophan | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 |
| Black soldier fly (larvae) | 0.0 | 305.0 | 0.0 | 0.0 | 0.0 |
| Spray dried plasma protein | 0.0 | 0.0 | 196.0 | 0.0 | 0.0 |
| Rape seed meal | 0.0 | 0.0 | 0.0 | 489.6 | 0.0 |
| Wheat gluten meal | 0.0 | 0.0 | 0.0 | 0.0 | 194.7 |
| Soybean meal | 342.7 | 0.0 | 0.0 | 0.0 | 0.0 |
| Total | 1000 | 1000 | 1000 | 1000 | 1000 |
| Composition, g/kg | | | | | |
| Dry matter | 908.5 | 932.5 | 914.8 | 957.1 | 915.1 |
| Ash | 55.1 | 59.0 | 57.4 | 53.3 | 53.0 |
| Crude protein | 160.0 | 160.0 | 160.0 | 160.0 | 160.0 |
| Ether extract | 52.6 | 50.0 | 37.4 | 52.0 | 28.8 |
| Starch | 320.8 | 393.3 | 439.5 | 223.4 | 457.1 |
| Sugar | 195.1 | 157.9 | 157.9 | 203.2 | 163.5 |
| Ca | 8.2 | 11.8 | 8.2 | 7.0 | 8.2 |

| Items | Diets[1] | | | | |
|---|---|---|---|---|---|
| | **SBM** | **BSF** | **SDPP** | **RSM** | **WGM** |
| P | 2.8 | 2.8 | 2.8 | 2.4 | 2.8 |
| K | 7.8 | 6.2 | 6.9 | 6.3 | 6.2 |
| Na | 2.0 | 2 | 4.7 | 2.0 | 2.0 |
| Cl | 2.5 | 2.5 | 7.3 | 2.5 | 2.5 |
| NSP[2] | 130.0 | 115.0 | 64.0 | 228.0 | 55.0 |
| EB[3], Meq/kg | 215.0 | 175.0 | 175.0 | 177.0 | 175.0 |
| Digestible energy, MJ/kg | 10.7 | 10.7 | 10.7 | 9.2 | 10.7 |
| Amino acids | | | | | |
| Lys | 8.7 | 9.0 | 12.5 | 6.2 | 7.0 |
| Met | 2.3 | 4.1 | 2.3 | 2.5 | 2.4 |
| Cys | 1.9 | 0.1 | 4.9 | 2.7 | 3.3 |
| Met+Cys | 4.2 | 4.2 | 7.2 | 5.2 | 5.7 |
| Thr | 5.1 | 5.0 | 7.3 | 4.6 | 4.1 |
| Trp | 1.8 | 1.3 | 2.1 | 1.4 | 1.3 |
| Ile | 6.3 | 6.9 | 3.8 | 4.4 | 5.5 |
| Arg | 11.1 | 7.6 | 8.1 | 8.0 | 5.4 |
| Phe | 7.3 | 7.1 | 7.8 | 4.8 | 7.8 |
| His | 3.8 | 4.1 | 4.6 | 3.5 | 3.2 |
| Leu | 10.5 | 10.8 | 13.3 | 8.3 | 10.5 |
| Tyr | 5.1 | 9.5 | 8.4 | 3.5 | 5.0 |
| Val | 6.4 | 11.2 | 9.1 | 5.5 | 5.9 |

[1]Diets: BSF is black soldier fly, SDPP is spray dried plasma proteins, RSM is rape seed meal, WGM is wheat gluten meal and SBM is soybean meal.

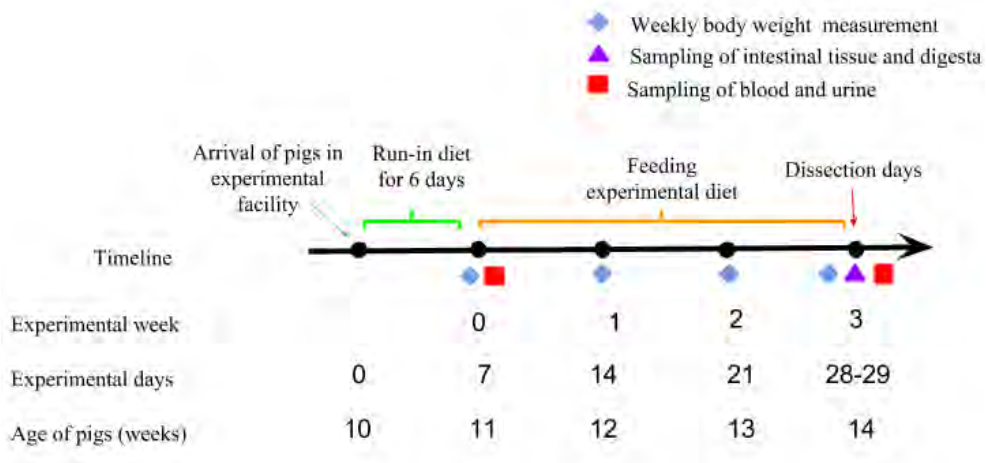
[2]NSP: non-starch polysaccharides

[3]EB: Electrolyte balance

**Supplementary Table 2: An overview of number of samples from animals per experimental dietary group used for analyzing amine profile, small intestinal microbiota and mucosa in this study.**

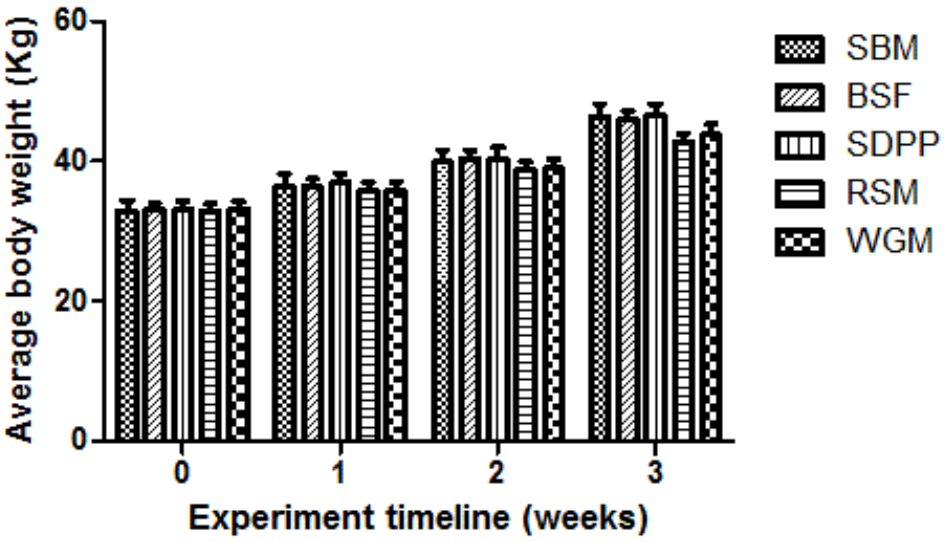
| Experimental diets | Plasma metabolites | Microbiota jejunum | Microbiota ileum | Transcriptomics jejunum | Transcriptomics ileum |
|---|---|---|---|---|---|
| SBM | 8 | 7 | 5 | 8 | 8 |
| BSF | 8 | 8 | 6 | 8 | 8 |
| SDPP | 7 | 6 | 6 | 6 | 7 |
| RSM | 8 | 6 | 6 | 8 | 8 |
| WGM | 8 | 6 | 6 | 8 | 8 |

Number of samples per experimental dietary group < 8 is due to either non-availability of the sample or poor quality for DNA and RNA from the collected samples, which are not suitable to perform sequencing or transcriptomics analysis.



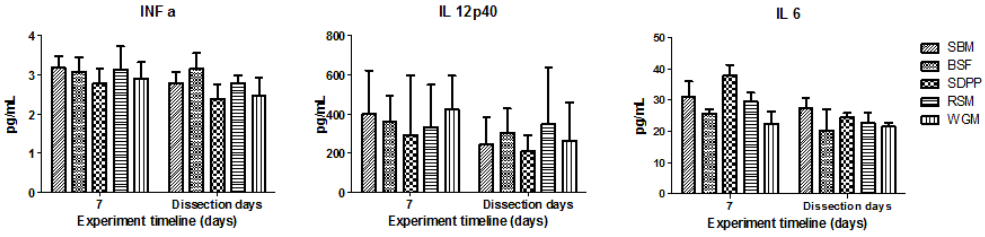

**Supplementary Figure 1: Design of experiment.**

The solid black dot in the timeline represents the corresponding experimental week/age of the pigs (weeks). Blood was collected for analysis of systemic immune signaling molecules and amine metabolites. Intestinal tissue and its digesta were used for transcriptome and microbiota analysis.



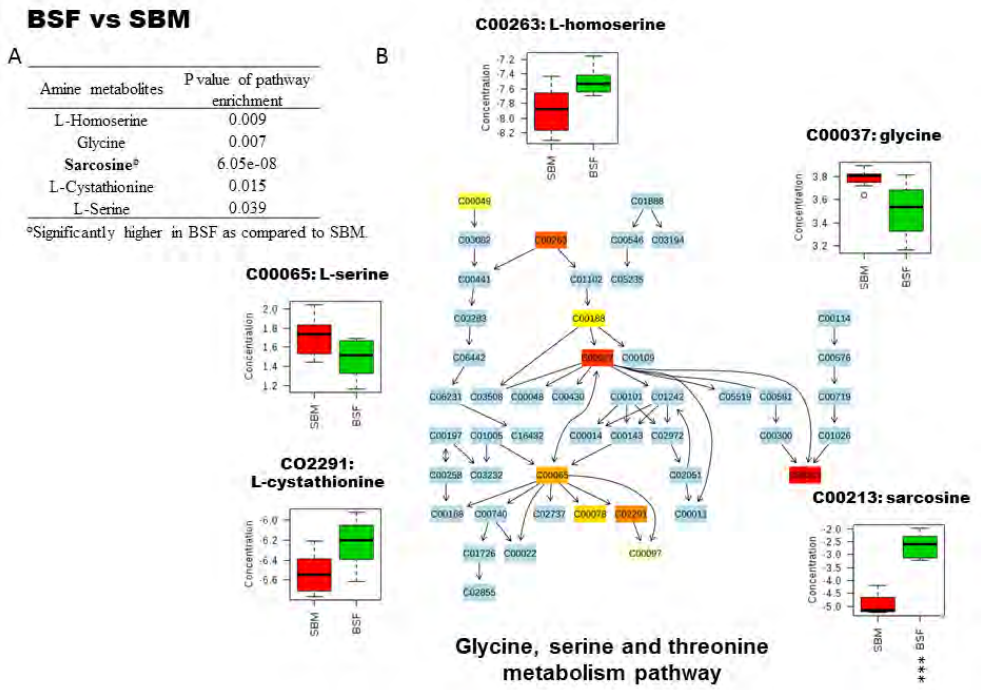

**Supplementary Figure 2: Avarage body weight in response to diets prepared with different protein ingredients.**

Bar plots are mean values of the body weight (in kg) in the treatment groups; whiskers are standard error mean. No significant (P < 0.05) difference was observed in the experimental diets compared to SBM based diet.



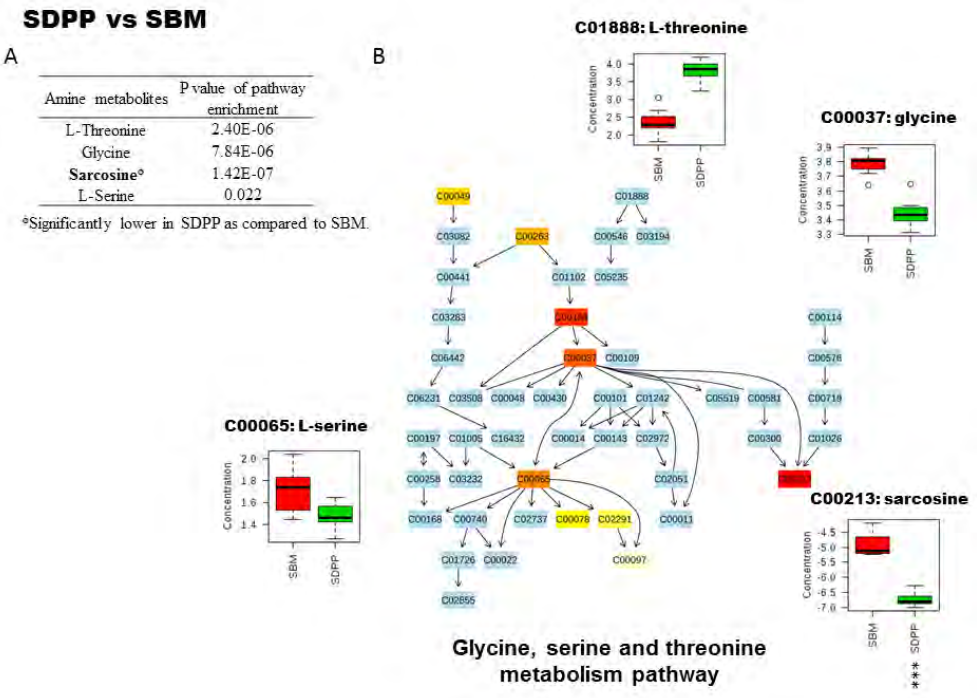

**Supplementary Figure 3: Concentrations of serum cytokines and chemokines in response to response diets prepared with different protein ingredients.**

Bar plots are mean values of the concentration level of cytokines/chemokines in the treatment groups; whiskers are standard error mean. No significant (P < 0.05) difference was observed in the experimental diets compared to SBM based diet.



