Modelling genetic selection for resistance to gastrointestinal parasites in small ruminants

Fabrizio Assenza

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Sciences Ecologiques, Vétérinaires, Génétiques et Bioingégneries

---

Présentée et soutenue par :
Assenza Fabrizio

le : mardi 30 septembre 2014

Titre :

Modelling genetic selection for resistance to gastrointestinal parasites in small ruminants

---

Ecole doctorale :
Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :
GenPhySE 1388

Directeur(s) de Thèse :

Moreno Carole

Rapporteurs :
Conington Joanne
Arranz Juan José
Cabaret Jacques

Membre(s) du jury :

Table of contents

1. General Introduction

Living organisms are self-sustained biochemical systems capable of undergoing Darwinian evolution. Indeed, all of them contain heritable information stored in nucleic acids, which allows the system to reproduce itself and to evolve. Concerning multicellular sexual organisms, the cells that constitute the whole system can be grouped in two categories: somatic cells and germ cells. The genetic information of each organism determines how the system reacts to the environment it lives in. This process resulted in all the varieties of different species we can observe on our planet. Some of these species have found a successful evolutionary pathway resulting in parasitism, i.e. their life cycle features a necessary stage hosted within another living organism.

In the context of small ruminant industry, gastrointestinal nematodes constitute a limiting factor to production and food security. Furthermore, the most popular control strategy, anthelmintics, resulted in shifting the evolutionary pathway of gastrointestinal parasite to a population featuring increased resistance to the anthelmintics themselves. This prompted research to find alternative solutions for limiting the economic impact of gastrointestinal parasitosis in livestock production industry. One of the most promising candidates is genetic selection, which would allow to modify the genetic make up of a target population to feature a less favourable internal environment for parasite establishment. The latter, in turn, would also direct the evolutionary pathway of the parasite population towards the adaptation to the changes in the internal environment of the host. However, previous studies have shown that this influence would not suffice for the adaptation speed of the parasite population to keep up with the changes in the internal environment of its host - when the latter was target of a breeding scheme for increasing its resistance to the parasite [1].

The efficacy of genetic selection as a control strategy for limiting the economic impact of gastrointestinal parasites can be estimated by software simulation. In order to do so, it is however necessary to estimate some genetic parameters, such as: the heritabilities of the traits included in the global breeding goal, the genetic correlations among them, and the presence of marker allele loci possibly featuring a larger effect on the phenotype than the allelic substitution effect averaged over the whole genome. The latter information could be acquired thanks to the recent technological advance which allows for genotyping quickly and inexpensively with dense marker maps, for example by high density SNPchips. Indeed,

molecular information allows for reading the information written in the DNA at relatively small reading intervals. These fragmentary pieces of information are enough to result both in more precise estimates of genetic parameters and in the possibility of performing tests for significant statistical association between a DNA region and the variation of a continuous trait. This technology also had a positive impact on the accuracy and precision of genetic parameter estimates, such as coefficients of relatedness and, therefore, it also had an impact on both the estimates of additive genetic (co)variances and on the estimates of breeding values. Finally, the information conveyed by molecular markers might have a positive impact on the yearly genetic gain and might result in better parameters estimates for testing the efficacy of genetic selection as a control strategy for limiting the economic impact of gastrointestinal parasites *in silico.*

The objective of this thesis project was to estimate these genetic parameters from the phenotypes observed by following an experimental infestation protocol for the disease caused by *Hamonchu contortus* in small ruminants. Haemonchosis ranks as one of the most problematic gastrointestinal parasitosis of small ruminants which limits production worldwide [2]. The infestation protocol featured two subsequent infestations with 10000 L3 infesting larvae which allowed to observe both the unprimed immune response and the successive primed immune response, and to compare them.

The first study conducted during this thesis project resulted in estimates of the heritabilities of growth and parasite resistance traits in sheep, together with the estimates of their genetic and phenotypic correlations. Moreover, the protocol allowed to gather some clues about the possible effect of genotype by environment interaction significantly affecting growth traits when expressed across non-contaminated and contaminated environments. Finally, it also allowed to compare the precision of these estimates, obtained by a model including pedigree information only, with those obtained by the same model but including molecular marker information as well.

The second study concerned instead the estimation of the marker loci allelic substitution effects on phenotypes collected by the same infestation protocol as before but applied on creole goats – in order to test the statistical association between the marker loci of a 50kSNP chip and parasite resistance traits such as faecal egg count and packed cell volume.

The results obtained from these studies can help with drawing the guidelines which an efficient breeding plan for selecting small ruminant for parasite resistance must be based upon. The parasite resistance was measured by faecal egg counts and packed cell volume in

goat, while only on faecal egg counts in sheep. The production was measured by the average daily gain before the unprimed infestation and by the average daily gain during the unprimed infestation in sheep. The results obtained feature the estimates of the following parameters. The heritabilities were computed for faecal egg counts measured during the unprimed and primed immune responses for both sheep and goat; for packed cell volume measured during both the unprimed and primed immune responses in creole goat; for the average daily gains measured both before the unprimed infestation (in a non-contaminated environment) and during the primed immune response (contaminated environment); and for the packed cell volumes measured during both infestations in goat. Moreover, the phenotypic correlations between the traits within each study were computed in both sheep and goat, while the genetic correlations among growth and parasite resistance traits were computed in sheep only. The study on goats also featured a genome-wide association study for faecal egg counts and packed cell volumes, measured during both infestations.

The information extracted from the two observational studies might be useful for computing an educated estimate of both the achievable yearly genetic gain of a breeding scheme for parasite resistance and of the possible correlated selection responses among the traits included in the global breeding goal. Finally, these two estimates can be used for further testing the efficacy of genetic selection as a strategy for controlling gastrointestinal nematodes in the context of an epidemio-genetic model in silico, i.e. by software simulation.

2. Literature review

The aim of this chapter is twofold. The first part is a literature review focusing on both the economic impact of gastrointestinal parasites on small ruminant production and on what is known about the pathophysiology of the parasitosis. The former aims at contextualizing the phenomenon of parasitic diseases within the small ruminant production industry. The latter aims at putting together what is known about the interaction between the host and the parasite, from different disciplines. Because of the variability of economic weights depending on the particular consumer target, the estimates computed across countries will be reported separately. Estimates concerning thespecies-specific impact of *H.contortus*, when available, and the cumulative impact of multi-species gastrointestinal nematode infestation, when species-specific estimates are not available, are reported. In order to understand how gastrointestinal parasitosis results in this production loss and what is hidden behind the phenotypes, which the genetic models applied later treat as a "black-box", a literature review on what is known about the biological determinism of haemonchosis will follow. This section will feature an essential inventory of the organs, cellular populations and molecules, mainly involved in the pathophysiology of the disease. This inventory might prove helpful in interpreting the results of the genetic analyses.

The second part is a literature review about the most popular practices used for the analysis of quantitative traits applied to livestock improvement. Its aim is both to explain the statistical tools underlying genetic analyses and to illustrate how the results obtained can be used for predicting the theoretical outcome of genetic selection. Furthermore, the impact that the availability of molecular information has on the estimation of genetic parameters, as well as on the prediction of the response to selection, will be highlighted.

2.1 Helminthiasis in small ruminants production industry

2.1.1 Gastrointestinal helminths of domestic small ruminants.

Parasitism is a widely observed form of non-mutual symbiotic relationship between organisms. The parasite is an organism that depends on the host for completing its life cycle. On the other hand, the host does not depend on the parasite by any means and normally reacts to the parasite in order to clear it from its body. From an evolutionary point of view, when a parasite population is able to overcome the clearance reaction of a host, it becomes endemic and a path of co-evolution between the two organisms eventually begins [3]. However, in the context of small ruminant production systems, parasitism, as well as any other disease that alters the production performance, is considered a problem because it creates a significant economic loss. The mechanisms by which infectious diseases cause an economic loss are varied and mainly depend on the disturbances caused by parasites on the utilization of the nutrients fed to the host (FIG.1).

As to the helminths of the gastrointestinal tract of the domestic small ruminants, they experience their reproductive stage as adults within the intestinal lumen. The helminths of veterinary importance infesting the intestinal tract of domestic small ruminants are numerous and they are localised in different tracts of it. The helminths of the oesophagus and of the omasum are: *Cotylophoron spp, Gonylonema pulchrum,* and *Paramphistomum spp.* The helminths of the abomasum are: *Haemonchus contortus, Marshagallia marshalli, Teledostertagia circumcinta, Ostartagia trifurcata, Parabonema spp.,* and *Trichostrongulus axei.* The helmints of the small intestine are: *Avitellina centripunctata, Bunostomum trigonocephalum, Cooperia curticei, Cooperia surnabada, Gaigeria pachyscelis, Moniezia expansa, Nematodirus battus, Nematodirus filicollis, Nematodirus spathiger, Strongyloides papillosus, Trichostrongylus capricola* and *Trichostrongylus vitirinus.* The helmints of the large intestine are: *Chabertia ovina, Oesophagostomum columbianum, Oesophagostomum venulosum, Skjabinema ovis, Trichuris ovis* and *Trichuris skrjabini.* However, most of the research in control strategies for containing the impact of these worms on small ruminant production is focused only on the most problematic nematodes, because of both their worldwide prevalence and of the extent of the economic loss they cause. These species are *H.contortus, Teladostertagia circumcinta* and *Trichostrongylus spp,* concerning the abomasum, and *Trichostrongylus spp.,* concerning the small intestine [4]. *H.contortus* in particular is reported as being one of the most problematic because of its worldwide

endemicity, the copious haemorrhage it causes and its marked resistance to anthelmintics [2, 5–7]. These features make it a good model organism for the study of gastrointestinal nematodiasis.

2.1.2 Economic impact of gastrointestinal nematodes on small ruminant industry

Gastrointestinal nematodes of small ruminants are raising growing concern across small ruminant production systems because of several reasons: the extent of the economic loss they cause is significant, their prevalence is extending outside tropical regions, and the growing number of parasites populations express resistance to anthelmintics.

The economic loss has been estimated on the order of millions of dollars per year in many countries. For example, the impact of gastrointestinal nematodes on the Australian sheep production system has been estimated to reach 1 billion dollar per year [8]; the cost of parasite control in New Zealand has been estimated as 29.3 million per year [9]; similar pictures appear in studies focused in Asia [10, 11]. The economic impact of gastrointesninal parasites is becoming relevant also in regions where its prevalence is not as high as in the tropical regions, such as Sweden, Netherlands, Denmark [12], France (Hoste et al., 2002) and UK [14]

Due to the occurrence of free-living stages during their life cycle, which ensures transmission between hosts, these parasites are exposed to different environmental conditions. This feature resulted in a picture of the endemicity of these parasites which located them mostly within warm and humid environments, i.e. subtropical and tropical environments [5]. However, due to both their marked ability of adaptation [7], and to the current climate change, these parasites are recently adapting to temperate regions up to the neighbourhood of the polar circle. Indeed, the current picture of the prevalence of infectious diseases is very likely to change due to the current climate change [15, 16].

Until recently, small ruminant's gastrointestinal parasites have been successfully controlled by the use of anthelmintics. However, many studies report an increasing resistance to these drugs among different populations of gastrointestinal parasites worldwide. This phenomenon further increases the economic loss due to gastrointestinal parasitism [14, 17–20]. Since most of the anthelmintics target single proteins, they are inevitably bound to lose efficacy because of the evolutionary potential of the parasites and their genetic variability. Resistance to anthelmintics, as well as the influence of the growing public concern for the use of drugs in food, production systems [21] have created a need for new strategies for controlling these parasites.