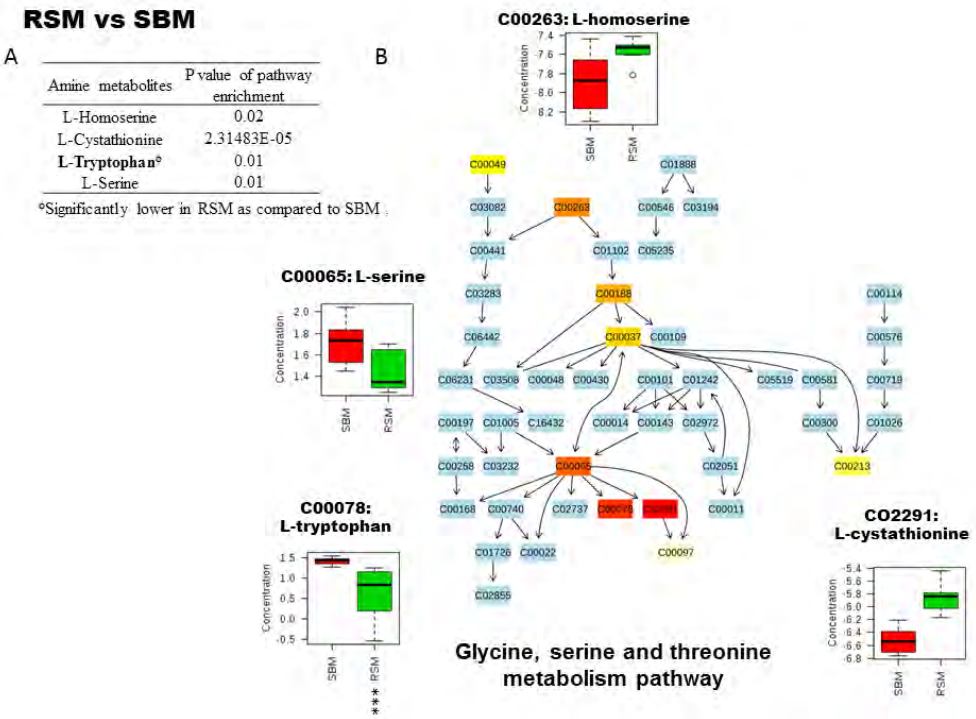

**Supplementary figure 4A: Glycine, serine and threonine metabolic pathway as represented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enriched in metabolic pathway analysis based on the amine profile in the plasma samples collected at the dissection days from pigs fed BSF and SBM diet.**

A. Amine metabolites that are significantly ($P < 0.05$) enriched in the pathway enrichment analysis. Within the significantly enriched amine metabolites, sarcosine is significantly high in BSF compared to SBM. Statistical significance was defined as *** is $P < 0.0001$. B. The glycine, serine and threonine metabolic pathway along with it significant enriched amine metabolites. "Alpha-numeric" representations are the KEGG identifiers. The colour (on gradient scale: cream to red, where towards red denotes less P value) of the rectangular-nodes in the KEGG pathway are all matched amines according to P values from pathway enrichment analysis. The sky blue rectangular nodes are the un-matched amines in the KEGG pathway. Black boxes represents the identified amines along with their concentration for the amine in each experimental diets represented in the box-plot. Box plots are mean value of the normalized concentration in the treatment groups; whiskers are standard error mean (for number of animal per dietary group, see supplementary table 3).



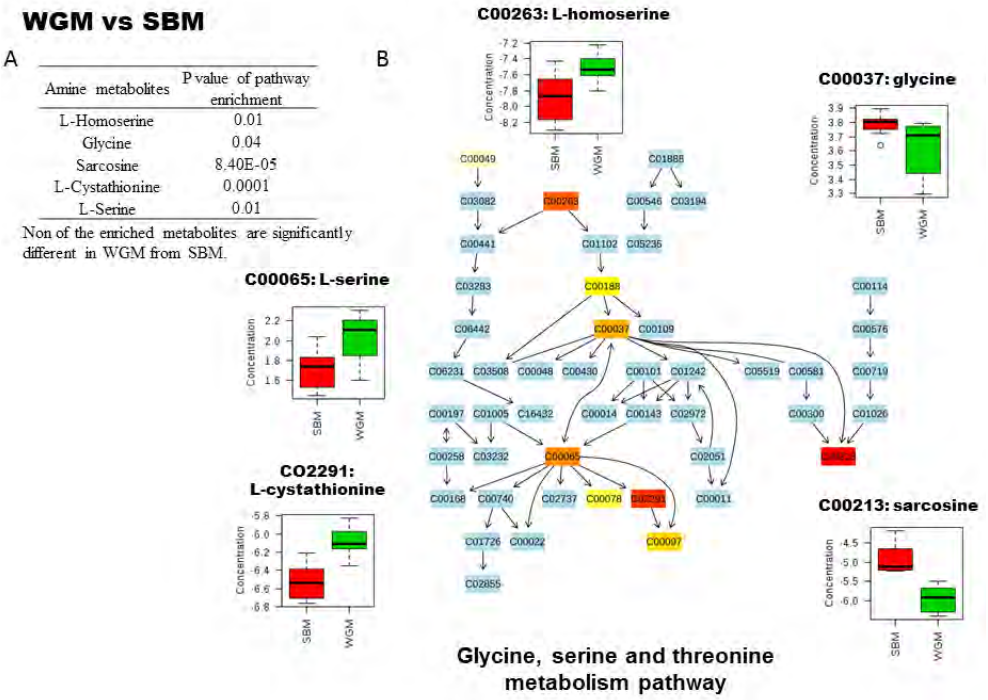

**Supplementary figure 4B: Glycine, serine and threonine metabolic pathway as represented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enriched in metabolic pathway analysis based on the amine profile in the plasma samples collected at the dissection days from pigs fed SDPP and SBM diet.**

A. Amine metabolites that are significantly ($P < 0.05$) enriched in the pathway enrichment analysis. Within the significantly enriched amine metabolites, sarcosine is significantly low in SDPP compared to SBM. Statistical significance was defined as *** is $P < 0.0001$. B. The glycine, serine and threonine metabolic pathway along with it significant enriched amine metabolites. "Alpha-numeric" representations are the KEGG identifiers. The colour (on gradient scale: cream to red, where towards red denotes less P value) of the rectangular-nodes in the KEGG pathway are all matched amines according to P values from pathway enrichment analysis. The sky blue rectangular nodes are the un-matched amines in the KEGG pathway. Black boxes represents the identified amines along with their concentration for the amine in each experimental diets represented in the box-plot. Box plots are mean value of the normalized concentration in the treatment groups; whiskers are standard error mean (for number of animal per dietary group, see supplementary table 3).



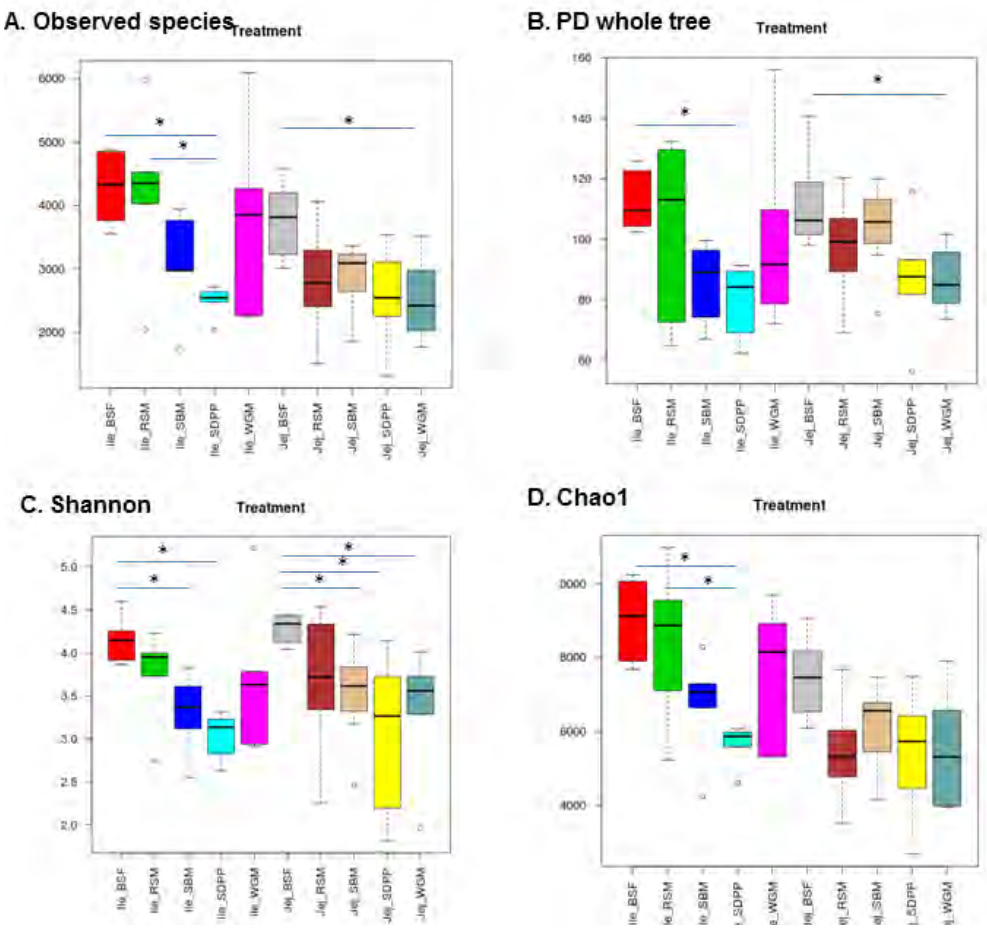

**Supplementary figure 4C: Glycine, serine and threonine metabolic pathway as represented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enriched in metabolic pathway analysis based on the amine profile in the plasma samples collected at the dissection days from pigs fed RSM and SBM diet.**

A. Amine metabolites that are significantly (P < 0.05) enriched in the pathway enrichment analysis. Within the significantly enriched amine metabolites, L-tryptophan is significantly low in RSM compared to SBM. Statistical significance was defined as *** is P < 0.0001. B. The glycine, serine and threonine metabolic pathway along with it significant enriched amine metabolites. "Alpha-numeric" representations are the KEGG identifiers. The colour (on gradient scale: cream to red, where towards red denotes less P value) of the rectangular-nodes in the KEGG pathway are all matched amines according to P values from pathway enrichment analysis. The sky blue rectangular nodes are the un-matched amines in the KEGG pathway. Black boxes represents the identified amines along with their concentration for the amine in each experimental diets represented in the box-plot. Box plots are mean value of the normalized concentration in the treatment groups; whiskers are standard error mean (for number of animal per dietary group, see supplementary table 3).




**Supplementary figure 4D: Glycine, serine and threonine metabolic pathway as represented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enriched in metabolic pathway analysis based on the amine profile in the plasma samples collected at the dissection days from pigs fed WGM and SBM diet.**

A. Amine metabolites that are significantly ($P < 0.05$) enriched in the pathway enrichment analysis. None of the enriched metabolites significantly differ in WGM from SBM. B. The glycine, serine and threonine metabolic pathway along with it significant enriched amine metabolites. "Alpha-numeric" representations are the KEGG identifiers. The colour (on gradient scale: cream to red, where towards red denotes less P value) of the rectangular-nodes in the KEGG pathway are all matched amines according to P values from pathway enrichment analysis. The sky blue rectangular nodes are the un-matched amines in the KEGG pathway. Black boxes represents the identified amines along with their concentration for the amine in each experimental diets represented in the box-plot. Box plots are mean value of the normalized concentration in the treatment groups; whiskers are standard error mean (for number of animal per dietary group, see supplementary table 3).




**Supplementary figure 5: Summary of diversity analyses from intestinal digesta of pigs fed diets containing proteins from different sources.**

*P < 0.05 is considered as significant.

**Supplementary Table 3: Taxon-Treatment association based on microbiota composition of genus level at jejunum and ileum in pigs fed diets containing protein from different sources.**

| Sl. No. | Taxon | Intestinal location | Treatment [a] |
|---|---|---|---|
| 1 | p__Actinobacteria;c__Actinobacteria;f__Actinomycetaceae;g__Actinomyces | Ileum | ↑BSF |
| 2 | p__Actinobacteria;c__Actinobacteria;f__Brevibacteriaceae;g__Brevibacterium | Ileum | ↑BSF |
| | p__Actinobacteria;c__Actinobacteria;f__Brevibacteriaceae;g__Brevibacterium | Jejunum | ↑BSF |

| Sl. No. | Taxon | Intestinal location | Treatment [a] |
|---|---|---|---|
| 3 | p__Actinobacteria;c__Actinobacteria;f__Corynebacteriaceae;g__Corynebacterium | Ileum | ↑BSF |
| | p__Actinobacteria;c__Actinobacteria;f__Corynebacteriaceae;g__Corynebacterium | Jejunum | ↑BSF |
| 4 | p__Firmicutes;c__Bacilli;Other;Other | Ileum | ↑BSF |
| | p__Firmicutes;c__Bacilli;Other;Other | Jejunum | ↑BSF |
| 5 | p__Firmicutes;c__Bacilli;f__Aerococcaceae;g__Facklamia | Ileum | ↑BSF |
| | p__Firmicutes;c__Bacilli;f__Aerococcaceae;g__Facklamia | Jejunum | ↑BSF |
| 6 | p__Firmicutes;c__Bacilli;f__Bacillaceae;Other | Ileum | ↑BSF |
| | p__Firmicutes;c__Bacilli;f__Bacillaceae;Other | Jejunum | ↑BSF |
| 7 | p__Firmicutes;c__Bacilli;f__Bacillaceae;g__Bacillus | Ileum | ↑BSF |
| | p__Firmicutes;c__Bacilli;f__Bacillaceae;g__Bacillus | Jejunum | ↑BSF |
| 8 | p__Firmicutes;c__Bacilli;f__Carnobacteriaceae;g__Carnobacterium | Jejunum | ↑BSF |
| 9 | p__Firmicutes;c__Bacilli;f__Enterococcaceae;Other | Ileum | ↑BSF |
| | p__Firmicutes;c__Bacilli;f__Enterococcaceae;Other | Jejunum | ↑BSF |
| 10 | p__Firmicutes;c__Bacilli;f__Lactobacillaceae;g__ | Jejunum | ↑BSF |
| 11 | p__Firmicutes;c__Bacilli;f__Staphylococcaceae;g__Staphylococcus | Jejunum | ↑BSF |
| 12 | p__Firmicutes;c__Bacilli;f__Streptococcaceae;g__Streptococcus | Jejunum | ↑SDPP |
| 13 | p__Firmicutes;c__Erysipelotrichi;f__Erysipelotrichaceae;g__[Eubacterium] | Jejunum | ↑RSM |
| 14 | p__Proteobacteria;c__Betaproteobacteria;f__Neisseriaceae;g__ | Jejunum | ↑WGM |
| 15 | p__Actinobacteria;c__Actinobacteria;f__Propionibacteriaceae;g__ | Jejunum | ↓WGM |
| 16 | p__Cyanobacteria;c__Chloroplast;f__;g__ | Jejunum | ↑SBM |
| | p__Cyanobacteria;c__Chloroplast;f__;g__ | Jejunum | ↑SBM |
| | p__Cyanobacteria;c__Chloroplast;f__;g__ | Jejunum | ↑SBM |
| | p__Cyanobacteria;c__Chloroplast;f__;g__ | Jejunum | ↑SBM |

[a]Significantly different in the treatment vs SBM

↑ is high and ↑ is low relative abundance.