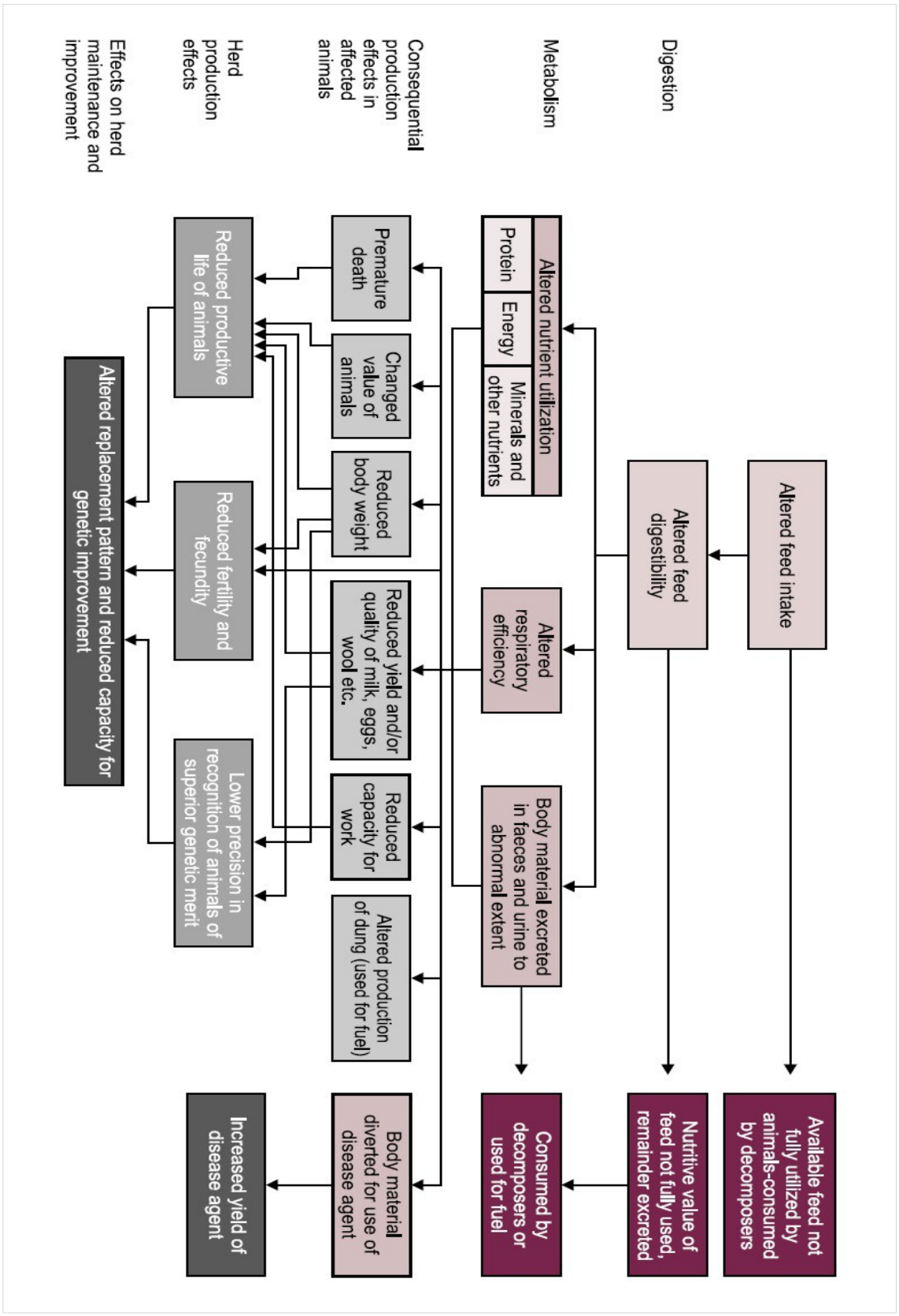


Figure 1: Impact of parasitosis on livestock production systems [5]

2.2 Pathophysiology of haemonchosis

In order to understand what are the known biological determinants of abomasal nematodiais and what is their influence on the alteration of the productive performance of the parasitized animal, this chapter aims at briefly describing the organs and the cellular populations mainly involved during the interaction between the abomasal parasites and their host. An inventory of the known biological mediators playing a major role in the regulation of this interaction will also emerge.

2.2.1 Morphofunctional description of *Haemonchus contortus*

H.contortus is the pathogenic agent of Haemonchosis. It belongs to the phylum Nematoda, which includes worms featuring the following characteristics. Nematoda are commonly known as round worms because their body has a cylindrical shape thinning at the extremities. The body is covered with a transparent cuticle, secreted by the hypoderm, which can also form various structures depending on the species. The cuticle of *H.contortus* features two cervical papillae, which fulfil both sensory and mechanic functions. The hypoderm deepens within the muscular tissue below to enclose the two excretory grooves along the sides and both the dorsal and ventral nerves inside the respective cords. The innermost membrane is composed of muscular cells, which form the celomatic cavity filled with fluid. The celomatic cavity contains the filamentous organs of the digestive and reproductive systems. The digestive system is formed by the oral cavity, the oesophagus, the intestine and the anus. The oral cavity of *H.contortus* features a lancet, which enhances the haemorrhage from the blood vessels of the host. The reproductive system of the female is formed by the ovary, the uterus and the vulva. The reproductive system of the male is formed by one testicle, and a ductus deferens which ends into the cloaca. The vulva of *H.contortus* presents a vulvar flap, which facilitates fecundation together with the spicules and the asymmetric bursa of the male's reproductive system (Fig.2, Fig 3, [6]. A large inventory of excretory/secretory products (ESP) is also involved in the interaction between the parasite and the host. Most of them have been characterized as proteases [22], whereas others have been hypothesized to play a crucial role in the regulation of the host response but have not been identified yet.

2.2.2 The interactions between the host and the parasite

2.2.2.1 Life cycle of *H.contortus*

This paragraph depicts the life cycle of *H.contortus*, which explains how the parasite and the host come into contact for their interaction to occur. The life cycle of *H.contortus* is similar to the direct life cycle of Trichostrongylidae and includes stages outside the host's digestive lumen and inside of it. The stages outside the host's digestive tract are: the egg, and the free-living larval stages. The egg is deposed by the adult females reproducing in the host's abomasum. After having been excreted together with the faeces, it develops to an L1 larva. L1 feeds on the bacteria encountered within the faeces and develops to L2. During the L1 and L2 stages, the individual stores energy which allows the development to the infesting L3 larva, as well as its survival, because the protective enclosing the L3 does not allow it to feed anymore. The timing of the whole development process and the percentage of egg successfully developing to L3 depends on the environmental conditions it occurs in [23–26]. The L3 larva migration patterns are driven both by passive and active transport; however, negatively geotropic patterns are only explained by active migration [27]. This increases the likelihood to be ingested by a potential host. After having been ingested by a suitable host, the L3 turns to L4: the L3 cuticle is shred and the intestine is developed, allowing for the beginning of the histotrophy within the abomasal mucosa. At this point, the cycle can either continue to the development of the adult stage or undergo a phase of hypobiosis. Hypobiosis is a state of arrested development which keeps the L4 to an early stage of development. The factors governing hypobiosis include both the environmental conditions experienced by the free-living stages and the environment occurring inside the host's digestive tract [28–30]. Once the L4 develops to an adult, it establishes in the abomasum and feeds by disrupting the blood vessels of the abomasal mucosa by its lancet. When the adults are sexually mature, they mate for producing eggs and the cycle begins again.

2.2.2.2 Immune response associated with resistance to *H.contortus*

The immune response of the host is due to many cellular populations derived both from myeloid stem cells and lymphoid stem cells. These cells differentiate into several subpopulations, fulfilling different roles when activated by the contact with antigens of non-self organism. The immune response is classified into two main types: the innate immune response, which refers to the reaction of the organism to any antigen recognised as non-self, and the acquired immune response, which refers to the reaction of the organism aimed at

previously encountered antigens. The former mostly involves myeloid-derived cells, whereas the latter is mostly due to lymphoid-derived cells. The immune response is regulated by a vast inventory of mediators, produced by a variety of cellular populations, which drive the type of immune cells recruited against the antigen, the healing of damaged tissues and build the "memory" of the acquired immune response. In general, the role of the effector cells is to cause damage to the neighbouring cells and the role of the mediators molecules is to maximise the localization of the activity of the effector cells to any organism expressing non-self antigens. Furthermore, the mediators molecules (cytokines) also regulate the mobility and the activation of some sub-populations of effector cells and stimulate the tissue repair pathways. Indeed, especially in the case of parasites which have co-evolved together with their host, not all immune response mechanisms have an effect on controlling the parasite population [31]. Some of the effective mechanisms which have been associated to resistance to *Haemonchus contortus* will be discussed in the following section.

2.2.2.2.1 Innate immune response

The innate immune response is mainly activated by one class of antigens referred to as pathogen-associated molecular patterns (PAMPs), which include molecules featuring high steric redundancy across various pathogens [32]. The very first defence line inhibiting the establishment of larvae is the mucous secreted by the surface mucous cells of the abomasum. The mucus reduces larval motility by mechanical impedance and by factors such as leukotriens [33, 34], secreted by mast cells and globule leukocytes. Mast cells also release histamine, which both increases peristalsis for mechanically clearing the parasites and initiates the inflammation process. Gastric secretions and motility result in reduced establishment and reduced fecundity of the worms [35].

The innate immune response is basically the inflammation which any tissue builds when damaged, which results in increased vascular permeability and increased blood circulation in the neighbourhood of the damage. In consequence, the concentration of circulating molecules, among which the complement has an effect on the larvae of *H.contortus*, increases in the surrounding of the larvae. The complement is a complex of many polypeptides which bind together by a cascade of covalent bounds. This cascade can, in turn, be triggered by one of the complement's polypeptides binding either to the antibodies coating the invading organism's surface (classical activation pathway) or directly to the carbohydrate structures of invading organisms (alternate activation pathway). The activation of the complement results in direct

cell membrane damage and release of citokines. *H.contortus* causes the activation of the complement via the alternate way, which results in the generation of chemokines such as C3a and C5a which attracts eosinophils to the sorroundings of the larvae [36–38]. Eosinophils, together with neutrophils, are also attracted by the ESP products of *H.contortus.*

The effector cells of the innate immune response are scattered across the connective tissues of the body and those mainly involved in the innate response to nematodes are: eosinophils, basophils, mast cells and antinflammatory macrophages. The mucosal mast cells release several cytokines, among which IL-13, IL-4 and IL-5 result in the following consequences: increasing intestinal motility, acting as a mechanical defense for the expulsion of the parasite. Mucosal mast cells are also involved in the regulation of the IEC cyokines release and when they are activated to infiltrating mast cells, they recruit and act together with eosinophils and neutrophils by releasing the content of their cytosolic granules [39]. The cytosolic granules contain several compounds. Reactive oxigene species, which inflicts direct oxidative damage to the parasites cuticle. Histamine, which increases intestinal motility and vascular permeability. Proteases, such as MCP-1 reducing fecundity of the adults and enhancing the activation of other effector cells. Chemokines, which recruit circulating basophils and eosinophils [40].

Other cell types involved in the innate immune response are the γδ T cells, by secretion of IL-5 and IL-13 [41]. In addition to this contribution to the innate immunity, the cytokines of the IEC, mast cells, DC and natural killer (NK) also contribute to steering the develompent of the acquired immune response towards a type 1 hypersensitivity response, by means of activating the T helper 2 lymphocytes (Th2) and B cells [42].

2.2.2.2.2 Acquired immune response

The acquired immune response is mediated by receptors, coded by MCH genes, which bind to specific antigens of the invading host and allow the lymphoid cells to focus their activity on the pathogen. These receptors can be both expressed on the surface of the lymphoid cells (cell mediated immune response), and released from their surface (antibody-mediated immune response) [43]. An effective specific immune response is activated against helminths when the MHC class II receptor of an antigen presenting cell, DC2 being specially effective, contacts a $CD4^+$ T cell receptor. IL-4 secreted by DC2 cells, together with IL-1 secreted by macrophages, also contribute to the activation of Th2 cells. The activated Th2 cells secrete a number of interleukines which contribute to building up an antibody-mediated humoral immune response

by inducing the B cells to shift into antibody producing producing plasma cells in the lymph nodes [31, 32]. The antibodies produced by the plasma cells are of different types and functions. As far as the response to helminths is concerned, immunoglobulins G (IgG) are mostly associated to reduced worm burden [44–46]; the circulating isoform of IgA has been associated to reduced worm growth and fecundity, whereas its mucosal and fecal form has been associated to reduced worm burden and ESP [45, 47]; IgE allow the localized degranulation of basophils and eosinophils to the worm surface by its affinity to th Fc receptor of these cells and have been associated with reduced worm burden but possibly also with immune- mediated tissue damage [48, 49]. DC cells also produce IL-10 and TGF-β, which induce T cells to differentiate into regulatory T cells (Treg), involved in the regulation of the immune response and the inflammation in the gastrointestinal tract [50].