**Supplementary Table 4: Differential enriched gene-sets in the jejunal mucosa of pigs fed diets with different protein sources relative to a diet with SBM.**

| | Up-regulated | Down-regulated |
|---|---|---|
| **BSF** | KEGG_METABOLISM_OF_ XENOBIOTICS_ BY_CYTOCHROME_P450 | KEGG_ECM_RECEPTOR_INTERACTION |
| | | KEGG_FOCAL_ADHESION |
| | | POTASSIUM_ION_TRANSPORT |
| | | GENERATION_OF_NEURONS |
| | | KEGG_RIBOSOME |
| | | NEURITE_DEVELOPMENT |
| | | NEUROGENESIS |
| | | CELL_CELL_ADHESION |
| | | NEURON_DEVELOPMENT |
| | | NEURON_DIFFERENTIATION |
| | | SYNAPTIC_TRANSMISSION |
| **SDPP** | NONE | LOCOMOTORY_BEHAVIOR |
| | | KEGG_COMPLEMENT_AND_COAGULATION_ CASCADES |
| | | CATION_HOMEOSTASIS |
| | | KEGG_ECM_RECEPTOR_INTERACTION |
| | | BEHAVIOR |
| | | CELLULAR_CATION_HOMEOSTASIS |
| | | MUSCLE_DEVELOPMENT |
| | | GENERATION_OF_NEURONS |
| | | NEURON_DIFFERENTIATION |
| | | NEURITE_DEVELOPMENT |
| | | ION_HOMEOSTASIS |
| | | NEUROGENESIS |
| | | AXONOGENESIS |
| | | NEURON_DEVELOPMENT |
| | | NERVOUS_SYSTEM_DEVELOPMENT |
| | | RESPONSE_TO_EXTERNAL_STIMULUS |
| | | RESPONSE_TO_WOUNDING |
| | | CELLULAR_MORPHOGENESIS_DURING_ DIFFERENTIATION |
| | | DNA_RECOMBINATION |

| | Up-regulated | Down-regulated |
|---|---|---|
| | | KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION |
| | | SYSTEM_DEVELOPMENT |
| | | KEGG_FOCAL_ADHESION |
| | | CELL_PROLIFERATION_GO_0008283 |
| | | KEGG_PRION_DISEASES |
| | | CELLULAR_HOMEOSTASIS |
| | | IMMUNE_RESPONSE |
| | | KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION |
| | | ANATOMICAL_STRUCTURE_DEVELOPMENT |
| | | KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS |
| | | G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY |
| | | MULTICELLULAR_ORGANISMAL_DEVELOPMENT |
| | | KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION |
| | | KEGG_ASTHMA |
| | | KEGG_BASAL_CELL_CARCINOMA |
| | | CELL_MIGRATION |
| | | KEGG_CELL_ADHESION_MOLECULES_CAMS |
| **RSM** | KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450<br>KEGG_RETINOL_METABOLISM<br>KEGG_TRYPTOPHAN_METABOLISM | NONE |
| **WGM** | KEGG_DRUG_METABOLISM_CYTOCHROME_P450<br>KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY<br>KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY | NONE |

| | Up-regulated | Down-regulated |
|---|---|---|
| | KEGG_PRIMARY_ IMMUNODEFICIENCY | |
| | KEGG_JAK_STAT_SIGNALING_ PATHWAY | |
| | KEGG_T_CELL_RECEPTOR_ SIGNALING_PATHWAY | |
| | KEGG_METABOLISM_OF_ XENOBIOTICS_ BY_CYTOCHROME_P450 | |
| | KEGG_FC_GAMMA_R_ MEDIATED_PHAGOCYTOSIS | |

**Supplementary Table 5: Differential enriched gene-sets in the ileal mucosa of pigs fed diets with different protein sources relative to a diet with SBM.**

| | Up-regulated | Down-regulated |
|---|---|---|
| **BSF** | KEGG_TRYPTOPHAN_ METABOLISM | KEGG_ECM_RECEPTOR_INTERACTION |
| | KEGG_PPAR_SIGNALING_ PATHWAY | NEURON_DIFFERENTIATION |
| | | GENERATION_OF_NEURONS |
| | | NEUROGENESIS |
| | | NEURITE_DEVELOPMENT |
| | | NEURON_DEVELOPMENT |
| | | AXONOGENESIS |
| | | NEGATIVE_REGULATION_OF_CELLULAR_ PROTEIN_METABOLIC_PROCESS |
| **SDPP** | NONE | NONE |
| **RSM** | NONE | NONE |
| **WGM** | KEGG_RETINOL_ METABOLISM | CELL_CYCLE_PROCESS |
| | | M_PHASE |
| | KEGG_DRUG_ METABOLISM_ CYTOCHROME_P450 | CELL_CYCLE_PHASE |
| | KEGG_ABC_TRANSPORTERS | DNA_DEPENDENT_DNA_REPLICATION |

| Up-regulated | Down-regulated |
|---|---|
| KEGG_PROXIMAL_TUBULE_BICARBONATE_RECLAMATION | KEGG_BASE_EXCISION_REPAIR |
| KEGG_SPHINGOLIPID_METABOLISM | MITOTIC_CELL_CYCLE |
| SPHINGOLIPID_METABOLIC_PROCESS | DNA_REPLICATION |
| | DNA_REPAIR |
| | DNA_METABOLIC_PROCESS |
| | M_PHASE_OF_MITOTIC_CELL_CYCLE |
| | KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION |
| | CELL_CYCLE_GO_0007049 |
| | KEGG_CELL_CYCLE |
| | DNA_RECOMBINATION |
| | CELL_CYCLE_CHECKPOINT_GO_0000075 |
| | RESPONSE_TO_DNA_DAMAGE_STIMULUS |
| | MITOSIS |
| | CHROMOSOME_SEGREGATION |
| | DNA_INTEGRITY_CHECKPOINT |
| | REGULATION_OF_DNA_METABOLIC_PROCESS |
| | KEGG_HOMOLOGOUS_RECOMBINATION |
| | KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS |

# 8

# CHAPTER 8

## General Discussion

## Background and Outline

The world population is expected to increase from 7.3 billion in 2015 to 11.2 billion in 2100 (UN, 2015). The expected increase in per capita income will result in an increased demand for livestock and dairy products (FAO, 2004; FAOSTAT, 2009). Protein-containing ingredients are needed to produce such products resulting in an increased demand for protein (Chadd et al., 2004). The protein used in diet for livestock nutrition originates from a variety of sources including forages, grains, legumes, animal meals and various by-products (Boland et al., 2013). However, cultivatable arable land is already limited and there is competition between protein used for human food and for animal feed, which ultimately could result in protein scarcity (FAO, 2009). To avoid the so-called feed vs food competition, there is a strong need to improve the efficiency of use of current and new protein sources for animal feed (FAO, 2004; Aiking, 2011; Boland et al., 2013; Jansman, 2016).

Healthy gut is prerequisite for sustainable and efficient animal production. Investigating the functioning of the intestine as a system is instrumental to understand how the diverse and closely connected processes in the intestinal tract are influenced by internal and external factors (Figure 1 of **Chapter 6**). In this context, it is important to understand which processes of the intestinal tissue are modulated by external (e.g. dietary) factors. The intestine is a tissue harbouring dynamic ranges of multiple cell types that strongly interact and communicate with each other. These interactions are influenced by diet and microbial constituents in the intestinal lumen. Recent developments in the areas of genomics, proteomics, metabolomics and computational sciences now provide us with tools and methods to start studying the behaviour of biological systems as a whole. In the research described in this thesis we applied genomics, proteomics, metabolomics (~omics) and computational approaches to describe (molecular) processes that are associated with the responses towards dietary protein sources of intestinal epithelial cells, the gut microbiota and the gut as a system using both *in vitro* and *in vivo* approaches.

This research started with characterisation of the composition of different protein sources using proteomic and bioinformatic procedures and developing a novel in *silico* approach to predict bioactive peptides and amino acids (AA) composition within the protein source (**Chapter 2**). The primary objective of this thesis was to investigate the effects of existing and novel protein sources in animal diet on changes in local (intestinal mucosa) and several systemic molecular and cellular parameters, thereby focussing on parameters assumed to be important for immune competence and immune homeostasis in the gut. Therefore the focus was on evaluating the responses of intestinal epithelial cells, gut microbiota and the gut as a system to different protein sources using both *in vitro* and *in vivo* approaches. We applied different ~omics techniques (**Chapters 3-5,7**) and a

systems biology approach (**Chapter 6**) to identify and describe the major components and processes that are influenced by the protein sources, in order to get a global view of the physiological activity of the intestinal mucosa. A new *in vitro* enteroid model was used to investigate its potential to study, in the absence of luminal microbiota, the immediate early effects of dietary ingredients on primary cells of the intestinal epithelium (**Chapter 3**). Thereafter, an *in vivo* study was performed to gather knowledge on the long-term effect of existing and novel protein sources. Changes in various local and systemic physiological parameters in mice were evaluated as induced by diets containing common and new protein sources (**Chapter 4 and 5**). Systemic effects were measured by the concentrations of blood cytokines and chemokines along with amine metabolites. Further, novel "nutritional-systems biology" approaches were used to demonstrate that relationships exist between distinct biological scales by integrating multi-scale quantitative (~omics) data, originating from the mice experiment (**Chapter 6**). Finally, an *in vivo* experiment with pigs (**Chapter 7**), was carried out to evaluate the effect of existing and novel protein sources on intestinal functioning and health of the target animal. The major findings described in the experimental chapters of this thesis and their perspectives are discussed in this chapter. Finally, recommendations for future research are given.

***Protein sources are potentially rich in bioactive peptides***

The capabilities of modern-day proteomic techniques were employed to characterise and quantify individual protein components within complex protein-containing feed ingredients. Such an approach allows information to be obtained on the collective group of individual proteins. In **Chapter 2**, we detected and semi-quantified 37, 58, 85, 188, 113 and 33 different individual proteins in casein (CAS), partially delactosed whey powder (DWP), spray dried porcine plasma (SDPP), soybean meal (SBM), wheat gluten meal (WGM) and yellow meal worm (YMW), respectively. On repeating the same experiment, one may get higher number of individual proteins for certain ingredients, particularly for YMW, due to the expected increase in information of the YMW protein composition submitted to the proteomic databases over time. The bioactive peptides potentially generated by digestion of the proteins was predicted in *silico* using bioinformatic approaches. The potential bioactivities were antithrombotic, antihypertensive, immunomodulating, antioxidative, antimicrobial and ileum contracting. These findings may be of practical relevance since several reports provided evidence that protein motifs derived from protein sources exert biological functions, although this is mainly investigated under *in vitro* conditions (Yamamoto et al., 1994; Maeno et al., 1996; Mullally et al., 1997; Lindmark-Mansson and Akesson, 2000; Teixeira et al., 2000; Gibbs et al., 2004; Pihlanto, 2006; Moller et al., 2008; Gao et al., 2011; Burris et al., 2014; Shinmoto et al., 2014). The results as described in **Chapters 4, 5** and **7**, however, did not provide clear evidence for the in *silico* predicted bioactivities of the protein sources in

mice and pigs (**Chapter 4, 5** and **7**). Possible explanations for the discrepancy between the *in silico* predicted effects of bioactive peptides and the *in vivo* observations (data generated in **Chapter 4, 5** and **7**) are that the active peptides were not absorbed or did not attain sufficient concentration in blood and target tissues to have their biological effects. Additionally the peptide bioactivity may be modulated due to molecular interactions with constituents from other ingredients of the diets. Furthermore the data generated *in vivo* may not have been sufficient to reveal all the predicted bioactivities.

While working on the in *silico* approach (**Chapter 2**), we have used the "BIOPEP" database (Minkiewicz et al., 2008; Iwaniak and Dziuba, 2011) as a tool for the evaluation of protein as the precursors of bioactive peptides. This freely available database hosts the sequence databases of bioactive peptides which are being mapped with the peptides derived from the protein sequences present in the protein source after in *silico* proteolytic process. We noticed that the database of bioactive peptides was last revised or updated in 2009. However, scientific discoveries in the field of bioactive peptides continue to grow (Correa et al., 2016; Gevaert et al., 2016; Gupta et al., 2017) and an update of the repository databases containing the sequence of bioactive peptides is needed.

By using the *in silico* approach described in **Chapter 2**, one can detect individual proteins in protein rich feed ingredients, thereby providing more detailed information on the composition of complex protein sources compared to conventional nutritional analytical approaches. Such information characterises the potential functional properties of ingredients that can be used in the future for formulating diets with specific functional properties. This topic is further discussed in the future perspective section of the general discussion.

***Intestinal organoids (enteroids) as a tool to study diet-host interaction***

Absence of microbiota in the enteroid model system can be of advantage as it offers the possibility to investigate *in vitro* the direct effects of dietary components on a near physiological model of the intestinal epithelium. In **Chapter 3** the utility of the system was tested with different dietary protein ingredients. In this study the protein sources were deliberately exposed to the enteroids in an undigested form, in order to investigate activities of solubilised components present in the protein source, e.g. fats, glycans and small proteins or modified peptides. As demonstrated in **Chapter 3** the enteroid model system contained all the cell types present in the tissue of origin and displayed responses specific to each protein source. These results obtained for SBM, about which most is known, were consistent with *in vivo* studies on the hypotriglycerdemic effect of soy protein and its effect on retinoid receptors in the liver suggesting that a component of SBM protein also negatively regulates cholesterol and lipid biosynthetic pathways through down regulation of retinoic acid receptors in the intestinal epithelium. This

result highlights organoids as a promising new model to evaluate complex interaction between food and feed ingredients and the intestinal epithelium.

The use of enteroids as an *in vitro* test system has several advantages over mono-cellular *in vitro* cell culture assays or the *in vivo* system. Unlike *in vitro* cell culture assay, the enteroid system consists of multiple cell types which are present in the epithelial lining and mucosa of the intestine. Being a relatively less complex system, enteroids can be used to study the specific interaction between the test-ingredient and the host mucosa. The read-out allows the identification of both signalling and response molecules within and between different cell types. Such information may enrich the knowledge on the functional properties of dietary ingredients. Moreover, enteroids can be used as a tool to gain insight into the molecular understanding of events that regulate complex phenotypes such as feed-efficiency or disease resistance varying across genotypes (Knap, 2005; Rauw, 2012; Bishop and Woolliams, 2014; Rauw and Gomez-Raya, 2015) in livestock species.