2.2.2.3 Pathogenesis

2.2.2.3.1 The anatomy of the abomasum

The abomasum of small ruminants is the organ of the digestive system most similar, both for its morpholgy and for its physiology, to the stomach of monogastric species. Its main function concerning digestive process is the proteolysis. It is a luminal organ composed of four main layers: the tunica serosa, the tunica muscularis, the tunica submucosa and the tunica mucosa. Parasitism concerns mainly the tunica mucosa, the innermost of them, where the histotrophic phase of the L4 and the haematophagic phase of the adults occur. The tunica mucosa is composed of three further layers: the lamina muscularis, the lamina propria and the lamina epithelialis. The lamina muscularis is composed of smooth muscle tissue. The lamina propria is composed of connective tissue which contains the blood vessels which the adults feed from. The lamina epithelialis is composed of heterogeneous populations of cells, which determine the functional subsetting of the tunica mucosa in two different regions: the fundic region and the pyloric region. These regions are identified according to the cellular populations found inside the gastric glands, the latter formed by the introversions of the lamina epithelialis deep within the lamina propria. In fact, the cells of the lamina epithelialis found outside these glands, which form the luminal surface of the abomasum, are mostly surface mucous cells and do not differ much across the three regions. These cells produce the mucous covering the luminal surface of the abomasum, which is mostly composed of mucin (MUC5AC) and forms the so-called mucosal barrier, protecting the mucosa itself from the gastric juice.

The pyloric is region located close to the pylorus and is characterised by the presence of the

pyloric glands. These type of glands further fulfill a regulatory function by producing both endocrine and paracrine mediators. The characteristic cells of the pyloric glands are the G cells (gastrin produing), the D cells (somatostatin producing) and the enterochromaffine cells (atrial natriuretic peptide producing).

The fundic region covers most of the remainder luminal surface and is characterised by the presence of the oxyntic glands, which produce the gastric juice. The cells found in the oxyntic glands are: the parietal cells (hydrocloric acid producing), the chief cells (pepsinogen producing), the enterochromaffine-like cells (histamine producing), the D cell and the enterochrome affine cells.

The neurons of the enteric nervous system (ENS) also play a key role in regulating the abomasum's functions. These neurons mediate the stimuli of the vagus in order to transit them to the cells of the lamina epithelialis through the release of the following neurotransmitters: acetylcholine (ACh), gastrin-releasing peptide (GRP), vasoactive intestinal polypeptide (VIP), nitric oxide and substance P. The sensory calcitonin gene-related neurons also contribute to the regulation of the gastric functions [51, 52].

2.2.2.3.2 The physiology of the abomasum and its regulation.

Digestion of proteins occurs within the abomasum due to the secretion of the proteolytic enzymes diluted in the gastric juice, which functions best at the low pH^+ achieved by the simultaneous secretion of hydrocloric acid. Low pH^+ also enhances the absorption of iron, calcium, vitamin b-12 and, by digesting the microorganisms of the prestomachs, allows the ruminants to use them as a protein source rather than letting them colonize the intestine. The mucosal barrier normally protects the abomasal mucosa from the proteolytic action of the gastric juice; however, disturbances in the regulation of gastric secretions can result in mucosal damage and impair the digestive function of the abomasum [53].

The main agonists of acid secretion are: histamine (produced by the ECL cell), gastrin (produced by the G cells) and ACh (produced by the ENS). These mediators influence the parietal cells by binding to its H_2, CCK_2 and M_3 receptors, respectively, and stimulate the activity of the parietal cell's proton pump (H^+ K^+ ATPase) via different signal transduction pathways. The histamine acts through the adenilate cyclase, the others act by inducing the release of intracellular Ca^+. The main antagonist is the somatostatin, produced by the D cells, which supposedly acts through the somatostatin type 2 receptor that activates the G protein-mediated transduction pathway [52]. The basal state of the abomasum is dominated by the

somatostatin released by the D cells, which inhibits the parietal cells directly and indirectly via both the ECL and the G cells, resulting in low activity of the proton pump. The shift to the activated state is initialized mainly by neurocrine impulses from the colinergic neurons, which both stimulate the parietal cells directly and remove the inhibition of somatostatin by inhibiting the activity of D cells. The inhibition of D cells results in the removal of both their inhibitory effects on the histamine release from the ECL cells and on the gastrin release from the G cells, which, in turn, stimulate the parietal cells. Histamine, together with gastrin, further inhibits the activity of the D cell and it also stimulates the parietal cell to activate its proton pump [54]. The basal interdigestive state is restored by several feedback mechanisms: the stimulatory effect of gastrin on antral D cells [55], mediated by gastrin-related peptide from the ENS; the colinergic neurons gradually reducing their input along with the increased distension of the abomasal walls; the lack of the anticipation stimulus; the reduction of proteins as the meal flows through the pylorus. Distension also induces the ENS neurons to release vasoactive intestinal peptide (VIP), which stimulates D cells [56]. The buffering activity of the meal decreases with time, exposing the ENS sensory neurons to perceive lower pH^+ and resulting in stimulation of D cells [57]. Finally, the amylin released by D cells further enhances the release of somatostatin [58].

2.2.2.3.3 Symptoms of haemonchosis

The symptoms of Haemonchosis are generally similar to those of other abomasal infestations by parasites featuring high haematophagy, they include: anorexia, anaemia, hypergasrtinaemia and hyperpepsinogenaemia. Chronic haemonchosis can also result in oedema (typically submandibular), due to persistent alteration of the osmotic pressure of the blood. At the anatomopathological examination, the abomasal mucosa shows the occurrence of adult worms, ulcerations and typical signs of inflammation. The bone marrow shows erythroid shift. At microscopic level, L4 can also be found and the cellular populations of the abomasal mucosa appear modified [59, 60]. The severity of the symptoms depends both on the host's response to the parasite, on the parasite itself, and on environmental factors.

A reduction of feed intake is observed in many infectious diseases. Anorexia is a behavioural trait which is induced by many mechanisms [61, 62]. A major role is possibly played by the immunity mediators, such as interleukin-1 (IL-1) [63] and tumour necrosis factor alpha (TNF-α) [64], secreted by activated monocytes [43], which have been found to depress appetite in mice and humans. Reduced gut motility also influences the appetite of the host because of the

increased stimulation of the tension receptors in the abomasal walls [65]. The reduction in gut motility is seen as a consequence of the parasite's excretory/secretory products, such as cholinesterases [60]. Anorexia is also induced by the increase in haematic concentrations of gastrin, secretin and cholecystokinin [65].

Anaemia is essentially due to the direct damage inflicted by the adult worms to the blood vessels, which results in copious haemorrhage because of the anticoagulants secreted by the parasite itself [22]. Each adult parasite can drain up to 0.05 ml of blood per day and, given that the number of adults feasting on one host's abomasal wall is normally in the order of thousands, the total blood loss can reach dangerous levels. Massive infestations can indeed result in sudden death of the host. The haemorrhage stimulates erythropoiesis quickly and, depending on the nutritional status of the host, the balance between the production of new herytrocytes and the blood loss can result in different levels of anaemia. Modifications of the bone marrow can also be observed accordingly with an increase of the erythroid tissue over the myeloid tissue.[66].

Achlorhydria is the increase of pH^+ within the abomasum observed during many abomasal nematodiais. The mechanisms leading to achlorhydria can be both direct and mediated by excretory/secretory compounds. The direct mechanism is based essentially on the damage made to the H^+ secreting parietal cells of the fundic region, where the adults *H.contortus* are most likely to establish [67, 68]. The indirect mechanism is mediated both by the excretory/secretory compounds of the nematode and by the neurocrine, endocrine and paracrine mediators released by the host. The former could inhibit the proton pump (H^+, K^+-ATPase) of the parietal cells directly or indirectly through the effect of the mediators released by the host in response the presence of the parasite. The parasite's secretagogues can actively stimulate the enterochromaffine-like cells to release histamine, which is also released by the host's mast cells as a response to the parasite's antigens. Other mediators, such as tissue growth factor alpha and the epidermal growth factor, also inhibit the parietal cell [52]. Achlorhydria is also considered to be related to the hypergastrinaemia observed during abomasal nematodiasis, especially during *Ostertagia spp* infestations.

Hypergastrinaemia is the increase of gastrin concentration in blood. Gastrin is an endocrine peptide produced by the G-cells which is largely involved both in the normal digestive function and in the cellular development of the abomasum [69]. Hypergastrinaemia is explained as a consequence of the increased abomasal pH^+, which remove the acid-related inhibition of gastrin secretion. Other factors contributing to hypergastrinaemia are ESP

secreted by the adults [70], TNF-α [71, 72], histamine [73] and IL-1β [72].

Before the basification of the abomasum lumen and the increase of gastrin in plasma, an increase in pepsinogen is normally observed in paratitised animals; however, it can also occur in the absence of hypergastrinaemia [74]. It has been related to different causes, such as: increased leakage trough the tight junctions of regenerating tissues [75]; inhibition of its conversion to pepsin, due to increased abmasal pH^+ [76], secretion of leukotriens [77]; increased secretion by the chief cells [70].

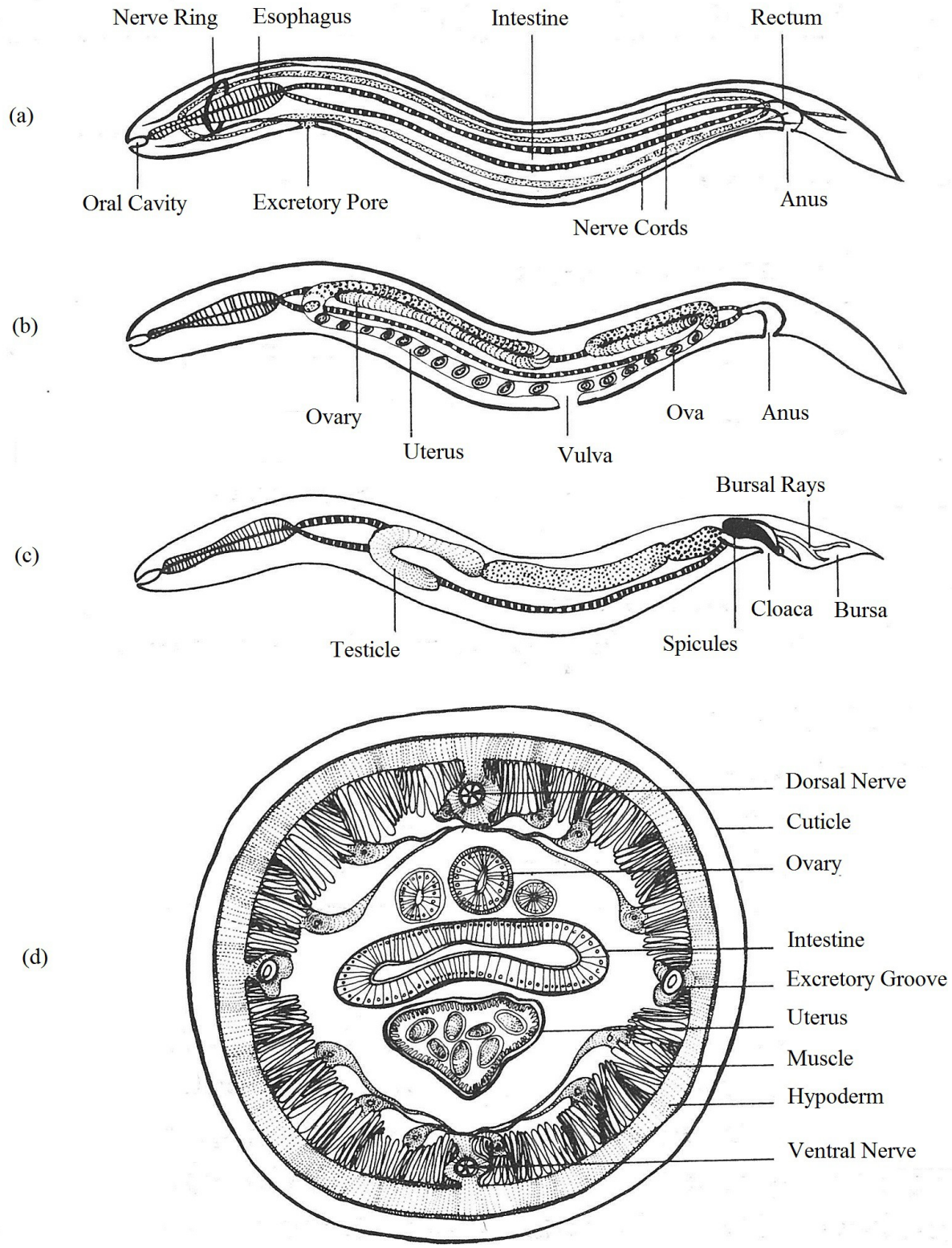


Figure 2: Schematic sections of a nematode. Longitudinal sections: (a) Digestive system; (b) Female reproductive system; (c) Male reproductive system. (d) Transversal section of a female [6]

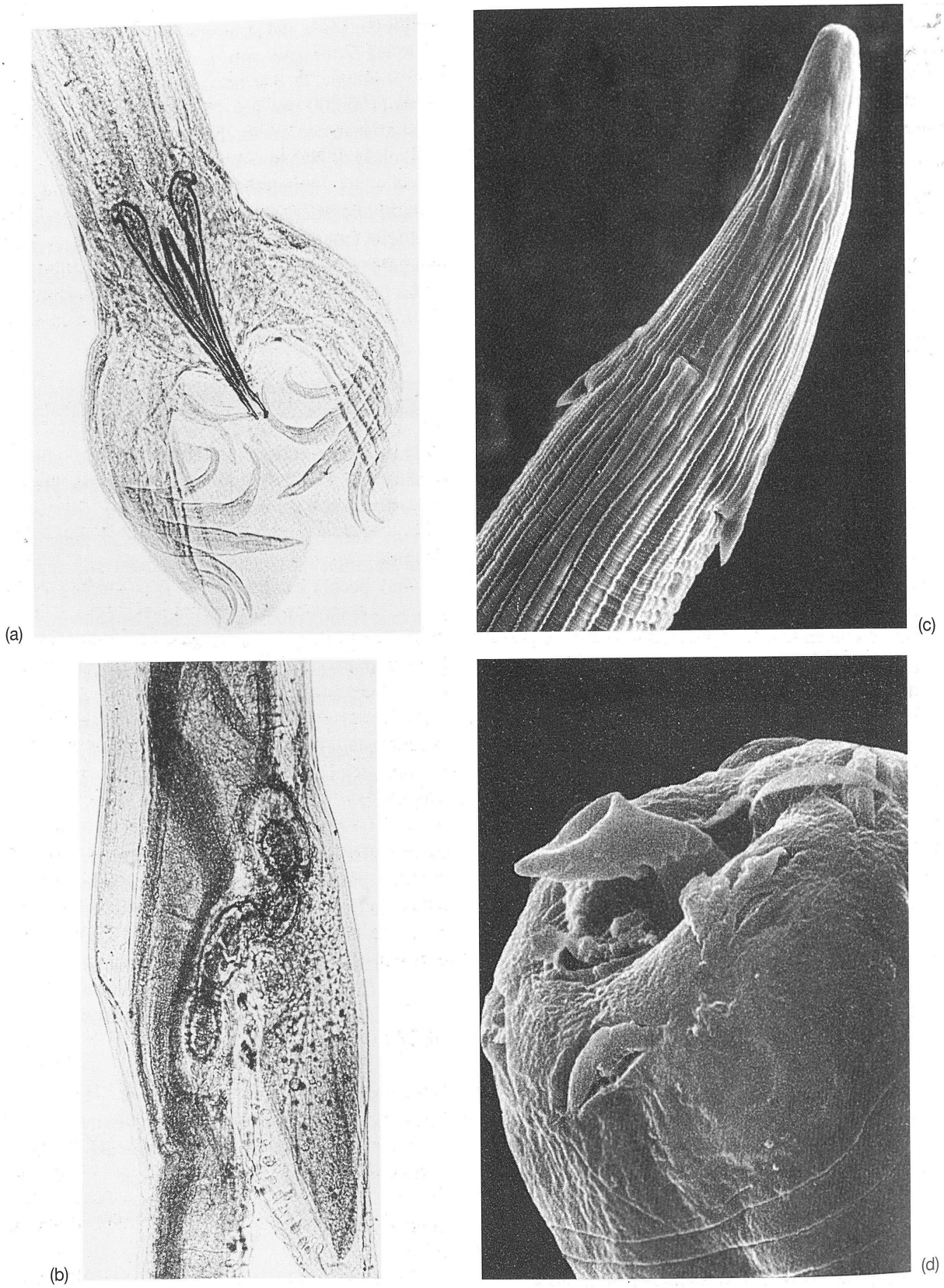
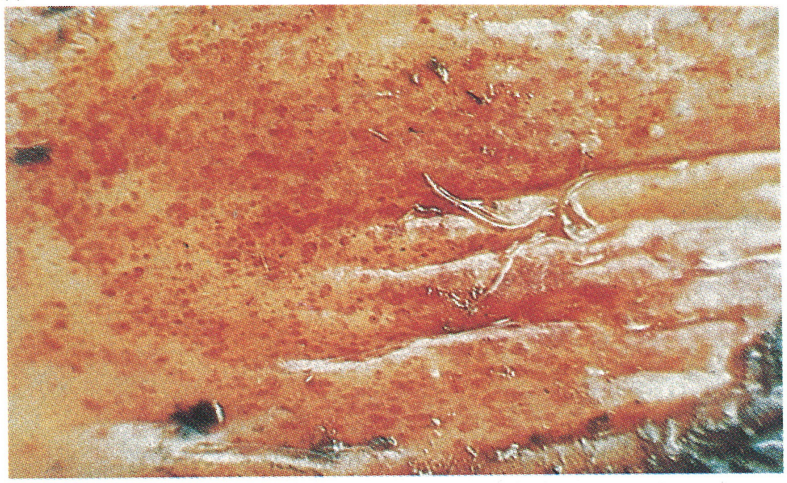


Figure 3: Microscopy of *H.contortus.* (a) Spicules and bursa of a male; (b) Vulvar flap of a female; (c) Cervical papillae (EM); (d) Lancet (EM). [6]

(a)



(b)

(c)

Figure 4: Haemonchosis a) Adults of *H.contortus* on the mucosa of the abomasum and mucosal ulcerations. b) Erithroid shift of the bone marrow.

c) Anaemia and sub mandibular oedema. [6]

2.3 Control strategies

Gastrointestinal parasitic diseases have been traditionally controlled in small ruminants industry by the use of anthelmintics. However, this has led to the selection of parasite populations which are much less affected by these drugs than back when their efficacy was assessed to be satisfactory. The problem arises because no anthelmintic can reach 100% efficacy in practice [6], namely: the individuals of the parasitic population bearing those alleles whose effect results in a higher likelihood to be killed by the anthelmintic will most likely die and the frequency of those alleles will decrease accordingly; by contrast, the individuals bearing those alleles whose effect results in a lower likelihood to be killed by the anthelmintics will most likely survive and contribute to the allelic pool for the following generations. This results in a gradual loss of the alleles which composed the genetic pool of the parasite population used for assessing the efficacy of the anthelmintic. Simultaneously, the alleles, once rare, which grant resistance of parasite population to the anthelmintic itself became common [78, 79].

These factors have risen the need to find alternative control strategies. Focusing on different factors affecting the parasite's fitness, researchers have proposed many alternatives to anthelmintics: targeted treatments, vaccines, farm management, biological control and genetic selection. However, most of them are still at an experimental stage and are not very widespread in the production context.

Targeted treatment proposes to limit the treatment to the heavily parasitised individuals only, which is achieved by following the FAMACHA method. This method proposes to measure the individual's extent of parasitism by measuring its anaemia level at the conjunctiva, based on a colour scale. Another targeted treatment strategy proposed to introduce refugia within the flock, i.e. a group of animals which receives no treatment. This follows the observed distribution of natural parasitosis as normally these parasites tend to heavily infest a few individuals in the flock while the others – only mildly [80]. By creating a refugia, this control strategy aims at reducing the selective pressure towards anthelmintic resistance on the parasite population. However, these strategies would require total compliance of farmers and veterinarians with accepting some production loss for the sake of reducing the occurrence of anthelmintic resistance, which is hardly achievable.

The development of vaccines is based on enhancing the host's specific immune response before it comes into contact with the real parasite. Since the development of a fully active specific immune response takes long time, the vaccination aims at stimulating the production

of memory B cells against the parasite's antigens before the host comes into contact with the real parasite. This is achieved by inoculating. Despite the fact that this strategy has shown its efficacy against very virulent pathogens, the same efficacy is difficult to be obtained against parasites because these organisms have evolved their own efficient strategies in order to escape the host immune response and rather co-evolve with their hosts toward reciprocal tolerance [81].

Many options of farm management have been proposed, mostly aiming at reducing the larval contamination of the pasture. Rotational grazing is applicable where the surface available for the animals is larger than the maximal stocking rate desired. It consists in splitting the surface available into at least two sectors and using them alternatively in order to keep the animals on the surface which features the smallest larval challenge. The larval challenge of the free surface can be reduced by ploughing, for example [82]. Mixed farming has also been proposed as a control strategy for reducing the larval challenge, specially in extensive farming. This strategy is applicable only when the parasitic populations living on the pasture feature mutually exclusive host specificity, i.e they don't feature common hosts. When this condition occurs, the grazing of one host will clear the pasture from the other host larvae, and vice versa [83].

Biological control aims at introducing the natural predators of the parasite's larvae into the pasture. This practice has been experimented in laboratory environment using the micelium *Duddingtonia flagrans*. However, it has not spread in production systems because the effective dose is not easily achievable in production conditions [84, 85].

2.3.1 Focus on genetic selection

Genetic selection appears one of the most promising candidates for controlling gastrointestinal nematodiasis in livestock. Genetic selection consists in setting up mating schemes aimed at modifying the genetic make-up of a population in order to shift the population's average of some phenotypes towards the desired direction. Currently, two main selection strategies have been used: selection for resilience and selection for resistance. The former aims at obtaining a population of individuals whose productive performance is not affected significantly by the presence of the parasite [86]. The latter aims at modifying the host's internal environment in order to make it less favourable for the parasite's proliferation [87].

Some arguments have been risen against the feasibility of controlling gastrointestinal

nematodes by genetic selection. They point to three main issues: the long-term reduction of the available genetic variation due to genetic selection, the possibility that the evolutionary potential of the parasite population can overcome the genetic response of the host population, and, finally, that some correlated responses to selection could increase the susceptibility of the host to other diseases or that they could slow down the genetic progress on other economically important traits [88]. The issue of long-term reduction of genetic variability concerns genetic selection in general and was already put forward by Fisher in 1930 together with his genetic theory of natural selection [89]. It is certainly true that continuous genetic selection on any trait will gradually fix the alleles involved in the determinism of the trait under selection and, hence, reduce the genetic variation within the target population. However, measuring the genetic variation underlying traits related to fitness, such as resistance to infectious diseases, which are an example of traits species have been naturally selected for since their origin, do not result in the absence of genetic variation [90]. The occurrence of genetic variation traits related to infectious disease resistance is first of all proven by the existence of different breeds featuring resistance to the same disease [31, 91], the existence of genetic variation for disease resistance within breed [92] and, finally, by the detection of QTLs for disease resistance at different locations [90, 93]. The explanations for this discrepancy between theory and observations can be several. Fisher's theory of natural selection refers to the behaviour of a system at equilibrium, which is not the case in reality because of its variation both in space and in time. Variation in space is due to the geographical distribution of the same species across different types of environment, whereas variation in time is due to the persistence of the same species in time and, therefore, to its exposure to the changes of the environment along time. An evidence of the latter can be found in the documented co-evolution between hosts and pathogens, which has also been reported to characterize the interaction between hosts and gastrointestinal nematodes [3]. The variation in time and space would cause the target alleles of selection to change and thus impede the fixation of all alleles affecting the phenotype under selection. Another possible explanation has been proposed by Ridley [94], whose theory has also been proven *in silico* [95], based on the concept of frequency-dependent selection. This theory is based on the fact that as the frequency of the resistant genotype increases, due to either natural or artificial selection, the prevalence of the pathogen decreases accordingly because its fitness to the host population is gradually lost. The selection pressure represented by the parasite decreases consequently until it becomes too weak to fix the alleles. Finally, genetic variation could be maintained due to the linkage disequilibrium of

the alleles for disease resistance to alleles of other traits [96].