Although enteroids mimic the cellular composition and function of the intestinal mucosa much better than mono-cellular systems, such as Caco-2 and IPEC, it is nonetheless a reductionist model. For example, these models lack the complex microbial community structures, which play a key role in maintaining intestinal homeostasis *in vivo*. Sub-populations of the intestinal microbiota interact with dietary components to form (bioactive) metabolites that may communicate to host mucosal cells through distinct mechanisms. A body of knowledge is accumulating that points to the gut microbiota as a mediator of dietary impact on the host metabolic status and the innate immune system (van Baarlen et al., 2013; Sonnenburg and Backhed, 2016; Thaiss et al., 2016). Furthermore, in the intestinal lumen there is a continuous "three-way" or "triangular" interaction of microbiota-host-diet that can modulate gut health and intestinal functions (Zhang et al., 2009; Ussar et al., 2015; Wells et al., 2017). Although the triangular interactions can be studied in *in vivo* experimental models (**Chapters 4, 5** and **7**), they are often too complex to fully interpret and do not always allow the identification of causal relationships. The knowledge generated by the two-way interactions in the enteroid model, can therefore contribute in a better understanding of the complex three-way interactions that occur *in vivo*.

***Nutrition regulates essential processes related to gut health***

Many complex traits in livestock are regulated at the level of expression of the genome (Cookson et al., 2009). Environmental factors can directly influence the transcription of the genome (Mitroi and Mota, 2008; Chou et al., 2012; Ganesh and Hettiarachchy, 2012). A recent review by te Pas and colleagues has shown the importance of the genotype and environmentally-induced epigenetic modifications for the transcriptomic

activity of genes in the genome (tePas et al., 2017). In addition, environmental factors are also shown to influence the transcription of the genome including dietary components (Efeyan et al., 2015). The findings as described in this thesis demonstrate that the choice of a dietary ingredient also has a significant impact on the expression of genes in the intestinal mucosa. In **Chapter 4** it is demonstrated that the activity of the mammalian (mechanistic) target of rapamycin (mTOR) pathway was diminished by using a SBM based diet in mice. The mTOR is a conserved protein kinase involved in a multitude of cellular processes including cell growth. The mTOR senses multiple upstream signals which include nutrients such as amino acids, growth factors and stress (Kim and Guan, 2011; Jewell and Guan, 2013). The mTOR downstream signalling pathway integrates both intracellular and extracellular signals and serves as a master regulator of proliferation and differentiation of cells including gut stem cells and T cells (Sarbassov et al., 2005; Dibble and Manning, 2013; Zarogoulidis et al., 2014; Weichhart et al., 2015). Modulation of this pathway is relevant for livestock (Qin et al., 2016; Ye et al., 2016; Wang et al., 2017), as it is likely to be involved in balancing the systems between renewal of intestinal epithelial (gut barrier) and proliferation of mucosal immune cells (regulation of immune tolerance and immune response) and overall host's metabolism. The results as described in **Chapter 4** demonstrate that protein ingredients have the potential to influence master regulators, like mTOR, that link nutritional conditions to (gut) health related aspects. It underscores the necessity to take into account the functional properties of protein sources in the formulation of diet for livestock.

In **Chapter 7**, the transcriptional inhibition of mTOR signalling and related response parameters with any protein ingredients could not be observed in pigs. Recently, Yang and colleagues have shown that the mTOR signalling pathway involved in energy metabolism, Golgi vesicle transport, protein amino acid glycosylation, cell secretion, transmembrane transport, ion transport, nucleotide catabolic process, translational initiation, epithelial cell differentiation and apoptosis, was reduced in post-weaned pigs compared to pre-weaned pigs (Yang et al., 2016). This suggests that the mTOR signalling activity in the gut of post-weaned piglets may be dampened which might be due to weaning-induced cell cycle arrest and apoptosis in the small intestine of pigs (Zhu et al., 2014). Further, the level of maturity of neonatal gut at birth differs between species (Tourneur and Chassin, 2013). The observed difference in mTOR signalling between pigs and mice might therefore be due to the dissimilarity in the level of maturity of intestine at a given age. However, to verify this hypothesis, further research is required.

***Amine metabolites endo-phenotype or biomarker***

The development of biomarkers of gut health is imperative to gain clarity in understanding of the patho-physiological events that influence the intestinal barrier, intestinal functionality particularly in relation to nutrient absorption and the ecology

of the GIT microbiota. While there is considerable knowledge in biomarkers that are indicative of the GIT ability to absorb, transport and secrete major macro and micro-nutrients, a large gap in the literature exists in relation to biomarkers of GIT permeability, GIT barrier function, or biomarkers that are indicative of the functional presence of beneficial microbiota or their metabolites in relation to dietary components (Wells et al., 2017). Profile of blood amine metabolites, not only reveals information about the metabolic processing of nutrients from protein sources, but also the metabolic and absorptive capacity of the gut, including the influence of the microbiome (Kogut and Arsenault, 2016). However, this has not been previously studied in relation to different protein sources. We provided evidence that measuring systemic amine metabolites could potentially fill this knowledge gap. Further, these profiles reveal biomarkers that could be used to assess the nutritional quality of different protein sources and the health status of the animals (**Chapter 5** and **7**).

In addition to peripheral blood, it might be interesting to measure the amine profile in hepatic-portal blood to further strengthen the role of amine metabolites as biomarkers for protein ingredients in relation to gut health. In future, amine-based endo-phenotypes can be employed for predicting dietary protein-associated phenotypes in relation to health e.g. renal disease, bone-calcium homeostasis, functioning of the liver (Martin et al., 2005; Delimaris, 2013) and to the efficiency of protein and amino acid utilization in systematic metabolism. This concept is similar to the current practice, for example, the use of elevated total cholesterol in serum as a consistent bio-indicator or biomarker of cardiovascular diseases and health risk associated with consumption of high fat diet (Siri-Tarino et al., 2015). However, further research is needed to establish this concept for protein containing feed ingredients. Moreover, a similar concept could be applied for fat- and carbohydrate based ingredients.

***Nutritional-systems biology approach as a tool to explore the nutritional-black box***

Most traits or phenotypes of livestock animals are found to be complex. Although we know of cases where a particular phenotype is caused by a single gene, there are no known examples where a phenotype is caused by a single (nutritional) factor. Due to the involvement of multiple genes and nutritional factors in the biological regulation of phenotypes, several molecular and cellular components and networks of components are expected to act together in the expression of a particular external phenotype. Among others, the molecular components include components of different biological levels, including transcribed mRNA, translated proteins and biosynthetic metabolites. The cellular components may comprise of different cell types and may also include specific taxa of the organism's microbiome. Establishment of a phenotype is contributed by both signalling and biosynthetic pathways of different cell types and communication between cell-types. In **Chapters 4, 5** and **7**, the effect of dietary protein interventions was studied

at different biological levels. It is tempting to speculate, however, that these biological levels do not function independently from each other and different feedback loops between the various biological levels regulate the expression of an external phenotype. In **Chapter 6**, first evidence is provided that such feedback loops exist. Several components within a particular biological level (connectivity hubs) were identified that directly or indirectly participate in processes of another biological level.

Integration of data derived from different biological levels is one of the main objectives of the emerging systems biology discipline, which is required to improve our fundamental knowledge on nutrient-microbiota-host interactions in the GIT. As shown in **Chapter 6,** such an approach successfully generated hypothesis for targeted research to identify causal relationships between the various biological scales. In future, hypothesis driven research based on the prediction made by systems biology approaches, will diminish our knowledge gap between dietary intervention and external phenotype.

***Gut Health and Nutrition 2.0***

Besides digestion and absorption of nutrients, the GIT is responsible for regulating physiological homeostasis that provides hosts the ability to withstand infectious and non-infectious stressors (Bartlett and Smith, 2003; Sansonetti, 2004; Garriga et al., 2006; Choct et al., 2010; Quinteiro-Filho et al., 2010; Crhanova et al., 2011; Maslowski and Mackay, 2011; Quinteiro-Filho et al., 2012; Arsenault et al., 2013). This statement and the term "Gut health" originate from the domain of human medicine field (Bischoff, 2011). The performance of production animals is hugely dependent on optimal gut health. There is increasing evidence that there is a direct relationship between animal performance and a "healthy" GIT (Collins et al., 2009; Ivarsson et al., 2014; Lindberg, 2014; Angelakis, 2016; Jha and Berrocoso, 2016). However, there is no clear definition of "gut health" as it covers multiple aspects of the GIT that encompasses a number of physiological and functional features, including nutrient digestion and absorption, host metabolism and energy generation, a stable microbiome, mucus layer development, barrier function and mucosal immune responses (Nurmi et al., 1992; van Der Wielen et al., 2000; Beckmann et al., 2006; Qu et al., 2008; Shakouri et al., 2009; Oakley et al., 2014). Although influencing gut health is an increasingly important topic in animal nutrition research, yet a clear scientific definition is still lacking. Recently two definitions of gut-health from the context of livestock have been proposed: 1) According to Kogut and Arsenault (2016), gut-health can be defined as the absence/prevention/avoidance of disease so that the animal is able to perform its physiological functions in order to withstand exogenous and endogenous stressors; 2) According to Jansman (2016), health of the GIT could be defined as its capacity to exert its different functions allowing the animal to achieve its potential productive performance under a variety of environmental

conditions. In short, gut health is the gatekeeper of health and the motor that drives animal performance (Kogut and Arsenault, 2016).

Different components of the intestinal microbial ecology are important in determining the gut health and growth in production animals (Figure 1). Clearly, this knowledge provides nutritionists with opportunities to develop feed additives, diet and feeding concepts that can promote conditions in the GIT, that would create and maintain an optimal balance between host mucosa and the microbiota in the lumen of the GIT and prevent disturbance of the structure and functionalities of the GIT. To improve "gut health" in livestock via nutrition, it is necessary to get in depth knowledge on the functionality (nutrient digestion and absorption) and immune status of the GIT that are important in determining animal performance.

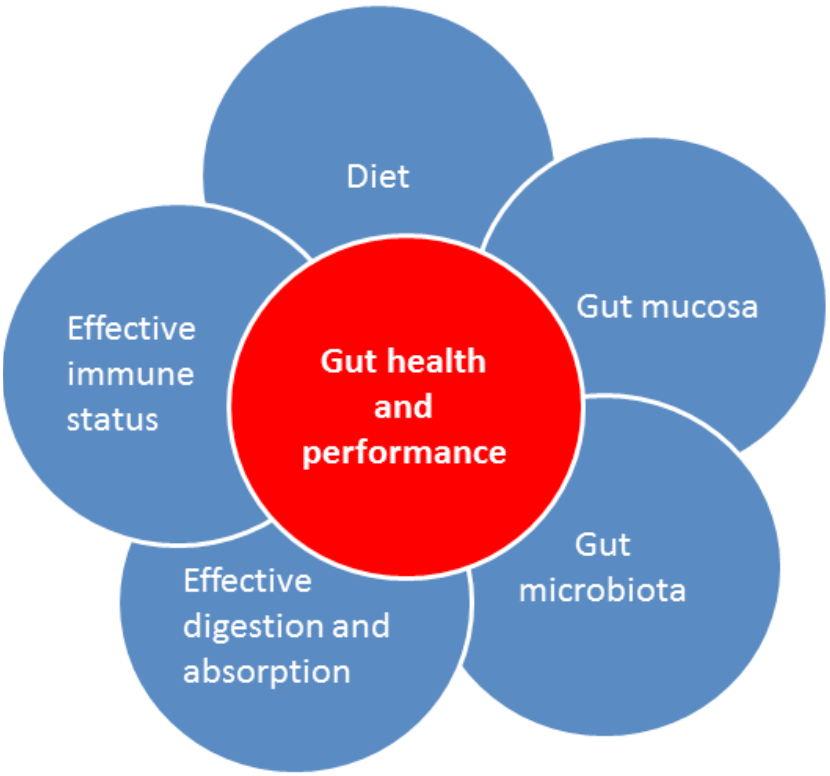



**Figure 1: Schematic representation of the different components of the intestinal ecology important in determining gut health and growth in production animals.**
Each component interacts with the other in order to maintain a dynamic equilibrium, a state defined as gut health. Source: Pietro Celi, DSM Nutritional Products.

This type of research requires a shift from traditional animal nutrition research that was largely focused on defining animal requirements for various nutrients, based on their roles as anabolic substrates, metabolic cofactors and energy sources. Evolving methods in molecular biology and physiology have now revealed far more diverse and far-reaching impacts of dietary components which is also named as "Animal Nutrition 2.0". In this thesis, by adopting a FeedOmics approach we further add elements to build in the concept of Animal Nutrition 2.0. The outlook of Animal Nutrition 2.0 is discussed in the "perspectives and future research" paragraph presented below.

## Perspectives and Future Research

### *Protein profile analysis by proteomics*

Information of protein profiles of protein sources could be of great benefit for multiple stake holders, particularly for food/feed manufacturers and technologists, human health and livestock researchers. Our combined proteomic and bioinformatic approach may be exploited by industries as assessment of quality for protein based products such as dairy based products. Sampling can be done at multiple "check-points" in the production line during manufacturing. This enables the food/feed manufacturers or technologists to gather detailed information on the quality of the protein based products produced in each batch. In addition, one can make assessment on the performance of food/feed processing techniques applied during the manufacturing of protein based products. These certainly can pave the way towards the development of improved methods for "Quality Control" of protein-based products thus bringing added value to the stakeholders.

To study kinetics of protein degradation, time series of intestinal digesta samples taken from different GIT compartments can be analysed for protein, peptide and amino-acid composition by employing the proteomic approaches as described in this thesis. It is known that peptides are prone to aggregation processes that lead to indigestible peptide aggregates, a phenomenon that was observed for soy-derived peptides in the intestine of pigs (Fischer et al., 2007). Using proteomic approaches, these peptides aggregates can be identified and monitored during the digestion process. Furthermore, application of proteomic technologies on digesta and faeces can provide detailed information on digestion degradation kinetics of the "parent" dietary protein along with identifying the nature of "undigested dietary protein" and endogenous proteins.