The phenomenon of genetic correlation with other traits is also put forward as an issue possibly undermining the efficiency of genetic selection for resistance to gastrointestinal nematodes. Previous studies have indeed reported positive genetic correlation between parasite resistance traits, such as faecal egg count, and production traits, such as growth [97, 98]. This suggests that a trade-off between resistance and production might exist [99]. A mechanism of allocation of resources is likely to cause a competition for resources between production traits and immunity. However, such a mechanism would result in a trade-off only if not enough resources were available to fulfil the total demand of nutrients [100]. Coop and Kiriazakis have indeed found that proteins' availability can have a strong impact on the extent of production loss caused by gastrointestinal nematodes in sheep [101–103]. Furthermore, appropriate breeding plans can tackle the occurrence of unfavourable genetic correlations between two traits of interest by splitting the breeding nucleus into different genetic lines and achieve the desired genetic progress on both traits anyway, although this strategy normally implies smaller genetic gain [104]. Another way to handle unfavourable genetic correlation between parasite resistance traits and production traits would be to breed for resilience rather than to breed for resistance. The former breeding objective results indeed in lower genetic pressure on the alleles which would cause undesired selection responses and maximise the genetic gain for production traits in the context of an environment featuring pathogen challenge [86]. This could eventually result in the establishment, or even increase the prevalence, of the pathogen in the environment. However, as long as the control strategy concerns a pathogen such as *H. contortus* and other gastrointestinal nematodes causing major economic loss in small ruminants production system, this wouldn't make genetic selection for resilience fail because these pathogens do not cause zoonoses, nor do they appear among the list of pathogens to be treated by extermination of the livestock. Undesired selection responses could also occur with regard to the susceptibility to other diseases, based on the evidence that resistance to gastrointestinal nematodes are mostly associated with humoral immune response [31, 105, 106]. The experimental evidence suggests, on the contrary, that susceptibility to one disease causes the animal to be more susceptible to superinfection, as well as that the resistance to one disease is normally related to a more effective general immune protection [107]. The last concern about the effectiveness of genetic selection as a control strategy is based on the possibility that the evolutionary potential of the parasite population could overcome the genetic progress of the host. Evidence against this argument can be found both

in natural populations, for example the Red Maasai sheep in east Africa [108], and in experimental trials, such as in the studies by Kemper et Al. and by Woolaston et Al. where they reported no ability of both *H. contortus* and *T. colubriformis* to overcome the genetic gain obtained in experimental sheep populations selected for parasite resistance [1, 109].

Breeding plans are based on the estimation of the breeding value of the individuals belonging to the population under selection according to the information on both their genetics (genealogical and/or genomic) and their phenotypic performance. The treatment of genetic information will be covered in the next chapter, while the issues concerning phenotypic information will be illustrated in this paragraph. The ideal phenotype for computing a breeding value would feature the following characteristics: it is highly correlated with the breeding goal, it is associated with high heritability, it is easy to be collected *in vivo*, its measurement can be automated and it is unexpensive. The phenotype featuring the highest correlation to the breeding objective for parasite resistance would be the worm burden, i.e. the number of adult worms established in the abomasum; however, the observation of it is not feasible in practice because it can only be measured by autopsy. An indirect measure of the worm burden can be found in the feacal egg count, which is related both to the worm burden [110] and to the pasture contamination. While the collection of the faecal samples is relatively easy to perform, the count of the number of eggs per gram of faeces is quite and does not allow automation. Furthermore, faecal egg count can be influenced by drenching with anthelmintics, the consistency of faeces, the cycles of egg excretion, the gastrointestinal transit time of the ingesta and selective grazing, which lowers its repeatability. This results indeed in low repeatability [111]. Nevertheless, faecal egg count remains one of the most popular phenotypes used to measure resistance to parasites. Other phenotypes related to parasite resistance can be found among both the mediators and the cellular populations involved in the immune response. The collection of the biological samples necessary for the measurement of the immunological phenotypes can be more or less easy, depending on the specific case. Considering that for a phenotype to be useful for breeding purposes in practice, its biological sample needs to be easily collected in vivo. The phenotypes related to immunity useful for selecting for parasite resistance are those measurable in blood samples and in saliva, such as: humoral and salivary immunoglobulins, peripheral eosinophil count. Some research has been focused on exploring the feasibility of using immunity-related phenotypes for breeding gastrointestinal nematodes-resistant sheep [112]. These phenotypes feature some advantages over faecal egg count: their measurement in the biological sample is more prone to

automation and their heritability estimates are higher than those of faecal egg count. However, while the immunoglobulin concentration in blood have been associated with parasite resistance in natural population [113], their effectiveness as a phenotype to base a breeding plan upon has been tested only *in silico [114]*. Both faecal egg count and the phenotypes related to immune response feature estimated genetic correlations with production traits, such as meat and wool production, which suggest the existence of a trade-off between disease resistance and production [99]. Although it seems that the competition for resources between disease resistance and production can be influenced by nutrition [103], the eventual limitation to genetic gain posed by this unfavourable genetic correlation can also be tackled by selection for resilience rather than for resistance. By definition, selection for resilience is based on computing the breeding values according to the production traits of interest from the phenotypes measured on individuals which are exposed to the environmental constraint, including its endemic pathogens. Selection based on such a breeding goal is useful for obtaining livestock populations specialized for coping with the constraints of some special local environment. In order to drive the genetic make-up of the population to a special resilience to gastrointestinal nematode, particularly to highly haematophagous nematodes such *H.contortus*, it is feasible to include packed cell volume in the breeding goal. This phenotype is measured on blood samples and its measurement is likely to be automated. Packed cell volume measures the volume of red blood cells relative to the total volume of the sample, which measures both how many parasites are escaping the immune response mounted by the host and how effective is the host's compensation to the anaemia caused by them. It has indeed negative correlation with faecal egg count and positive correlation with growth traits [115].

The evidence that natural selection, i.e. selective pressure for fitness, results in increasing the frequency of the alleles which endow the populations established in a certain environment with the resistance to its endemic pathogens, is broadly documented in the literature [88, 90, 92]**.** However, these local livestock breeds are also reported not to feature a productive performance as high as the breeds obtained by artificial selection do, even when measured in the same conditions as the latter [116]**.** Nevertheless, the performance obtained by artificial selection is partly due to the fact that these breeds have been selected in controlled environments, where impact of infectious diseases on their production performance, including gastrointestinal nematodiasis, has been limited by the use of chemicals, until the selective pressure applied by these chemicals on their target pathogen populations resulted in their loss

of efficacy [14, 18]. The fact that within the same livestock species, such as sheep or goat, genetic selection can lead to opposite results suggests on one hand that sufficient evolutionary potential is available for genetic selection, as proven by the heritability estimates of both production and disease resistance traits [110, 117], and on the other hand-that productive performance and disease resistance might not co-exist within the same breed. However, it is worth noticing that the natural selective pressure which the disease resistant breeds were selected by did not put as much weight on production as the artificial selective pressure did. As well as the latter did not put much weight on disease resistance because the breeding goals have been designed for selecting within an environment where pathogens could be controlled efficiently by other means. The estimates of the genetic correlations between production traits and parasite resistance traits reported in literature vary from negative to positive and they are likely to be influenced by the environment which the observation used for their estimation have been collected from [90]. Furthermore, appropriate breeding plans have already proven their power to mix in the same breed traits featuring unfavourable genetic correlations between them, such as the creation of the Romane breed, which features both enhanced growth and reproductive performance [118]. Genetic selection for parasite resistance within breeds previously selected for production only has been experimented and the results obtained seem encouraging. The studies by Woolaston and Kemper in Romney sheep [1, 109] is specially interesting as it shows that sheep selected for resistance to gastrointestinal nematodes feature an internal environment which the parasite could not adapt to. Recently the availability of affordable dense molecular markers, such as SNPs, has made it possible to consider marker assisted selection as well as possible strategy for enhancing genetic selection in the context of production systems. The existence of genetic variance have been confirmed further by recent studies on resistance to gastrointestinal parasites based on molecular information [90, 93]. The genetic resistance to gastrointestinal parasites can be linked to the following specific mechanisms. Differential expression of genes involved in blood coagulation, tissue repair, gastrointestinal motility, abomasal immune cell migration have been reported. Additive genetic variation have been observed in the alleles of the ovine MHC. Differential expression of the genes coding for Th2/Th1 immune response mediators, with the interleukins promoting a Th2 response being more expressed in resistant sheep confirmed previous in vitro studies. Real time PCR detected differential expression of genes coding for tight junctions proteins, proteases, enzymes involved in reactive oxygen species production, PAMPs recognition receptors (TLR2 and 3), cytokines (such as IL1β, IFN-γ, TGF-β, IL-10,

IL-8), tissue repair genes, Treg cell marker CD25 [31, 39, 106].

2.4 Quantitative genetics applied to livestock improvement

2.4.1 Genetic variation

The computation of the breeding values used for setting up breeding plans is commonly performed by applying various formulations of the linear mixed animal model derived from the original model model proposed by Fisher in 1918 [119]. This is also known as *infinitesimal model* because it is based on the assumption that the influence of the genotype on any phenotype observed on an individual can be modelled as the cumulative effect of a theoretically infinite number of alleles, each of which features an infinitely small effect on the phenotype. This model represented the first breakthrough in the partitioning of the observed phenotypic variance into its genetic and environmental components: $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$. Where σ_p^2 is the phenotypic variance, σ_g^2 is the genetic variance, σ_e^2 is the environmental variance and the covariance between the genetic and environmental components is assumed to be null.

Broadly speaking, the genetic component would include both additive and epistatic components. However, it is common practice in the application of this model to account for the former component only. There are several reasons behind this practical choice. First of all, the progress obtained by genetic selection, based on additive genetic components only, did result in sufficiently large livestock improvement to satisfy the demand of animal products, since its first implementation after World War Two. Furthermore, since the largest stakeholder of the benefit deriving from the progress in the genetics of animal's quantitative traits is the livestock improvement industry, i.e. private breeding companies or national consortia. Their aim is not really to explain the genetic determinism of quantitative traits but to increase their income for the former and to increase the national food security for the latter.

Both objectives can be achieved by setting up breeding plans based on the estimated breeding values of livestock, the efficiency of which can be predicted by simulation based on evolutionary genetics theory [104, 120] Breeding plans allow to control the gene flow across successive generations of the target population, which determines which alleles persist over time in the gene pool and which alleles do not. This results in a change of the allele frequencies observed over time, which ultimately determines the genetic progress of the population. The additive component of the genetic variation represents an estimate of the impact of this genetic progress on the average phenotype of the population. On the contrary,

the other components of the genetic progress estimate the impact due to epistatic effects, which are related to the particular combination of the alleles coming together in the gamete of the same individual and also on the particular combination of the two gametes coming together to form the zygote from two individuals. Therefore, their effect is lost over generations because these particular combinations are broken down by recombination and Mendelian sampling. However, it is possible to exploit it outside of the reproducers' nuclei for the production of commercial hybrids, which are used for production purposes only [121].

The most popular method for computing the breeding values and the additive component of genetic variation from real data is to solve an appropriate linear mixed model [122] by Best Linear Unbiased Prediction (BLUP) [123], using the genetic variance obtained by Restricted Maximum Likelihood (REML) [124] In order to obtain an unbiased estimate of the additive genetic variation, it is necessary to account for the gene flow occurring in the population under study [125, 126]. This is achieved by including in the variance structure of the linear mixed model a genetic relationship matrix containing estimations of the relatedness between the individuals of the population, the phenotypes of which were recorded, and their inbreeding coefficients [127]. However, since the advent of inexpensive genotyping technology, a great deal of research in quantitative genetics has been dedicated to including molecular information into the computation of genetic parameters, which resulted in the development of new methods. [128–131]

2.4.2 Definitions of inbreeding and relatedness

Relatedness and inbreeding estimators are necessary for a number of applications spanning across different fields of research, such as conservation genetics, phylogeny and quantitative genetics. The specific questions that each discipline aims at answering resulted in the development of various coefficients, which feature different properties according to what the questions to be answered define. As far as quantitative genetics is concerned, these estimators are used in practice to set up the genetic relationship matrix among the individuals of the population under study, which is used for the estimation of various genetic parameters, such as genetic variances, breeding values and allelic substitution effects. The off-diagonal elements of the genetic relationship matrix used in quantitative genetics are the estimates of the relatedness coefficients between individuals, whereas the diagonal elements contain the estimates of the relatedness of each individual with itself, i.e; the inbreeding coefficient.

These quantities were first defined by Wright [132] at the beginning of the last century, when

the only information available on the genetics of animals was the population's pedigree. According to the original definition, the inbreeding coefficient of an individual is the correlation between the homologous alleles borne by the gametes coming together to form that individual, relative to the total array of homologous gametes obtained by random derivatives of the reference population. The coefficient of relatedness between two individuals was defined as the inbreeding coefficient of a hypothetical offspring between the two individuals. Later on Malécot [133] proposed to compute these coefficients as probabilities based on the concept of identity by descent (IBD), which is the state of two alleles that are identical and inherited by a common ancestor. Namely, the inbreeding coefficient is defined by Malécot as the probability that two homologous alleles within an individual are identical by descent, whereas the coefficient of relatedness is defined as the probability that the homologous alleles of two individuals are identical by descent.

It is worth noticing that both approaches require the definition of a reference population whose individuals are assumed to be unrelated. This necessity comes from merely practical issues, because theoretically speaking all individuals derived from a common ancestor are related. This means that in fact all individuals belonging to the same species are related and their genealogy could be traced back at least as far as when the speciation of their common ancestor arose along the phylogenetic tree. When relatedness is estimated for genetic studies on real data, it is common practice to truncate the genealogy of the population under study to the oldest known ancestor according to pedigree information. This implies that these estimates are somewhat arbitrary, because the distance in time of the reference population from the individuals which these parameters are estimated for have indeed an influence on the estimates themselves. As a matter of fact these estimators measure the information available on the relatedness rather than the true relatedness, as Jacquard pointed out in 1974 [134]. Indeed, the estimators of inbreeding and relatedness based on pedigree information only cannot account neither for mutation, nor for recombination events because the pedigree does not contain this information. Therefore, the values obtained from these estimators should rather be interpreted as an average over all loci of the genome because they are not locus-specific. Hence, all the estimates for individuals sharing the same genealogy, i.e. full sibs, will be the same.

Advances in genotyping technology allowed to observe the allelic configuration of animals at some marker loci. When two alleles, either at a marker locus within an individual or between individuals, are identical, they are defined identical by state (IBS), with no reference to the

genealogical origin of the alleles. Both IBD and IBS describe the identity state between the alleles at a certain locus and apply both to the alleles of the DNA of a non-haploid individual and to the alleles born by two individuals at the same locus. The difference between the IBD state and the IBS state resides in the definition of the identity: two alleles are defined identical by descent if they are identical and have been inherited by a common ancestor (neglecting mutation); whereas they are defined identical by state if they carry the same molecular information. Therefore, it is always true that two IBD alleles are also IBS (if mutation does not occur), but two IBS alleles are not necessarily IBD. Furthermore, the information on the IBS state conveyed by molecular markers is locus-specific. Indeed, the coefficients of inbreeding and relatedness based on molecular information allow to account for both mutation and recombination events. However, at the current state of the genotyping technology, the information obtained by molecular markers does not cover all the loci of any genome but is limited to some marker loci. Despite the fact that the density of the marker loci has increased quite rapidly in recent years [135], the optimal strategy to obtain the most accurate estimates of relatedness is to compute coefficients which allow to combine the molecular information with the genealogical information in order fill the gaps due to incomplete genotyping [136, 137].

2.4.2.1 Computing the genetic relationship matrix from pedigree information only

Different methods for computing the genetic relationship matrix based on pedigree information have been proposed. Together with his definitions of inbreeding and relatedness coefficients, Wright computed it by path analysis [132]. However, when it comes to practical application, the pedigree normally includes a very large number of animals and complex genealogical paths, which makes the computation by path analysis infeasible. Furthermore, common practices in quantitative genetics such as BLUP require the inversion of the genetic relationship matrix, which is a task whose computation burden increases exponentially with the dimension of the matrix to be inverted. It is necessary for the method to be useful in practice to allow its translation into efficient machine language.

In 1975, Henderson proposed a method allowing to obtain the inverse of the genetic relationship matrix directly [138]. This method is based on a recursive process which allows it to be implemented in a very fast and memory-efficient computer code. It is based on the definition of the coefficient of relatedness proposed by Wright [132], which implies that the coefficient of relatedness between two individuals equals the inbreeding coefficient of a hypothetical offspring between them and on the assumption that in the reference population

all individuals are assumed to be non-inbred and unrelated. This results in all the individuals of the base population to have inbreeding coefficient equal to 1.

This method requires all the animals to be sorted such as the parents precede their offspring, then a matrix **L** can be defined as the matrix featuring **LL`**=**A**. **A** is the genetic relationship matrix, with diagonal elements a_{ii} equal the inbreeding coefficient of each individual and off-diagonal elements a_{ij} equal the coefficient of relatedness between individuals i and j. Let the individuals 1 to $t<n$, where n id the total number of individuals, be the unrelated and non-inbred individuals belonging to the defined base population with unknown parents. Then the upper left t^2 submatrix of **L** equals an identity matrix **I**. Now, let p & q, with $p<q$, be the parents of individual i, with $q<i$-1. If both parents of i are known, then the elements of **L** are:

$$l_{ij}=(l_{pj}+l_{qj})/2 \text{ for } j=1 \text{ to } p;$$

$$l_{ij}=l_{qj}/2 \text{ for } j=\text{p}+1 \text{ to } q;$$

$$l_{ij}=0 \text{ for } j=q+1 \text{ to } i-1;$$

$$l_{ii}=\sqrt{1+0.5\sum_{j=1}^{p} l_{pj}l_{qj}-\sum_{j=1}^{q} l_{ij}^2}$$

If only one parent is known, for example p, then:

$$l_{ij}=l_{pj}/2 \text{ for } j=1 \text{ to } p;$$

$$l_{ij}=0 \text{ for } j=p+1 \text{ to } i-1;$$

$$l_{ii}=\sqrt{1-\sum_{j=1}^{n} l_{ij}^2}$$

Finally, if both parents are unknown, then: $l_{ij}=0$ and $l_{ii}=0$.

In order to obtain the $\mathbf{A}^{-1}$ the vector **d** is also required, with elements $d_i=1/l_{ii}^2$. The diagonal elements of $\mathbf{A}^{-1}$ equal $a^{ii}=d_i+0.25\sum_{k=1}^{k} d_k$, where k refers to the progeny of the ith sire. The off diagonal elements are $a^{ij}=-0.5\text{d}_j+0.25\sum_{k=1}^{k} d_k$, if j is a progeny of i or $a^{ij}=0.25\sum_{k=1}^{k} d_k$, if i and j are the parent of any progeny. This method have been implemented in a variety of software dedicated to genetic analysis, such as ASReml [139], AIReml [140]. Obtaining the $\mathbf{A}^{-1}$ matrix without having to invert it by the conventional algebra also avoids the problem eventually posed by a non positive definite **A** matrix.

2.4.2.2 Computing the genetic relationship matrix from marker information only

The computation of inbreeding and relatedness estimators based on marker information only, concerns mostly the study of natural populations, whose pedigree is very often unknown. In this case, using molecular information is the only option for estimating relationships between individuals [141]. The research in quantitative genetics applied to livestock improvement focuses rather on developing methods for enhancing the precision of pedigree-based coefficients by including molecular information into the genetic relationship matrix. However, all these methods require the genetic relationship matrix based on marker information to be set up first and then mingled with **A**.

The genomic relationship matrix is commonly computed following the methods proposed by VanRaden [142]. These methods are based on a *n* by *m* **M** matrix, where *n* is the number of individuals and *m* is the number of marker loci. The elements of **M** code for the genotype of each individual at each marker locus and can assume the following values: m_{nm}=-1 if individual *n* is homozygous for one allele at marker locus *m*; m_{nm}=0 if individual *n* is heterozygous at marker loci *m*; m_{nm}=1 if individual *n* is homozygous for the other allele at marker locus *m.* Let **P** be the *n* by *m* matrix containing the vectors of allele frequencies expressed as twice deviation of the second allele's frequency from 0.5. Such that column *m* of **P** = $2(p_m - 0.5)$, where p_m is frequency of allele 2 at locus *m* in the reference base population. Then matrix **Z** is computed as **Z** = **M-P**, which results in centering the elements of **M** to 0. The **Z** matrix can be used for computing the genomic relationship matrix **G** by different methods. One method computes **G** as

$$\mathbf{G} = \frac{\mathbf{ZZ`}}{2\sum_{m=1}^{m} p_m(1-p_m)}$$

The elements of this **G** matrix can be transformed into their homologous as defined by Wright [132] by subtracting 1 to the diagonal elements of **G (**in order to obtain the inbreeding coefficient) and by dividing the off-diagonal elements of **G** by the square roots of their respective diagonal elements (in order to obtain the relatedness coefficients between individuals).

The other method for computing **G** from **Z** was developed for human genetics [143] and computes the **G** matrix as $\mathbf{G} = \mathbf{ZDZ`}$, where **D** is diagonal and contains the reciprocal of the marker's expected variance. In studies based on real data the allele frequencies in the base population p_m are often unknown and they are estimated from population which underwent

selection. This introduces a bias in the coefficient estimates, because they should be referred to the allele frequencies in the base population [132], which results in a bias of other parameters derived from the genetic relationship matrix. Nevertheless, it has been shown that the impact of this bias on fundamental parameters used in quantitative genetics, such as breeding values, can be reduced by applying a correction derived from information available on gene flow through generations [144].

There exists another method for computing **G** which does not require the knowledge of the allele frequencies [145]. It requires only the **A** matrix and the matrix **MM`** to be computed as $\mathbf{MM'} = g_0 \mathbf{11'} + g_1 \mathbf{A} + \mathbf{E}$, where **E** is the matrix containing the differences between the expected and the observed proportions of shared DNA. The solution for g_0 and g_1 are obtained by solving the model as:

$$\begin{bmatrix} n^2 & \sum_j \sum_k A_{jk} \\ \sum_j \sum_k A_{jk} & \sum_j \sum_k A_{jk}^2 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \end{bmatrix} = \begin{bmatrix} \sum_j \sum_k (MM`)_{jk} \\ \sum_j \sum_k (MM`)_{jk} A_{jk} \end{bmatrix}$$

The **G** matrix is then obtained as $\mathbf{G} = \frac{\mathbf{MM`} - g_0(\mathbf{11`})}{g_1}$. The **G** matrices obtained by these methods are likely to result non positive definite, which limits their application to most of quantitative genetics studies. Therefore it is common practice to weight them by the coefficients of the **A** matrix, such that $\mathbf{G}_w = w\mathbf{G} + (1-w)\mathbf{A}$ [142], with w endowing the genomic relationship matrix with inversibility. Furthermore, the **G** matrix can be used for computation of genomic breeding values following multi-step procedures [146] or non parametric models, such as reproducing kernel Hilbert spaces regression [130].