### *Functional properties of protein sources*

The functional properties of protein sources should be taken into account during formulation of diet. Here, in this thesis, we have shown that apart from the "strict-nutritional" value, diet along with their ingredients and constituents have other "non-strict-nutritional" functional properties. However, the functional properties differ

among the protein sources. In *in silico* approach, we observed differences in diversity of bioactive peptides predicted to be available during enzymatic degradation. Spray dried plasma protein (SDPP) was predicted to give rise to a higher diversity of bioactive peptides compared to other protein sources used in the study. The *in vitro* approach of enteroids also suggested different functional properties of protein sources through induction of unique biological processes. SDPP was found to upregulate several biological pathways associated with processes of cell migration, movement, related to increased cell turnover, in enteroids. Among other protein sources, CAS down-regulated glutathione metabolism, SBM down-regulated biological processes involved in triglyceride, phosphatidylcholine biosynthesis and pathways. In *in vivo* experiment with mice, SBM based diet was distinctly different from other protein sources by reducing mTOR pathway expression and strongly increasing abundance of *Bacteroidales* family *S24-7* (as mentioned in **Chapter 4**). In blood, we also detected differences in amine profile of mice and pigs fed protein from different sources (**Chapter 5 and 7**). These indicates that FeedOmics approach can be used as toolbox that can measure the impact of the protein ingredients from different sources on host's physiology in particular to metabolism and immunity. Characterisation of the functional properties of protein sources can thus be done using FeedOmics approach.

In current practice, diet formulation for livestock is based on the provision of ileal or faecal digestible nutrients and derived provision of metabolisable or net energy by feed ingredients. Thus, it considers only the "strict-nutritional" value of feed ingredients which relate to proteins/ AA, starch and sugars, fats, fermentable non-starch-polysaccharides, minerals and vitamins. From the context of protein ingredients, we provide the elements/ technique/methods to measure the presence of protein-based functional properties in some common and new protein sources for pigs, using in *silico*, *in vitro* and *in vivo* approaches. To use this information for practical purposes, efforts are required towards the quantification of these functional properties of feed ingredients. By adopting several of the here described approaches, one can preselect groups of proteins and/or peptides with a particular bio-functional property that are present in the ingredients. These can be further quantified by employing targeted ~omics approaches in different models (*in silico*, *in vitro, in vivo*) (Gallien et al., 2012; Wienkoop and Staudinger, 2013; Shi et al., 2016). Moreover, the quantified values of such functional properties could be included in the existing feed tables (CVB, 2011; NRC, 2012) as "non-strict nutritional" values alongside the existing "strict nutritional values". Combining unbiased ~omics approaches with targeted approaches appears to have true potential to contribute for quantifying the "non-strict nutritional" values. This will support the field of animal nutrition to consider both strict- and non-strict nutritional value of feed ingredients while composing animal diet.

***Further development of Animal Nutrition 2.0***

Animal Nutrition 2.0 leads to continued hope that dietary alterations remain a fruitful area of investigation in the quest for improved gut health and function. Advancement of methods in molecular biology and physiology can further aid studies to get a clear picture. This will enable nutritionists to develop more targeted diets with specific functional properties directed towards supporting gut health and towards creating immune competent and resilient animals, for example via nutrition-based early life programming of later life metabolism and immune competence. The development of these products need to take into account the functional properties of feed ingredients that can impact animal biology by influencing microbial communities, activating cellular receptors, or by directly altering activity of digestive enzymes in the gut.

The approach taken in this thesis enabled to gather knowledge at the tissue and cellular level by investigating the response of its molecular components as influenced by diets containing proteins from various sources. Different biological levels were represented by different molecular components such as transcribed mRNA, translated proteins and biosynthetic metabolites of host cells and the intestinal bacterial genome. Expanding such approaches by incorporating the dynamics of time dependent changes in the profiles of the various molecular and cellular components by measuring along the whole gut might further diminish the knowledge gap on the functional properties of protein sources. Moreover, validation experiments are required to demonstrate effects of the functional properties of protein sources on animals that are kept under the challenging condition of current husbandry systems for livestock.

Each of the components studied in this research has an established legitimacy in biological science which is collectively named as ~omics. They are often used to conduct large-scale studies of pathways and networks of the molecular components for a better understanding of the biology as a system., However, certain shortcomings were encountered while using them in the FeedOmics approach, particularly in relation to metabolomics. The majority of the metabolic databases that are freely assessable are having detailed information about small molecule metabolites found in the human body (e.g. The Human Metabolome Database, HMDB and The BiGG database) or in micro-organism (e.g. SYSTOMONAS (SYSTems biology of pseudOMONAS). No metabolic database is currently available with detailed information about metabolites found in the livestock species. Unless livestock species specific databases are being created, livestock researchers have to rely on the human databases thereby compelling researchers to extrapolate the knowledge gained from human metabolism to livestock species. If livestock metabolic databases are being developed in future then, enrichment of the database could be done with data of metabolic profiles of digesta and blood, considering production purpose and age of the livestock species involved.

In the current approach of FeedOmics, we have used 16S rRNA sequencing of microbial DNA to answer the question, "who are present in the gut?". As an answer, we obtained a catalogue of microbes residing in the gut. However, it is also important to recognise the potential of the residing microbes as a genetic content and what they do. Recent technological developments allow to provide answers to both the questions by means of metagenome sequencing and meta (-transcriptomics, -proteomics, -metabolomics), respectively (Deusch et al., 2015). Employing such extended methodologies in a FeedOmics approach would certainly enhance its resolution towards understanding of the existing complexity in the gut microbiota in relation to nutrition.

Further, intestinal organoids of targeted species, for example from pigs or poultry can also be used to study the efficacy of the digested and non-digested components of feed ingredients. This approach potentially can unveil novel (species-specific) functional properties of ingredients that can impact gut biology by various means, for example activating cellular receptors or directly altering cellular metabolic or functional activity. This approach can truly be explored for enriching the knowledge base towards the functional properties of ingredients, which is still in its infancy stage.

## References


1. Aiking, H. 2011. Future protein supply. Trends in Food Science & Technology 22: 112-120. doi:10.1016/j.tifs.2010.04.005
2. Angelakis, E. 2016. Weight gain by gut microbiota manipulation in productive animals. Microbial Pathogenesis. doi:http://dx.doi.org/10.1016/j.micpath.2016.11.002
3. Arsenault, R. J., S. Napper, and M. H. Kogut. 2013. Salmonella enterica Typhimurium infection causes metabolic changes in chicken muscle involving AMPK, fatty acid and insulin/mTOR signaling. Veterinary Research 44: 35. doi:10.1186/1297-9716-44-35
4. Bartlett, J. R., and M. O. Smith. 2003. Effects of different levels of zinc on the performance and immunocompetence of broilers under heat stress. Poultry science 82: 1580-1588.
5. Beckmann, L., O. Simon, and W. Vahjen. 2006. Isolation and identification of mixed linked beta -glucan degrading bacteria in the intestine of broiler chickens and partial characterization of respective 1,3-1,4-beta -glucanase activities. J Basic Microbiol 46: 175-185. doi:10.1002/jobm.200510107 [doi]
6. Bischoff, S. C. 2011. 'Gut health': a new objective in medicine? BMC Medicine 9: 24. doi:10.1186/1741-7015-9-24
7. Bishop, S. C., and J. A. Woolliams. 2014. Genomics and disease resistance studies in livestock. Livestock Science 166: 190-198. doi:http://dx.doi.org/10.1016/j.livsci.2014.04.034
8. Boland, M. J., A. N. Rae, J. M. Vereijken, M. P. M. Meuwissen, A. R. H. Fischer, M. van Boekel, S. M. Rutherfurd, H. Gruppen, P. J. Moughan, and W. H. Hendriks. 2013. The future supply of animal-derived protein for human consumption. Trends in Food Science & Technology 29: 62-73. doi:10.1016/j.tifs.2012.07.002
9. Burris, R. L., H. P. Ng, and S. Nagarajan. 2014. Soy protein inhibits inflammation-induced VCAM-1 and inflammatory cytokine induction by inhibiting the NF-kappaB and AKT signaling pathway in apolipoprotein E-deficient mice. Eur J Nutr 53: 135-148. doi:10.1007/s00394-013-0509-7
10. Chadd, S. A., W. P. Davies, and J. M. Koivisto. 2004. Practical production of protein for food animals Protein sources for animal feed industries. p 77-124. FAO of UN, Rome, Italy.
11. Choct, M., Y. Dersjant-Li, J. McLeish, and M. Peisker. 2010. Soy Oligosaccharides and Soluble Non-starch Polysaccharides: A Review of Digestion, Nutritive and Anti-nutritive Effects in Pigs and Poultry. Asian-Australasian Journal of Animal Sciences 23: 1386-1398.
12. Chou, C. J., M. Affolter, and M. Kussmann. 2012. A nutrigenomics view of protein intake: macronutrient, bioactive peptides, and protein turnover. Progress in molecular biology and translational science 108: 51-74. doi:10.1016/B978-0-12-398397-8.00003-4
13. Collins, J. W., R. M. La Ragione, M. J. Woodward, and L. E. J. Searle. 2009. Application of Prebiotics and Probiotics in Livestock Prebiotics and Probiotics Science and Technology. p 1123-1192. Springer New York, New York, NY.
14. Cookson, W., L. Liang, G. Abecasis, M. Moffatt, and M. Lathrop. 2009. Mapping complex disease traits with global gene expression. Nature reviews Genetics 10: 184-194. doi:nrg2537 [pii] 10.1038/nrg2537 [doi]

15. Correa, R. C. G., R. M. Peralta, C. W. I. Haminiuk, G. M. Maciel, A. Bracht, and I. Ferreira. 2016. The past decade findings related with nutritional composition, bioactive molecules and biotechnological applications of Passiflora spp. (passion fruit). Trends in Food Science & Technology 58: 79-95. doi:10.1016/j.tifs.2016.10.006
16. Crhanova, M., H. Hradecka, M. Faldynova, M. Matulova, H. Havlickova, F. Sisak, and I. Rychlik. 2011. Immune response of chicken gut to natural colonization by gut microflora and to Salmonella enterica serovar enteritidis infection. Infection and immunity 79: 2755-2763. doi:IAI.01375-10 [pii] 10.1128/IAI.01375-10 [doi]
17. CVB. 2011. Chemical compositions and nutritional values of feed ingredients., Centraal Veevoeder Bureau, Lelystad, the Netherlands.
18. Delimaris, I. 2013. Adverse Effects Associated with Protein Intake above the Recommended Dietary Allowance for Adults. ISRN Nutr 2013: 126929. doi:10.5402/2013/126929 [doi]
19. Deusch, S., B. Tilocca, A. Camarinha-Silva, and J. Seifert. 2015. News in livestock research - use of Omics-technologies to study the microbiota in the gastrointestinal tract of farm animals. Computational and Structural Biotechnology Journal 13: 55-63. doi:10.1016/j.csbj.2014.12.005
20. Dibble, C. C., and B. D. Manning. 2013. Signal integration by mTORC1 coordinates nutrient input with biosynthetic output. Nat Cell Biol 15: 555-564.
21. Efeyan, A., W. C. Comb, and D. M. Sabatini. 2015. Nutrient-sensing mechanisms and pathways. Nature 517: 302-310. doi:nature14190 [pii] 10.1038/nature14190 [doi]
22. FAO. 2004. Protein sources for the animal feed industry., Food and Agriculture Organisation, Bangkok.
23. FAO. 2009. Global Agriculture Towards 2050, Food and Agricultural Organization of the United Nations, Rome, Italy.
24. FAOSTAT. 2009. FAO Statistical Databases, Food and Agriculture Organization of the United Nations, Rome, Italy.
25. Fischer, M., A. G. J. Voragen, S. R. Piersma, L. V. Kofod, C. I. Joergensen, P. Guggenbuhl, C. S. Nunes, and H. Gruppen. 2007. Presence of indigestible peptide aggregates of soybean meal in pig ileal digesta residue. Journal of the Science of Food and Agriculture 87: 2229-2238. doi:10.1002/jsfa.2948
26. Gallien, S., E. Duriez, C. Crone, M. Kellmann, T. Moehring, and B. Domon. 2012. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. Molecular & Cellular Proteomics : MCP 11: 1709-1723.
27. Ganesh, V., and N. S. Hettiarachchy. 2012. Nutriproteomics: a promising tool to link diet and diseases in nutritional research. Biochimica et biophysica acta 1824: 1107-1117. doi:S1570-9639(12)00127-6 [pii] 10.1016/j.bbapap.2012.06.006 [doi]
28. Gao, Y. Y., Z. Y. Jiang, Y. C. Lin, C. T. Zheng, G. L. Zhou, and F. Chen. 2011. Effects of spray-dried animal plasma on serous and intestinal redox status and cytokines of neonatal piglets. J Anim Sci 89: 150-157. doi:10.2527/jas.2010-2967

29. Garriga, C., R. R. Hunter, C. Amat, J. M. Planas, M. A. Mitchell, and M. Moreto. 2006. Heat stress increases apical glucose transport in the chicken jejunum. American journal of physiology Regulatory, integrative and comparative physiology 290: R195-201. doi:00393.2005 [pii] 10.1152/ajpregu.00393.2005 [doi]
30. Gevaert, B., L. Veryser, F. Verbeke, E. Wynendaele, and B. De Spiegeleer. 2016. Fish Hydrolysates: A Regulatory Perspective of Bioactive Peptides. Protein and Peptide Letters 23: 1052-1060. doi:10.2174/0929866523666161102122637
31. Gibbs, B. F., A. Zougman, R. Masse, and C. Mulligan. 2004. Production and characterization of bioactive peptides from soy hydrolysate and soy-fermented food. Food Res. Int. 37: 123-131. doi:10.1016/j.foodres.2003.09.010
32. Gupta, R. K., K. Gupta, A. Sharma, M. Das, I. A. Ansari, and P. D. Dwivedi. 2017. Health Risks and Benefits of Chickpea (Cicer arietinum) Consumption. Journal of Agricultural and Food Chemistry 65: 6-22. doi:10.1021/acs.jafc.6b02629
33. Ivarsson, E., S. Roos, H. Y. Liu, and J. E. Lindberg. 2014. Fermentable non-starch polysaccharides increases the abundance of Bacteroides-Prevotella-Porphyromonas in ileal microbial community of growing pigs. Animal : an international journal of animal bioscience 8: 1777-1787. doi:S1751731114001827 [pii] 10.1017/S1751731114001827 [doi]
34. Iwaniak, A., and J. Dziuba. 2011. BIOPEP-PBIL Tool for the Analysis of the Structure of Biologically Active Motifs Derived from Food Proteins. Food Technology and Biotechnology 49: 118-127.
35. Jansman, A. J. M. 2016. Health and functions of the gastrointestinal tract in pigs: Effects of functional ingredients and feed and ingredient processing. J Anim Sci 94: 12-21. doi:10.2527/jas2015-9886
36. Jewell, J. L., and K.-L. Guan. 2013. Nutrient Signaling to mTOR and Cell Growth. Trends in biochemical sciences 38: 233-242.
37. Jha, R., and J. F. D. Berrocoso. 2016. Dietary fiber and protein fermentation in the intestine of swine and their interactive effects on gut health and on the environment: A review. Animal Feed Science and Technology 212: 18-26. doi:10.1016/j.anifeedsci.2015.12.002
38. Kim, J., and K.-L. Guan. 2011. Amino acid signaling in TOR activation. Annu Rev Biochem 80: 1001-1032. doi:10.1146/annurev-biochem-062209-094414 [doi]
39. Knap, P. W. 2005. Breeding robust pigs. Australian Journal of Experimental Agriculture 45: 763-773.
40. Kogut, M. H., and R. J. Arsenault. 2016. Editorial: Gut Health: The New Paradigm in Food Animal Production. Frontiers in Veterinary Science 3: 71. doi:10.3389/fvets.2016.00071
41. Lindberg, J. E. 2014. Fiber effects in nutrition and gut health in pigs. Journal of Animal Science and Biotechnology 5: 15.
42. Lindmark-Mansson, H., and B. Akesson. 2000. Antioxidative factors in milk. British Journal of Nutrition 84: S103-S110.