2.4.2.3 Computing the genetic relationship matrix from both pedigree and marker information.

Modern practices in quantitative genetics aim at including both genealogy and molecular information at marker loci for maximizing the accuracy of genetic selection. Legarra et al [147] proposed a method for obtaining a genetic relationship matrix including both pedigree and marker information. Let the subscript 1 denote non-genotyped animals and the subscript 2 denote the genotyped ones, then the genetic relationship matrix can be partitioned as $\mathbf{A} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{bmatrix}$ as well as its inverse $\mathbf{A^{-1}} = \begin{bmatrix} \mathbf{A^{11}} & \mathbf{A^{12}} \\ \mathbf{A^{21}} & \mathbf{A^{22}} \end{bmatrix}$. Then from the conditional

distribution of the breeding values of the non-genotyped animals ($\mathbf{u_1}$) conditioned by the one of the genotyped animals ($\mathbf{u_2}$)

$$p(\mathbf{u_1}|\mathbf{u_2}) = N(\mathbf{A_{12}A_{22}^{-1}u_2}, \mathbf{A_{11}} - \mathbf{A_{12}A_{22}^{-1}A_{21}})$$

it follows that

$\mathbf{u_1} = E(\mathbf{u_1}|\mathbf{u_2}) + \epsilon = \mathbf{A_{12}A_{22}^{-1}u_2} + \epsilon$, with $\mathrm{Var}(\epsilon) = \mathbf{A_{11}} - \mathbf{A_{12}A_{22}^{-1}A_{21}} = (\mathbf{A^{11}})^{-1}$.

Recalling the **Z** matrix of the individual's marker genotypes by VanRaden [142], centred by the allele frequencies, $\mathbf{u_2}$ can be expressed as $\mathbf{u_2} = \mathbf{Za}$, where **a** is the vector of the marker effects the vector of the genotyped animal's breeding values and $\mathrm{Var}(\mathbf{u_2}) = \mathbf{ZZ`}/k = \mathbf{G}$,

where $k = 2\sum_{m=1}^{m} p_m(1-p_m)$. Therefore $\mathbf{u_1} = \mathbf{A_{12}A_{22}^{-1}Za} + \epsilon$ and

$$\mathrm{Var}(\mathbf{u_1}) = \mathbf{A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21}} + \mathbf{A_{11}} - \mathbf{A_{12}A_{22}^{-1}A_{21}} = \mathbf{A_{11}} + \mathbf{A_{12}A_{22}^{-1}(G - A_{22})A_{22}^{-1}A_{21}}$$

Given that $\mathrm{Cov}(\mathbf{u_1}, \mathbf{u_2}) = \mathbf{A_{12}A_{22}^{-1}G}$ the **H** matrix of genetic relationship including pedigree and genomic information can be written as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H_{11}} & \mathbf{H_{12}} \\ \mathbf{H_{21}} & \mathbf{H_{22}} \end{bmatrix} = \begin{bmatrix} \mathbf{A_{11}} + \mathbf{A_{12}A_{22}^{-1}(G - A_{22})A_{22}^{-1}A_{21}} & \mathbf{A_{12}A_{22}^{-1}G} \\ \mathbf{GA_{22}^{-1}A_{12}} & \mathbf{G} \end{bmatrix}$$

By using the equivalence $\mathbf{A_{12}A_{22}^{-1}} = -(\mathbf{A^{11}})^{-1}\mathbf{A^{12}}$ it is possible to avoid the inversion of the submatrices of **A** and compute their inverse directly, following Henderson [138].

2.4.3 Estimation of the additive genetic variance

The additive genetic variance is a measure of the genetic variability in the population under study, i.e. of the allelic polymorphism present in its genetic pool, and it plays a fundamental role in several practices such as: the prediction of the response to genetic selection, the estimation of breeding values and the estimation of heritability [104, 120]. According to the genetic model applied for its estimation, its value is obtained by different procedures.

In the most classical, yet most common, infinitesimal model it is normally computed by restricted maximum likelihood [124]. The infinitesimal model can be expressed as $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$, where **y** is a vector of observed phenotypes, **b** is the vector of estimated fixed environmental effects and **X** its incidence matrix, **a** is the vector of the estimated random effects and **Z** its incidence matrix, finally **e** is the vector of residuals. The expectations are assumed to equal E(**y**)=**Xb** and E(**a**)=E(**e**)=0; the variances are assumed to equal

$\mathrm{Var}(\mathbf{e})=\mathbf{I}\sigma_e^2=\mathbf{R}$, $\mathrm{Var}(\mathbf{a})=\mathbf{A}\sigma_a^2=\mathbf{G}$ and $\mathrm{Var}(\mathbf{y})=\mathbf{ZGZ`}+\mathbf{R}$, because it is also assumed that the covariance between **a** and **e** is null. Where **A** is a genetic relationship matrix. Note that **G** here is not the genomic relationship matrix but it's the genetic variance (covariance, for multi-traits models) matrix. The prediction of the **a** effects is commonly performed by BLUP [148], which requires estimates of σ_a^2 and σ_e^2 .

The same author proposed a restricted maximum likelihood estimation based on log-likelihood, the term "restricted" comes from the fact that the likelihood function is maximized for the random parameters only. The idea behind maximum likelihood procedures is to define the conditional distribution of the model's outcome given the parameters of the model and search for the combination of parameter that most closely approaches the outcome of the model to the observed data, which corresponds to maximizing the likelihood or the log-likelihood function of the model. The log likelihood function for the linear mixed animal model defined above reads as follows:

$L\alpha(1/2)\{-(\mathbf{y}-\mathbf{Xb})`\mathbf{V^{-1}}(\mathbf{y}-\mathbf{Xb})-\mathrm{logdet}(\mathbf{V})-\mathrm{logdet}(\mathbf{X`V^{-1}X})\}$, where $\alpha=\sigma_e^2/\sigma_a^2$ [149]. The first term of is a weighted sum of squares of the residuals, the second is the term depending on the variance matrix of the random term and the third can be interpreted as a penalty for estimating fixed effects. The solutions for the parameters of interest, in this case σ_a^2 and σ_e^2 , can be obtained by different methods. One of the most popular have been proposed by Gilmour et al. and it is also implemented in software for variance components estimations [139, 140].

The solutions of the linear mixed model for the parameters in **a** and **b** can also be approached by Bayesian or non-parametric methods [150]. In the definition above, the genetic relationship matrix **A** is assumed to be computed from pedigree information only, however it can be substituted with one of the genetic relationship matrices presented above in order to perform the estimation of the genetic variance with more accurate relatedness coefficients.

2.4.4 Estimation of breeding values

The estimated breeding value is a genetic parameter used for setting up breeding plans. The breeding value of an individual is defined as twice the mean deviation of its progeny from the population mean, its progeny being obtained by random mating of that individual with other individuals of its population and the means being computed from the phenotypes of the trait under selection or the phenotypes of multiple traits combined in a selection index. The

breeding value of an individual represents the sum of the allelic substitution effects of the alleles its gametes can bear, therefore, the breeding value of an individual also equals half the sum of its parents' breeding values because meiosis results in only half of the alleles of each parents to form a zygote. The cumulative effect of all alleles was indeed the only estimable quantity, back when no affordable genomic information was available. This resulted in the assumption that all loci have the same infinitesimal effect on the phenotype, to be necessary for the infinitesimal model. Nevertheless, this assumption resulted in the development of a very flexible framework of analytical procedures for breeding values estimation, namely: the linear mixed animal model, the best linear unbiased prediction (BLUP) and the restricted maximum likelihood [148, 151].

The linear mixed animal model presented in the paragraph above can be also expressed in the following matrix notation:

$$\begin{bmatrix} \mathbf{X}\mathbf{X}` & \mathbf{X}`\mathbf{Z} \\ \mathbf{Z}`\mathbf{X} & \mathbf{Z}`\mathbf{Z}+\mathbf{A}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}`\mathbf{y} \\ \mathbf{Z}`\mathbf{y} \end{bmatrix} \text{ , where } \alpha = \sigma_e^2/\sigma_a^2$$

The solutions for the vectors $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ can be obtained simultaneously by BLUP, given that the solutions for the σ_a^2 and σ_e^2 variance components, obtained by restricted maximum likelihood, are considered the true values of the environmental and genetic variance, respectively.

As a matter of fact, the estimation of breeding values based on pedigree only method have been recently outperformed, in terms of accuracy, by a genomic approach which allows to account simultaneously for both the family structure and the IBD probabilities of haplotypes. Indeed, in the original formulation of the linear mixed animal model, the genetic relationship matrix was computed according to pedigree only, since its elements represent the relatedness and inbreeding coefficients averaged over all loci. This implies that all alleles were considered to contribute to their cumulative effect estimated as the breeding value in $\hat{\mathbf{a}}$. Despite the fact that genetic selection based on this assumption allowed significant genetic progress in livestock populations, especially when used in synergy with artificial insemination [152], advances in life sciences, such as biochemistry, physiology and quantitative genetics itself, suggest that some loci might play a larger role than others in the observed variation of the phenotypes under study [153]. The information conveyed by inexpensive molecular markers allows both to observe which alleles were actually inherited from each parent and to estimate their allelic substitution effect at each marker locus, by several strategies. The benefit of these

estimates to genetic selection programs can be twofold: it increases the accuracy of the estimated breeding values and it helps in optimizing the choice of selection candidates, especially for sex-related phenotypes or those which cannot be observed in vivo [154]. The former comes from allowing both to compute a more accurate relationship matrix than what is achievable with pedigree information only [155] and to enhance the accuracy of ungenotyped animals' breeding values with the information extracted from genotyped ones. This strategy was originally proposed by Meuwissen *et al.* in 2001, and relies on the availability of more and more dense SNPchips developed for livestock species [129]. This ground-braking article was developed further by numerous research groups. Furthermore, these methods allow both to obtain the estimated position and effect of the putative QTLs and to integrate this estimation into the estimate of the breeding values, simultaneously; which eventually increases the accuracy of the latter estimate compared to the estimate obtained by pedigree information only [156]. The definition of the breeding value estimated by these models remains the cumulative effect of the marker loci, except that the effect of each of them is computed. This results in the model reformulation of the classical linear mixed animal model as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \sum_{k=1}^{k} z_k a_k + \mathbf{e}$$

Where k is the number of marker loci, $\mathbf{z_k}$ is the vector of genotypes at locus k and a_k is the effect of marker k. The other terms correspond to the terms of the classical linear mixed animal model: $\mathbf{y}$ is the vector of phenotypes; $\mathbf{X}$ is the incidence matrix of the vector of fixed environmental effects $\boldsymbol{\beta}$; $\mathbf{u}$ is the vector of individuals' polygenic effects remained unobserved from the molecular markers and $\mathbf{e}$ is the vector of residuals. Therefore, these methods aim at estimating as many as k allelic substitution effects, one for each marker locus, which can result in large mean square errors (MSE) of the estimates themselves because of the high density of modern SNPchips. Given a parameter estimated from a vector of observations, $\hat{\theta}(\mathbf{y})$, and the true unobservable value of that parameter, θ, the MSE of the parameter estimate equals: $\mathrm{MSE}(\hat{\theta}) = E[\hat{\theta}(\mathbf{y}) - \theta]^2 = Var[\hat{\theta}] + Bias[\hat{\theta}]^2$; i.e. the sum of the parameter's variance and its bias. One strategy to reduce the MSE is to reduce the variance term, for example, by shrinkage. Let $\alpha \in [0,1]$ and $\tilde{\theta} = \alpha\hat{\theta} + (1-\alpha)0 = \alpha\hat{\theta}$, the shrinkage parameter α results in $Var(\alpha\hat{\theta}) = \alpha^2 Var(\hat{\theta})$, which ensures $Var(\tilde{\theta}) < Var(\hat{\theta})$ as long as $0 < \alpha < 1$. However, $\tilde{\theta}$ can be biased when $\theta \neq 0$ [156]. Shrinkage is mostly applied to the models used in quantitative genetics by penalized regression or by several Bayesian

methods.

Penalized regression methods include those methods which rely on obtaining the parameters estimates as the solutions to an optimization problem, such as ridge regression [157], LASSO [158], and elastic net [159]. These methods differ between each other in the function chosen for minimizing the following equation according to $\hat{\mu}$ and $\hat{\boldsymbol{\beta}}$:

$$\left\{\sum_{i=1}^{n}\left(y_i-\mu-\sum_{j=1}^{p}x_{ij}\beta_j\right)^2+\lambda J(\boldsymbol{\beta})\right\}$$

Where the first term is the residual sum of squares of a linear mixed model, featuring the vector $\boldsymbol{\beta}$ as the marker effect, and λ is the shrinkage parameters as a function, *J,* of the complexity of the model, related to the number of parameters to estimate. In Ridge regression the $J(\boldsymbol{\beta})$ function is proportional to the sum of squares of the regression coefficients, the function used by LASSO allows both to zeroing out some parameters and to obtain shrunk estimates of the parameters left to estimate. The elastic net was designed as a combination of the two other penalized regressions [156].