43. Maeno, M., N. Yamamoto, and T. Takano. 1996. Identification of an antihypertensive peptide from casein hydrolysate produced by a proteinase from Lactobacillus helveticus CP790. J. Dairy Sci. 79: 1316-1321.
44. Martin, W. F., L. E. Armstrong, and N. R. Rodriguez. 2005. Dietary protein intake and renal function. Nutr Metab (Lond) 2: 25.
45. Maslowski, K. M., and C. R. Mackay. 2011. Diet, gut microbiota and immune responses. 12: 5-9.
46. Minkiewicz, P., J. Dziuba, A. Iwaniak, M. Dziuba, and M. Darewicz. 2008. BIOPEP database and other programs for processing bioactive peptide sequences. Journal of Aoac International 91: 965-980.
47. Mitroi, N., and M. Mota. 2008. Nutrigenomics/Nutrigenetics. Rom J Intern Med 46: 295-304.
48. Moller, N. P., K. E. Scholz-Ahrens, N. Roos, and J. Schrezenmeir. 2008. Bioactive peptides and proteins from foods: indication for health effects. European Journal of Nutrition 47: 171-182. doi:10.1007/s00394-008-0710-2
49. Mullally, M. M., H. Meisel, and R. J. FitzGerald. 1997. Identification of a novel angiotensin-I-converting enzyme inhibitory peptide corresponding to a tryptic fragment of bovine β-lactoglobulin. FEBS Lett. 402: 99-101. doi:10.1016/s0014-5793(96)01503-7
50. NRC. 2012. Nutrient Requirements of Swine. National Academic Press, Washington, D.C. 20418 USA.
51. Nurmi, E., L. Nuotio, and C. Schneitz. 1992. The competitive exclusion concept: development and future. International journal of food microbiology 15: 237-240.
52. Oakley, B. B., H. S. Lillehoj, M. H. Kogut, W. K. Kim, J. J. Maurer, A. Pedroso, M. D. Lee, S. R. Collett, T. J. Johnson, and N. A. Cox. 2014. The chicken gastrointestinal microbiome. FEMS Microbiol Lett 360: 100-112. doi:10.1111/1574-6968.12608 [doi]
53. Pihlanto, A. 2006. Bioactive peptides: functionality and production. Agro Food Industry Hi-Tech 17: 24-26.
54. Qin, C., K. Qiu, W. Sun, N. Jiao, X. Zhang, L. Che, H. Zhao, H. Shen, and J. Yin. 2016. A proteomic adaptation of small intestinal mucosa in response to dietary protein limitation. Scientific reports 6: 36888. doi:srep36888 [pii] 10.1038/srep36888 [doi]
55. Qu, A., J. M. Brulc, M. K. Wilson, B. F. Law, J. R. Theoret, L. A. Joens, M. E. Konkel, F. Angly, E. A. Dinsdale, R. A. Edwards, K. E. Nelson, and B. A. White. 2008. Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. PLoS One 3: e2945. doi:10.1371/journal.pone.0002945 [doi]
56. Quinteiro-Filho, W. M., A. Ribeiro, V. Ferraz-de-Paula, M. L. Pinheiro, M. Sakai, L. R. M. Sa, A. J. P. Ferreira, and J. Palermo-Neto. 2010. Heat stress impairs performance parameters, induces intestinal injury, and decreases macrophage activity in broiler chickens. Poultry science 89: 1905-1914. doi:89/9/1905 [pii] 10.3382/ps.2010-00812 [doi]
57. Quinteiro-Filho, W. M., M. V. Rodrigues, A. Ribeiro, V. Ferraz-de-Paula, M. L. Pinheiro, L. R. M. Sa, A. J. P. Ferreira, and J. Palermo-Neto. 2012. Acute heat stress impairs performance parameters and induces mild intestinal enteritis in broiler chickens: role of acute

hypothalamic-pituitary-adrenal axis activation. J Anim Sci 90: 1986-1994. doi:jas.2011-3949 [pii] 10.2527/jas.2011-3949 [doi]

58. Rauw, W. M. 2012. Immune response from a resource allocation perspective. Front Genet 3: 267. doi:10.3389/fgene.2012.00267 [doi]
59. Rauw, W. M., and L. Gomez-Raya. 2015. Genotype by environment interaction and breeding for robustness in livestock. Front Genet 6.
60. Sansonetti, P. J. 2004. War and peace at mucosal surfaces. 4: 953-964.
61. Sarbassov, D. D., S. M. Ali, and D. M. Sabatini. 2005. Growing roles for the mTOR pathway. Current opinion in cell biology 17: 596-603. doi:S0955-0674(05)00148-1 [pii] 10.1016/j.ceb.2005.09.009 [doi]
62. Shakouri, M. D., P. A. Iji, L. L. Mikkelsen, and A. J. Cowieson. 2009. Intestinal function and gut microflora of broiler chickens as influenced by cereal grains and microbial enzyme supplementation. Journal of animal physiology and animal nutrition 93: 647-658. doi:JPN852 [pii] 10.1111/j.1439-0396.2008.00852.x [doi]
63. Shi, T., E. Song, S. Nie, K. D. Rodland, T. Liu, W.-J. Qian, and R. D. Smith. 2016. Advances in targeted proteomics and applications to biomedical research. Proteomics 16: 2160-2182.
64. Shinmoto, H., S. i. Dosako, and I. Nakajima. 2014. Anti-oxidant Activity of Bovine Lactoferrin on Iron/Ascorbate Induced Lipid Peroxidation. Biosci. Biotechnol. Biochem. 56: 2079-2080. doi:10.1271/bbb.56.2079
65. Siri-Tarino, P. W., S. Chiu, N. Bergeron, and R. M. Krauss. 2015. Saturated Fats Versus Polyunsaturated Fats Versus Carbohydrates for Cardiovascular Disease Prevention and Treatment. Annual review of nutrition 35: 517-543. doi:10.1146/annurev-nutr-071714-034449
66. Sonnenburg, J. L., and F. Backhed. 2016. Diet-microbiota interactions as moderators of human metabolism. 535: 56-64.
67. Teixeira, S. R., S. M. Potter, R. Weigel, S. Hannum, J. W. Erdman, and C. M. Hasler. 2000. Effects of feeding 4 levels of soy protein for 3 and 6 wk on blood lipids and apolipoproteins in moderately hypercholesterolemic men. Am. J. Clin. Nutr. 71: 1077-1084.
68. tePas, M. F. W., O. Madsen, M. P. L. Calus, and M. A. Smits. 2017. The Importance of Endophenotypes to Evaluate the Relationship between Genotype and External Phenotype. Int. J. Mol. Sci. 18: 472. doi:doi:10.3390/ijms18020472
69. Thaiss, C. A., N. Zmora, M. Levy, and E. Elinav. 2016. The microbiome and innate immunity. 535: 65-74.
70. Tourneur, E., and C. Chassin. 2013. Neonatal Immune Adaptation of the Gut and Its Role during Infections. Clin Dev Immunol 2013.
71. UN. 2015. World population prospects: The 2015 revisions, key findings and advance tables, United Nations, Department of Economic and Social Affairs, Population Divisions.
72. Ussar, S., N. W. Griffin, O. Bezy, S. Fujisaka, S. Vienberg, S. Softic, L. Deng, L. Bry, J. I. Gordon, and C. R. Kahn. 2015. Interactions between Gut Microbiota, Host Genetics and Diet Modulate

the Predisposition to Obesity and Metabolic Syndrome. Cell metabolism 22: 516-530. doi:S1550-4131(15)00336-8 [pii] 10.1016/j.cmet.2015.07.007 [doi]

73. van Baarlen, P., J. M. Wells, and M. Kleerebezem. 2013. Regulation of intestinal homeostasis and immunity with probiotic lactobacilli. Trends Immunol 34: 208-215. doi:S1471-4906(13)00014-8 [pii] 10.1016/j.it.2013.01.005 [doi]
74. van Der Wielen, P. W., S. Biesterveld, S. Notermans, H. Hofstra, B. A. Urlings, and F. van Knapen. 2000. Role of volatile fatty acids in development of the cecal microflora in broiler chickens during growth. Applied and environmental microbiology 66: 2536-2540.
75. Wang, D., X. Wan, J. Peng, Q. Xiong, H. Niu, H. Li, J. Chai, and S. Jiang. 2017. The effects of reduced dietary protein level on amino acid transporters and mTOR signaling pathway in pigs. Biochem Biophys Res Commun. doi:S0006-291X(17)30375-3 [pii] 10.1016/j.bbrc.2017.02.084 [doi]
76. Weichhart, T., M. Hengstschlager, and M. Linke. 2015. Regulation of innate immune cell function by mTOR. Nat Rev Immunol 15: 599-614.
77. Wells, J. M., R. J. Brummer, M. Derrien, T. T. MacDonald, F. Troost, P. D. Cani, V. Theodorou, J. Dekker, A. Meheust, W. M. de Vos, A. Mercenier, A. Nauta, and C. L. Garcia-Rodenas. 2017. Homeostasis of the gut barrier and potential biomarkers. Am J Physiol Gastrointest Liver Physiol 312: G171-G193. doi:ajpgi.00048.2015 [pii] 10.1152/ajpgi.00048.2015 [doi]
78. Wienkoop, S., and C. Staudinger. 2013. Proteomics, Quantification-Unbiased and Target Approach. In: W. Dubitzky, O. Wolkenhauer, K.-H. Cho and H. Yokota (eds.) Encyclopedia of Systems Biology. p 1799-1800. Springer New York, New York, NY.
79. Yamamoto, N., A. Akino, and T. Takano. 1994. Antihypertensive Effects of Different Kinds of Fermented Milk in Spontaneously Hypertensive Rats. Biosci. Biotechnol. Biochem. 58: 776-778.
80. Yang, H., X. Xiong, X. Wang, B. Tan, T. Li, and Y. Yin. 2016. Effects of Weaning on Intestinal Upper Villus Epithelial Cells of Piglets. PLoS One 11: e0150216. doi:10.1371/journal.pone.0150216.
81. Ye, J.-L., C.-Q. Gao, X.-G. Li, C.-L. Jin, D. Wang, G. Shu, W.-C. Wang, X.-F. Kong, K. Yao, H.-C. Yan, and X.-Q. Wang. 2016. EAAT3 promotes amino acid transport and proliferation of porcine intestinal epithelial cells. Oncotarget 7: 38681-38692. doi:9583 [pii] 10.18632/oncotarget.9583 [doi]
82. Zarogoulidis, P., S. Lampaki, J. F. Turner, H. Huang, S. Kakolyris, K. Syrigos, and K. Zarogoulidis. 2014. mTOR pathway: A current, up-to-date mini-review (Review). Oncol Lett 8: 2367-2370. doi:10.3892/ol.2014.2608 [doi] ol-08-06-2367 [pii]
83. Zhang, C., M. Zhang, S. Wang, R. Han, Y. Cao, W. Hua, Y. Mao, X. Zhang, X. Pang, C. Wei, G. Zhao, Y. Chen, and L. Zhao. 2009. Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. 4: 232-241.
84. Zhu, L. H., J. X. Xu, S. W. Zhu, X. Cai, S. F. Yang, X. L. Chen, and Q. Guo. 2014. Gene expression profiling analysis reveals weaning-induced cell cycle arrest and apoptosis in the small intestine of pigs. J Anim Sci 92: 996-1006. doi:jas.2013-7551 [pii] 10.2527/jas.2013-7551 [doi]

# Appendices

## Summary

The challenge of creating a sustainable food supply for growing human population has attracted much attention in novel and economical sources of protein for animal feed. Dietary protein sources serve as one of the important constituents of animal nutrition providing strict nutritional (essential amino acids) and non-strict nutritional functional properties. However, there is a lack of assessment methods for evaluation of different functional properties of protein sources beyond their capacity to provide amino acids and other nutrients. The research objective of this thesis is to evaluate the functional properties of different protein containing feed ingredients. For evaluation we used multiple ~omics technologies (FeedOmics approach) to improve our understanding of the gut associated changes induced by the dietary protein sources, especially on the level of the intestinal microbiota and the mucosal gene expression. In addition, we also investigated the effects of protein sources on several systemic immune and metabolic parameters.

In **Chapter 2**, selected protein sources were characterised for functional properties. We aimed to characterise and quantify individual proteins present in various potential protein sources using advanced proteomic method. Subsequently, *in silico* procedures was used to predict amino acid compositions and potential bioactive properties of these protein sources. Casein (CAS), partially delactosed whey powder (DWP), spray dried porcine plasma (SDPP), soybean meal (SBM), wheat gluten meal (WGM) and yellow meal worm (YMW) were selected. By NanoLC-LTQ-Orbitrap-Mass Spectrometry (nLCMS), protein contents of the protein rich feed ingredients were detected and semi-quantified. Based on these data, the amino acid composition of the selected proteins sources were calculated and compared with the chemically determined amino acid composition. By using bioinformatics, we predicted the bioactive properties of these protein sources after *in silico* digestion with monogastric proteolytic enzymes. The analysis revealed that selected protein sources are rich in bioactive peptides, particularly in angiotensin-converting enzyme inhibitors and anti-oxidative properties. We demonstrated the added value of the use of proteomic and bioinformatic procedures to approximate the protein, peptide and amino acid composition as well as to predict the potential bioactive properties of the (novel) protein-containing feed ingredients.

Given that novel protein sources may contain bioactive components with activities beyond their strictly nutritional properties we investigated the use of intestinal organoids as a model to test the effects of different protein sources on the intestinal epithelium (**Chapter 3**). Mouse enteroids were exposed to different undigested protein sources (*viz.* soybean meal, SBM; casein, CAS; spray dried plasma protein, SDPP; and yellow meal worm, YMW) and RNA was isolated for genome-wide transcriptomics.

The different protein sources induced unique biological processes. YMW protein was predicted to down-regulate biological processes involved in lipid metabolism processes, CAS was predicted to down-regulate glutathione metabolism, SBM down-regulated biological processes involved in triglyceride, phosphatidylcholine biosynthesis as well as pathways associated with retinoid and retinol metabolism. Taken together these results highlight enteroids as a promising new model to evaluate interaction between feed/food ingredients and the intestinal epithelium.

The second part of the thesis focussed on understanding the functional properties of various protein sources in animal experiment *in vivo*. An experiment was conducted with laboratory mice fed with diets containing protein from various sources (same sources as used in **Chapter 2**). Here the aim was to understand the long term (relative to *in vitro* experiment performed in **Chapter 3**) *in vivo* effect of protein sources on intestinal functioning and health in mice. **Chapter 4** describes the effect of dietary protein sources in mice on local intestinal (immune) gene expression, microbial colonization and systemic immune responses. By genome wide transcriptome analysis, we identified fourteen high level regulatory genes that are strongly affected in SBM-fed mice compared to the other experimental groups. They are mostly related to the mammalian (mechanistic) target of rapamycin (mTOR) pathway linked to T cell activation, proliferation and cell fate along with antigen presentation by immune cells. In addition, an increased concentration of granulocyte colony-stimulating factor was observed in serum of SBM-fed mice compared to other dietary groups. Moreover, by 16S rRNA sequencing, we observed that SBM-fed mice had higher abundances of *Bacteroidales* family S24-7, compared to the other sources. We showed that measurements of genome-wide expression and microbiota composition in the mouse ileum reveal divergent responses to diets containing different protein sources, in particular for a diet based on SBM. In **Chapter 5** we studied the effect of dietary protein source on metabolic amine profiles in serum and urine of mice. We showed that the metabolism of protein within host is affected by the source of dietary protein. Dietary protein sources were found to have profound effects on host metabolism, particularly in systemic amine profiles, which is considered as considered as an endo-phenotype. Metabolites like alpha-aminobutyric acid and 1-methylhistidine are sensitive indicators of too much or too little availability of specific amino acids in the different protein diets. Furthermore, we concluded that amine metabolic profiles can be useful for assessing the nutritional quality of different protein sources.

In **Chapter 6**, a set of computational methods of systems biology was used for understanding the effect of protein sources with regard to their functional properties. We described potential correlations between molecular parameters as measures on various biological scales by integrating multi-scale quantitative (~omics) data obtained from mice experiment. For the first time, an integration of such heterogeneous data-types, arising from a single experiment, has been reported.

Final evaluation of protein sources was carried out in a feeding trial with pigs by feeding them with diets containing protein from various sources viz. SBM, black soldier fly larvae (BSF), SDPP, rapeseed meal (RSM) and WGM. Here, the aim was to understand the long term (relative to the *in vitro* experiment described in this thesis) *in vivo* effect of protein sources on intestinal functioning and health in the target animal pigs. **Chapter 7** describes the effects of dietary protein sources in pigs on both local intestinal (immune) gene expression, microbial colonization and systemic immune responses. We have shown that diets prepared with different protein sources, differ in their ability to modulate the physiology of pigs at the local as well as the systemic level. Different protein sources elicit specific responses in blood amine metabolite profiles by altering the composition and diversity of the intestinal microbiota as well as intestinal mucosal gene expression.

Finally, **Chapter 8** summarises the research undertaken in this thesis and discusses the key results, followed by describing methods to evaluate the functional properties of protein-containing feed ingredients by using (but not limited to) modern day (~omics) technologies and novel *in vitro* (intestinal models). Finally, a new potential application of using knowledge generated in the area of animal nutrition is discussed in this thesis.

This thesis presents FeedOmics approach as a toolkit, to evaluate (novel) protein containing feed ingredients of different origin considering both their nutritional and functional value in terms of their capacity to support or modify nutrient supply, the animal's physiology, tissue development and functioning. Such knowledge may contribute to introduce novel and/or alternative protein containing feed ingredients in the diet of livestock, thus creating a sustainable food supply for growing human population.

## Acknowledgement

"Arise, awake, and stop

not till the goal is reached"

**Swami Vivekananda.**

This quote continues to be the driving force as this has led me to come to this point. Everyone would certainly agree that finishing PhD project is a collective effort. Many have contributed directly or indirectly to make things happen towards finalizing this thesis. Thus, at the start, I would like to thank the 'un-named' persons whom I could not mention here. This would have been impossible without your generosity and help.

I would like to thank my promotor and supervisor, **Mari Smits**, for giving me the opportunity to join the *Investerings programma / ontwikkelings programma* (IPOP; investment programme / development programme) Customised Nutrition Project of Wageningen University & Research and work as a PhD candidate. Mari, I would like to thank you for believing in me and for all the help, continuous encouragement and guidance whenever needed. You inspired me both in professional and personal areas of life. I learnt a lot from you and from this experience of completing a PhD thesis at a foreign university. **Alfons Jansman**, my co-promotor and project leader of the IPOP Customised Nutrition project, thank you for providing me the opportunity to pursue a PhD at Wageningen University & Research. You were always there to support me, for instance on brain storming for new ideas in the field of animal nutrition, providing me with financial support whenever required by my research, and critical assessment of the manuscripts and drafts. I thoroughly enjoyed travelling with you in various conferences and learnt a lot from you. I would also like to thank my promotor, **Jerry Wells** for hosting me in his group and providing me with advanced facilities in the Host Mirobe Interactomics (HMI) lab. I want to thank you especially for your help and advice during the experiments with mice. Also, I really appreciate and thank you for giving me an opportunity to work with mice intestinal organoids. Your critical comments on the manuscripts were always fruitful. A big thanks goes to my co-promotor **Dirkjan Schokker** who was always there to support me on various bioinformatics matters, biostatistics and data interpretation and editing manuscript drafts. I am privileged to be his first PhD candidate as co-supervisor. Dirkjan, you made me feel very comfortable by being more of a buddy. I hope we continue to engage in future projects.

I would like to warmly thank all the collaborators and co-authors. Words fall short when its come to thank **Leo Kruijt** (Wageningen Livestock Research, WLR), who is also one of my paranymphs. You are a great guy and I really enjoyed working with you. Your

role was vital and I thank you for being a co-author in most of the manuscripts. You helped me to pull through the rigorous animal experiments and sample collections even at odd hours in the morning. It is not possible to jot down all the things that I have learnt from you, in particular to practical lab (bench) work. Your pragmatic and thoughtful views on multiple things (both academic and non-academic) helped me to acclimatize to this foreign country and the local work ethics. I am very grateful to **Sjef Boeren** (Laboratory of Biochemistry, WUR) for detecting peptides with nano-LC-MS. This was indeed a fruitful collaboration and hoping to engage with you more in future. I am thankful to **Mark Boekschoten**, **Jenny Jansen** of Human Nutrigenomics (WUR) and **Stephanie Vastenhouw** of Wageningen Bioveterinary Research (WBVR) for technical expertise and microarray scanning. I want to extend my gratitude to **Frank Harders** (WBVR) and **Javier Ramiro-Garcia** (Laboratory of Microbiology, WUR) for expertise and help with sequencing of gut microbiota and analysis. I would like to warmly thank **Amy C. Harms** and all other people at the Netherlands Metabolomic Centre at Leiden University for expertise and help with metabolomic analyses. Receiving two additional grants through ZonMw: Enabling Technology Hotels with you was fruitful. You were excellent collaborators. My heartfelt thanks goes to **Maria Suarez-Diez** (Systems and Synthetic Biology, WUR) for the lively discussions and brain-storming that we had during the meetings on systembiology-based manuscript. I want to thank **Vitor A.P. Martins dos Santos** (Systems and Synthetic Biology, WUR) for your inputs and useful advice on the systembiology-based manuscript. I thoroughly enjoyed the 'wetlab-drylab' collaboration.

My sincere thanks goes to the industrial partners (Nutreco and Darling Ingredients International) for showing faith and continuous support to this project. Thank you for your active participation during the steering committee meetings. **Coen Smits** (Nutreco), **Carine van Vuure, Fred Beekmans, Lourens Heres** and **Anke van Doremalen** (Darling Ingredients Internationals), your participation, comments, suggestions, feedbacks on this project helped to maintain the efficiency of the project in terms of deliverables.

A special thanks goes out to **Michiel Kleerebezem** (HMI, WUR), **Jurriaan Mes, Harry Wichers** (Food and Biobased Research, WUR), **Guido Hooiveld** (Human Nutrition, WUR), **Harry Gruppen, Peter Wierenga** (Laboratory of Food Chemistry, WUR), **Wouter Hendriks** (Animal Nutrition Group, WUR), **Huub Savelkoul**, **Joost van Neerven** (Cell Biology and Immunology, WUR), **Ruud Albers** (Nutrilead), **Annick Mercenier** (Nestle) for taking keen interest in this research and providing vital inputs and suggestions.

The work performed in this thesis could not have completed without the help of colleagues from Wageningen University & Research Animal Breeding and Genomics, Animal Nutrition of WLR, WBVR, all animal care takers and personnel of the Animal

Research Facility (DB) in Lelystad and the Research Facility in Carus, Wageningen. Especially **Gerrit Jan Deetman, Albert Ter Laak** (DB, WBVR, Lelystad), **Rob Steenmans** and **Bert Weijers** (Research Facility, Carus, WUR). I thank **Arjan Hoekman** (WLR) for placing the chemical and consumables orders in day to day basis during the ongoing animal experiments. Thank you **Piet Wikselaar** and **Ruud Dekker** (Animal Nutrition of WLR) for your help in conducting the animal experiments in Lelystad and collecting samples on dissection days which was of utmost importance. I appreciate your effort. I would also like to thank **Pieter Roskam** (WBVR) to help me with his expertise in dissection. And a special thanks goes to **Henk Schipper** (Experimental Zoology Group, WUR) for helping me with microscopy.

A very big thank you to all my (former) colleagues at HMI for the good times at work and all other activities: **Agnieszka, Anja, Avis, Bart, Berdien, Bruno, Edoardo, Ellen, Jonna, Jori, Jurgen, Juri, Linda, Loes, Marcela, Marjolein, Nadya, Nico, Nirupama, Nuning, Raka, Rogier, Stefan** and **Simon**. Thank you all for the lovely moments that I have spent with you all, it will be cherished forever. **Loes, Anja** and **Nico**, thank you for providing the much needed emotional support and you guys never let me feel away from home. **Loes**, I wish every office gets a secretary like you. You always provided solution to any problem that I stumbled upon. Your eagerness to help with a smiling face always reminded me of my mother. Indeed, just like a 'foster mother' you helped me to get integrated in the group and overall to the 'Dutch' society. **Anja** (my second paranymphs) **and Nico**, your relentless guidance, support and help (both in- and out-of-work) forms one of the foundation pillar in making of this thesis. For instance in-work, your support towards organoid work, tissue staining, participation during the sectioning days in several animal experiments are just a few to mention from the heap of favours. The ease and the comfort level that I attained in Dutch society is a reflection of what you both did for me out-of-work. I can never thank you enough! Apart from the technical competencies, I learnt a lot from both of you as a couple at a personal level. I wish you both a healthy, happy life ahead and continue to being an inspiration. **Ellen**, *bedankt* :) for helping me with the MagPix analysis and your participation during the sampling of mice experiment. It was always fun to spent time with you during coffee breaks, I specially admire you for your direct honest opinions. **Marjolein**, you were proactive and helped me with the DEC proposal of the mice experiment. I sincerely want to thank you for the humble gesture and I want to wish you all the very best with your future endeavours. **Linda**, our collaboration for the organoid paper was a success. I thoroughly enjoyed working with you. Thank you for motivating me which helped me to touch the finishing line. I will miss all the fun we had together with **Bruno, Rogier, Edo** and **Marcela**. Enjoy the new phase of life with your family and wish you all the very best for future. **Nirupama** (Niru)**,** it was really fun to share the office with you for the last 4 years. I learnt a lot from you through our collaborations which ultimately resulted in two manuscripts and many conference

proceedings. I thoroughly enjoyed the wetlab-drylab approach and hope to collaborate in future. You handled quite good with all my non-sense jokes. I must confess that you are the only colleague who senses all my 'cheap thrills' and I find it hard to apply one on you. Lately I also joined the 'Deadpool' fan club and hope we will continue to have nice conversations in the coming days. I want to wish you all the very best for your upcoming PhD defence and future endeavours. **Agnieszka** (Aga), **Nuning and Berdien**, it was great to share my office with you all. I will really miss our "interesting" chats. **Marcela, Edo, Bruno** and **Rogier** you were the fun element of HMI. If I have to jot down every instance and events that I shared with you all, then I will need to write equivalent pages as I have written for this thesis. So making a long story short, thank you for all the help that you provided me. Thank you for making me the part of the fun moments. **Jori,** I will miss the lunch sessions with you and **Simon**, I will miss your interesting presentations during the lab meetings. **Bart**, you joined our group when I was about to finish my PhD. However, in such a short time, we had quality time together and had a fruitful collaboration regarding the organoid story and thus you became one of the co-authors of Chapter 3. Apart from this, you helped me with designing the cover of this thesis! I really appreciate your contribution and want to heartily thank you for that.

**Hsuan** (Animal Nutrition Group, WUR), both of our projects belonged to the IPOP Customized Nutrition Research - line 3. In multiple events you helped me, for instance grinding meal worms, preparing DEC proposals, conducting animal experiments together, collecting samples on dissection days, feed composition analysis, presenting results to our industrial partners together and the list is never ending. Thank you for the nice time and I enjoyed working with you. **Tetske, Guido Bosch, Sergio, Sonja and Myrthe** (Animal Nutrition Group, WUR) - thank you for the nice discussion about dietary proteins, solubility, insect as alternative protein source, digestion kinetics, animal nutrition and digestive physiology. You all have enriched me with your knowledge and expertise. **Tetske, Hsuan and Sergio**, thanks for giving me a great company during the conferences we attended together. Hoping to see some of you in upcoming conferences. **Lonneke** (Human and Animal Physiology, WUR), **Geraldine** (Laboratory of Food Process Engineering, WUR), **Tetske, Hsuan** and **Sergio**, we will surely miss the buffet after IPOP meetings. I wish you all good luck with all future endeavours.

**Eva, Elsa, Olaf**- whenever I start a conversation with you all, time just flies. I will miss the endless conversation of PhD and life altogether. **Eva**, thank you for the road trip to Hungary and the hospitality that you showed. This has really helped me to get out from the stress accumulated due to the last stretch of PhD. **Elsa**, you somehow inspired me to live in the moment. Thanks a lot and hope to see you in Colombia soon. **Olaf**, you are such an amiable person. It would be great if we can collaborate in future projects. That will be awesome. Keep squashing and do count on me as well ;-) I wish all of you great success

with finishing your PhDs. **Balaji** and **Neeraj**, I started my Wageningen journey with you guys. It was great fun being your housemate and later on neighbours in Wageningen. I will cherish our cooking sessions, trips together, movies and doing crazy stuffs with both of you. Good luck with finishing your PhDs soon and then maybe we can plan a post-PhD trip somewhere! **Thibout**, I have learnt several things in life from you. Perhaps you don't even realize it. Thank you (and lets not talk about what I learnt from you! ;) :D). **Rini**, meeting you and **Zoey** is one of the best thing happened to me in Wageningen. Cheers to our friendship! A friend in need is a friend indeed- you truly justify this **Liana**. Thank you for motivating me when I needed the most. **Satyaki, Debarup, Jayanta, Neerjhar, Aman** thank you for being the continuous support. A big thanks to all my teachers- **Babul Sarmah, Probodh Borah, Soumen Naskar, Chanda Roy, Jyoti B. Dutta, Late Radha** mam and **Ranu** mam.

**Maa** (Pranati Kar) and **Baba** (Sankar Lal Kar), I find it hard to express the gratitude I feel for your effort and presence in my life. Thank you for your love, support and guide throughout my life. I will always need your blessings and faith for my future endeavours. Love you both! **Dada** (Santanu Pratim Kar), you always inspired me and gave me the push to challenge myself and extend my horizon. I cherish the endless scientific and philosophical talks we had, that also gave me a better perspective of things. Thank you for that and just being there for me always. **Papa** (Tridev Choudhury), I want to thank you for the encouragement and comfort you gave me whenever I needed it. I really value your wisdom and expect your continuous support in future. **Mamma** (Mausumi Choudhury) and **Bhai** (Apratim Choudhury), I love you the way you are. Thank you for making me a part of your life and guarding me with your love and confidence.

**Raka**, without you, it would have been very difficult to reach this point. We have made it!! Although it might sound cliché, you are the support system without whom finalising this thesis would have been very difficult. Thank you for your continuous support, believe and faith that you placed upon me. Your patience, especially during the last phase of the PhD, helped me stabilise and re-compose myself. You comforted me when it was needed the most. I hope I can provide you the same in future when you will be finishing your PhD. Cheers to the highs and lows that we spent together till now!

**Soumya Kanti Kar**

## Curriculum vitae

Soumya Kanti Kar was born in Assam, India, on the 4th of February 1986. He completed his high school from Kendriya Vidyalaya Khanapara, Guwahati, India. Thereafter, he started Bachelor in Veterinary Science & Animal Husbandry (BVSc & AH) at College of Veterinary Science Khanapara, where he obtained his BVSc & AH degree in 2010. He joined the Masters of Veterinary Science (MVSc) programme in 2010, sponsored by Department of Biotechnology, Govt. of India with studentship and financial aid for research. During his MVSc, Soumya studied the effect of dietary zinc on gene expression of lipogenic enzymes and regulatory factors in subcutaneous adipose tissue of growing pigs. In an additional project, he studied the effect of dietary zinc on the humoral immune response of piglets after classical swine fever virus vaccination at the Department of Microbiology, College of Veterinary Science, Khanapara, India, thereby completing MVSc (Animal Biotechnology) with first class (distinction) in 2012.

From 2013, Soumya started working as a PhD candidate at the Host-Microbe Interactomics group of Wageningen University and Animal Breeding and Genomics group of Wageningen Livestock Research, under the supervision of Professor Mari A. Smits and Professor Jerry M. Wells. The findings of his PhD research can be found in this thesis, titled "*FeedOmics, an approach to evaluate the functional properties of protein containing feed ingredients*". This research was part of the IPOP Customized Nutrition project of Wageningen University & Research. During his PhD tenure, in 2013 and 2015, Soumya received special research assignments by means of additional research grants from The Netherlands Organisation for Health Research and Development (ZonMw programme Enabling Technologies) which allowed him to conduct metabolomics studies with Netherlands Metabolomics Centre, Leiden University.

## Publication list based on this thesis

### *Peer reviewed scientific publications*

1. *S Kar, A Jansman, S Boeren, L Kruijt, M Smits. 2016.* Approximation of the amino acid composition and bio-functional properties of current and novel protein sources for pigs. *Journal of Animal Science*. 2016.94:30–39. doi:10.2527/jas2015-9677.
2. *N Benis*, S Kar*, V dos Santos, M Smits, D Schokker, M Suarez-Diez.* Multi-level integration of environmentally perturbed internal phenotypes reveals key points of connectivity between them. *Accepted in Frontiers in Physiology, 2017. 8:388. doi: 10.3389/fphys.2017.00388*

* Co-First authors: these authors contributed equally to the work.

### *Conference and symposia proceedings*

1. *S. K. Kar, A.J.M. Jansman , L. Kruijt , F.M. de Bree , F. Harders , E. Kuijt , A. Bossers , M. A. Smits.* Impact of diets comprising various protein containing feed ingredients on the microbiota composition in the small intestine of mice. Proceedings of NextGen Genomics and Bioinformatics Technologies (2014) conference, 17th-19th November 2014, Bangalore, India.
2. *S K Kar, A J M Jansman, S Boeren, L Kruijt, M A Smits.* Approximation of the amino acid composition and bio-functional properties of current and novel protein sources for pigs. Proceedings of the 13th Digestive Physiology of Pigs, 19-21 May, 2015, Kliczków, Poland.
3. *S K Kar, A J M Jansman, L Kruijt, E H Stolte, N Benis, D Schokker, M A Smits.* Alternative protein sources for monogastrics: composition and functional assessment. Proceedings of the 66th annual meeting of the European Federation of Animal Science (EAAP), 31st Aug to 4th of Sep, 2015, Warsaw, Poland.
4. *S K Kar, A J M Jansman, D Schokker, L Kruijt, M A Smits*. Effects of dietary protein sources on intestinal and systemic responses of pigs. Proceedings of the 67th annual meeting of the European Federation of Animal Science (EAAP), 29th Aug to 2nd of Sep, 2015, Belfast, United Kingdom.
5. *N Benis, M Suarez-Diez, D Schokker, S K Kar, V A P M dos Santos, Mari A Smits*. Integration of –omics data measured at different biological levels. Proceedings of the BioSB, 19-20th April, 2016, Lunteren, The Netherlands.
6. *N Benis, M Suarez-Diez, D Schokker, S K Kar, V A P M dos Santos, Mari A Smits*. High level data integration: Pathway analysis of murine intestinal gene expression. Proceedings of the European Conference on Computation Biology, 3-7th Sep, 2016, Den Haag, The Netherlands.
7. *S K Kar, A J M Jansman, L Kruijt, N Benis, M A Smits.* Effect of dietary proteins on immunity and metabolism in mice. Proceedings of the 5th International Symposium on Energy and Protein Metabolism and Nutrition, 12-15 September, 2016, Krakow, Poland.

8. *S K Kar, N Benis, D Schokker, J M Wells, A J M Jansman, L Kruijt, M A Smits.* Use of multi-omics approaches in the search for alternative dietary protein sources. Proceedings of the Protein for Life conference, 23-26 October, 2016, Ede, The Netherlands.
9. *M Hulst, S K Kar, C van Vuure, A van Doremalen, L Heres, A de Wit, A J M Jansman, M A Smits. In vivo* and *in vitro* study on the mode of action of spray dried plasma when used as feed additive. Proceedings of the Protein for Life conference, 23-26 October, 2016, Ede, The Netherlands.

## Training and Supervision Plan

(Completed in the fulfilment of the requirements for the education certificate of the Graduate School Wageningen Institute of Animal Sciences; One ECTS equals a study load of 28 hours).

| | |
|---|---|
| **Basic Package (3 ECTS)** | |
| WIAS Introduction Course (mandatory, 1.5 credits) | 2013 |
| Course on philosophy of science and/or ethics (mandatory, 1.5 credits) | 2013 |
| **Scientific Exposure (20.2 ECTS)** | |
| ***International conferences*** | |
| NGBT 2014: NextGen Genomics and Bioinformatics Technologies, Bangalore, India | 2014 |
| International conference on Digestive Physiology on Pigs, Kliczkow, Poland | 2015 |
| 66th EAAP annual meeting, Warsaw, Poland | 2015 |
| 67th EAAP annual meeting, Belfast, UK | 2016 |
| 5th EAAP ISEP, Krakow, Poland | 2016 |
| Protein for life, Ede, The Netherlands | 2016 |
| ***Seminars and workshops*** | |
| WIAS Science Day, Wageningen, The Netherlands | 2013-2015 |
| Bio-Plex user meeting, Grimbergen, Belgium | 2013 |
| International Workshop on Nutrition and Intestinal Microbiota, Berlin, Germany | 2013 |
| Seminar PacBio in Next Generation Sequencing, Wageningen, The Netherlands | 2014 |
| Schothorst Feed Research symposium: 23rd Intercoop Workshop on Poultry and Swine Nutrition, Ede, The Netherlands | 2015 |
| Technological advancement in food research, Maastricht, The Netherlands | 2016 |
| The health and robustness platform: TopigsNorsvin, Hamar, Norway | 2016 |
| ***Presentations*** | |
| Impact of diets comprising various protein containing feed ingredients on the microbiota composition in the small intestine of mice. Bangalore, India; **Poster** | 2014 |
| Influence of different sources of dietary protein in small intestinal microbiota of mice. Wageningen, The Netherlands; **Poster** | 2014 |
| Approximation of the amino acid composition and bio-functional properties of current and novel protein sources for pigs. Kliczkow; **Oral** | 2015 |

| | |
|---|---|
| Alternative protein sources for monogastrics: composition and functional assessment. Warsaw, Poland; **Oral** | 2015 |
| Feedomics. Ede, The Netherlands; **Oral** | 2015 |
| Effects of dietary protein sources on intestinal and systemic responses of pigs. Belfast, UK; **Oral** | 2016 |
| Effect of dietary proteins on immunity and metabolism in mice. Krakow, Poland; **Oral** | 2016 |
| Use of multi-omics approaches in the search for alternative dietary protein sources. Ede, the Netherlands; **Oral** | 2016 |
| Technological advancement in food research. Maastricht, The Netherlands; **Oral** | 2016 |
| Microbiome for health, immunity and using organoids as a tool. Hamar, Norway; **Oral** | 2016 |
| **In-Depth Studies (11.6 ECTS)** | |
| ***Disciplinary and interdisciplinary courses*** | |
| 5th International Advance course: Industrial food protein | 2013 |
| Analysis of microarray and RNA seq expression data using R/BioC and web tools, Rotterdam, The Netherlands | 2014 |
| Intestinal Microbiome and diet in human and animal health, Wageningen, The Netherlands | 2014 |
| Advanced Proteomics course, VLAG/NuGO association, Wageningen, The Netherlands | 2015 |
| Quality of protein in animal diet, Wageningen, The Netherlands | 2015 |
| Gut biology and health: GSST PhD course and COST Action FA1401 (PiGutNet) tranning school, Aarhus University, Denmark | 2015 |
| ***Advanced statistics courses*** | |
| Design of Experiment- Advance Statistics, Wageningen, The Netherlands | 2013 |
| **Statutory Courses (3 ECTS)** | |
| Use of Laboratory Animals, Utrecht University, The Netherlands | 2013 |
| **Professional Skills Support Courses (4.2 ECTS)** | |
| PhD Competence assessment, Wageningen, The Netherlands | 2013 |
| Information Literacy Including Endnote Introduction, Wageningen, The Netherlands | 2013 |
| Project and Time Management, Wageningen, The Netherlands | 2014 |
| Techniques for writing and presenting a scientific paper, Wageningen, The Netherlands | 2016 |
| The Final Touch: Writing the General Introduction and Discussion, Wageningen, The Netherlands | 2016 |

| | |
|---|---|
| **Research Skills Training (9 ECTS)** | |
| Preparing own PhD research proposal | 2013 |
| Special research assignments | |
| ZonMw Enabling Technologies: Metabolic effects of dietary proteins in mice | 2013 |
| ZonMw Enabling Technologies: Metabolic effects of dietary proteins in pigs | 2015 |
| **Didactic Skills Training (2.2 ECTS)** | |
| ***Lecturing*** | |
| Quality of protein in animal diets, Wageningen Academy, Wageningen, The Netherlands | 2015 |
| ***Supervising practicals and excursions*** | |
| HMI-50306 Microbial Disease Mechanisms | 2013 |
| HMI-50306 Microbial Disease Mechanisms | 2015 |
| **Management Skills Training (2 ECTS)** | |
| ***Organisation of seminars and courses*** | |
| WIAS Science Day, Wageningen 2014, The Netherlands | 2014 |
| **Education and Training Total** | **55.2** |

***Colophon***

The research described in this thesis is part of the IPOP Customized Nutrition project of Wageningen University & Research and was financially supported by Wageningen University & Research, the Dutch Ministry of Economic Affairs, Nutreco NV and Darling Ingredients International.

**Cover design:** Bart van der Hee, Raka Choudhury and Soumya Kanti Kar.
**Lay-out and Printing:** Digiforce.