Another approach to shrinking is found in Bayesian methods. These methods allow both the variable selection and shrinkage by setting up the corresponding prior density distribution of the markers effects; which are, in fact, drawn from the above mentioned prior distribution, by an algorithm such as the Gibbs sampler, until it converges to the combination of parameters which maximizes the goodness of fit of the model. This requires the modeller to define the hyperparameters whose values shape the prior distribution. The shape of the prior used to draw the allelic substitution effects at marker loci, is indeed what makes the difference between these Bayesian methods. All of these models can be summarized by the following Bayesian setting:

$$\begin{aligned}
&p(\mu,\boldsymbol{\beta},\sigma^2|\mathbf{y},\boldsymbol{\omega})\\
&\propto p(\mathbf{y}|\mu,\boldsymbol{\beta},\sigma^2)\,p(\mu,\boldsymbol{\beta},\sigma^2|\boldsymbol{\omega})\\
&\propto\prod_{i=1}^{n}N\left(y_i|\mu+\sum_{j=1}^{p}x_{ij}\beta_j,\sigma^2\right)\prod_{i=1}^{n}p(\beta_j,|\boldsymbol{\omega})\,p(\sigma^2)
\end{aligned}$$

Let $(\mathbf{y})$ be the vector of data, let $\{\mu,\boldsymbol{\beta},\sigma^2\}$ be the model's unknowns and let $\boldsymbol{\omega}$ be the vector of hyperparameters defining the prior distribution's density; then, $p(\mu,\boldsymbol{\beta},\sigma^2|\mathbf{y},\boldsymbol{\omega})$ is the posterior distribution density of the estimates, given the data and the hyperparameters;

$p(\mathbf{y}|\mu,\boldsymbol{\beta},\sigma^2)=\prod_{i=1}^{n}N\left(y_i|\mu+\sum_{j=1}^{p}x_{ij}\beta_j,\sigma^2\right)$ is the conditional density of the data, given the

unknowns; $p(\mu, \boldsymbol{\beta}, \sigma^2 | \boldsymbol{\omega}) \propto \prod_{i=1}^{n} p(\beta_j, | \boldsymbol{\omega}) p(\sigma^2)$ is the joint prior density of the model's unknowns. The values assigned to the hyperparameters **ω** define, among others, the thickness of the prior distribution around zero, which corresponds to the shrinkage and variable selection processes of the penalized methods. Indeed, the Bayesian methods can be grouped, according to the prior distribution they use, into the following groups. Models using a Gaussian prior, such as ridge regression BLUP [160]. Models using thick-tailed priors, such as the scalet-*t* of Bayes A [129] and the double exponential or Laplace distributions of the Bayesian LASSO [171]. Spike-slab priors, which are defined as the mixture of a large variance and small variance distribution; this is achieved for example by the stochastic search variable selection [162], by applying the Pareto principle [163]. The modeller can also choose to build the spike-slab prior by mixing non Gaussian distribution, such as the scaled-*t [164]*. Following the same principles of spike-slab modelling, the modelling of the prior can be pushed to the more extreme case of point of mass at zero and slab prior. This shape can be obtained either by a scaled-*t* distribution, as in Bayes B [129], or by a Gaussian distribution, as in Bayes C [165].

In the comparison of Bayesian models performed both on simulated and real data, Habier et al. 2011 [165] showed that the most appropriate formulation of Bayesian regression is Bayes Cπ because it allows to learn the prior distribution's parameter from some training data and because of its computational efficiency. Indeed, the former property can account, at least to a certain extent, for the unknown genetic architecture of the trait under study; whereas the methods that let the user define these parameters suffer from arbitrariness [165, 166]. The Bayesian LASSO [167] was not included in the study by Habier et Al. [165], but the authors themselves warn that the definition of the parameter λ must be treated with caution, because of its potential impact on the results. Whatever the method used for genomic prediction, all the methods mentioned above provide estimates of the breeding values for genotyped animals only. In order to include the genomic information into the estimate of ungenotyped animals' breeding values it is necessary to follow a multi-step procedure which could result in the loss of information which can nullify the potential benefit of investing in genotyping [168].

Aguilar et al. [155] proposed a method which allows to avoid multi-step procedures for genomic predictions and obtain comparable results in terms of accuracy. This method is based on the classical formulation of the linear mixed animal model. The difference is that the genetic relationship matrix between the individuals includes both pedigree and molecular

information [147]. The main advantage of this model is that it avoids the potential loss of information caused by the multi-step evaluation. Furthermore, it allows to rejoin easily the detailed modelling at the molecular level (by the choice of an appropriate scaling parameter for the elements of the genomic relationship matrix) to the flexibility of the well known framework developed for evaluations based on pedigree information only. Last, but not least, the computational burden is much lighter than the methods mentioned above.

Semi-parametric [128, 130] and non-parametric [169] approaches were also proposed by Gianola et Al. These methods make it feasible to rank individuals according to their genetic merit by taking into account also non-linear relationships between the genotypic and phenotypic variation, whose relationship is in fact very far from being linear [170]. Although this feature reflects closer the real genetic determinism of quantitative traits, these methods did not receive much attention in practice.

2.4.5 Estimation of the QTLs' allelic substitution effects

Before the genomic approach discussed above, which considers all marker information simultaneously, the methods used to analyse the genetic variability in its elementary components were based on the concept of quantitative trait locus (QTL). A QTL is a region of the genome carrying one or more polymorphic genes with alleles displaying a different effect on the quantitative trait of interest. Hence, a QTL would be a locus on the genome whose allelic polymorphism explains a major proportion of the phenotype under study. The existence of such loci have been proven for a number of phenotypes of commercial interest in different livestock species [153, 171], however the number of loci featuring an economically significant effect is expected to be small compared to the number of loci featuring small effects [172], specially when the additive effect only is considered.

The estimation of the allelic substitution effects at QTL loci can be estimated following different strategies: linkage analysis (LA), linkage disequilibrium analysis (LDA) and a combination of the two (LDLA). These approach allow to obtain information about the possibly significant association between the variability observed at marker loci in a region of the genome and the phenotypic variation of the trait under study. All strategies for detecting QTLs rely on the occurrence of linkage disequilibrium, which was first described by Bateson and Punnett in the early 1900 as a deviation from the segregation patterns between two traits expected according to the Mendelian law of independent segregation [173]. The level of linkage disequilibrium depends on several factors, such as: selection, mutation, migration and

drift. Its persistence across generations depends instead on the distance between the QTL and the marker locus (which can be reduced by using high density SNPchips, for example) [174]. Hill and Robertson proposed a measure of linkage disequilibrium, based on the D-statistic, called r² [175]. The D-statistic was defined as the difference between the right hand term and left hand term of the equation: $freq(AB)\,freq(ab) = freq(Ab)\,freq(aB)$; where AB, ab, Ab and aB are the genotypes observed in the population under study and *freq(.)* is the frequency of the genotype between brackets. The r² is derived from D as $r^2 = \frac{D^2}{p(1-p)q(1-q)}$; where p and q are the allele frequencies. The difference between LA and LDA resides in the fact that the LA approach relies on the between-generation transmission patterns of markers and putative QTLs, while LDA relies on long term linkage disequilibrium between marker and putative QTLs.

LA is based on the linkage disequilibrium observed between two loci located on the same chromatid, which is measured by the recombination rate between them. By observing the genotype at marker loci of parents and offspring it is possible to estimate, by defect, the number of chiasmata occurred between two loci, because they break the parental phase, and this, in turn, allows to infer the position of a putative QTL between the two marker loci according to an appropriate mapping function [176]. The presence of a QTL is postulated when the estimated allelic substitution effect of a marker loci is significant, based on the idea that the significance of its effect is due to the linkage disequilibrium between the marker locus and the QTL.

In LDA, the association between a marker locus and a putative QTL can be tested by several strategies. The simplest screening can be done by multiple testing each marker locus separately against the null hypothesis that its allelic substitution effect is not different from zero. In order to do so, the gene content at the marker locus is coded according to the number of copies of one of the alleles, and its effect is estimated as a fixed effect in a regression model. This method also allows to account for the population structure by adding a random component of polygenic effect similar to the linear mixed animal model described above [177]. Multiple testing on the same dataset requires a correction of the significance test for each locus which can be achieved by several methods [178]. Müller et al [179] proposed a method specially conceived for multiple testing of marker loci in genetic analysis that takes into account the linkage disequilibrium between the marker loci and allows to not to consider the tests as independent.

The power of LA and LDA can be improved by considering more than one marker bracket at a time. This approach was proposed by Lander & Botstein as interval mapping [180]. In the context of LA analysis: the parents' genome is phased (which requires the grand-parental origin of the alleles to be known, for each heterozygous marker) and the presence of a QTL between two adjacent markers is tested using the observations on the pairs of marker alleles transmitted to the offspring. In the context of LDA and LDLA, the association between QTLs and marker haplotypes is based on the prediction of the QTL allele carried by the chromosome conditional to the observed marker haplotypes.

In order to increase further the power to detect putative QTLs, it is necessary its alleles to feature the most heterozygosity as possible at population level but in real data analysis the alleles of the putative QTL are unknown, as well as its existence in the first place. Moreover, in LA, the haplotypes of the parents are rarely available; what is known are the genotypes of the offspring and their phenotype. However, in the context of LA analysis, it is possible to maximize the occurrence of informative marker brackets by appropriate experimental designs, such as: crossings of divergent lines, backcrosses, daughter design and grand daughter design [104]. Such kinds mating schemes are indeed designed for maximizing both the heterozygosity of the putative QTL alleles and the linkage disequilibrium between them and the marker positions available within the population.

2.4.6 Prediction of the response to selection

The estimation of breeding values and the partitioning of the phenotypic variance into its genetic (σ_a^2) and environmental (σ_e^2) components are useful for predicting the theoretical response to genetic selection. The response to selection is the expected shift of the offspring's population mean for the phenotype(s) included in the breeding objective from parents' population mean. Under the assumptions of linearity underlying the models currently applied in practice for estimating breeding values, either including molecular information or not, and of Gaussian distribution of the phenotype under selection, the regression between the mid-parent's values and the offspring' s averages is also linear. [181].

The basic equation for predicting the response to selection R reads $R=h^2S$, where h^2 is the estimated stricto sensu heritability $h^2=\sigma_a^2/\sigma_p^2$, the ratio between the additive genetic variance and the phenotypic variance; S is the selection differential $S=i\,\sigma_p$, the product of the selection intensity i and the phenotypic variation in the parent's population. The

heritability indicates how much information on the true breeding value of an individual is contained in its phenotype, the square root of the heritability is indeed the correlation between the phenotype and the true breeding value of an individual [120, 182]. The intensity of selection is a function of the proportion animals used as parents for the next generation over the whole population under selection. The aim of breeding plans is to maximize the selection response, which can be achieved by several strategies. The most intuitive solution would be to increase the selection intensity. However this is not very effective because by reducing the number of individuals contributing to the genetic pool for the next generation it increases the inbreeding, which results in long term inbreeding depression. Furthermore, the selection intensity is limited by the reproductive potential of the species under selection [120]. Another solution could be to increase the heritability estimate of the traits under selection. Given $\sigma_p^2=\sigma_a^2+\sigma_e^2$, the estimate of h^2 increases along with the reduction of the environmental variance. This can be achieved, for example, by applying appropriate rearing and management techniques on the individuals, the phenotypes of which are collected from; by collecting repeated measurements of the phenotype; by collecting phenotypes on the relatives of the selection candidates; and finally by the choice of the phenotype itself.

Molecular information can also enhance the selection response by increasing the accuracy and precision of both the estimated breeding values [183]. However, the benefit deriving from molecular information in terms of selection response augmentation, depends on the type of marker used [171], which results in two three scenarios. Gene assisted selection, which can be done when the causal mutation affecting the phenotype is known (for example the marker linked to the mutation of myostatin for double muscling in cattle [184]). Linkage equilibrium marker assisted selection, based on QTL markers in population wide linkage disequilibrium with the causal mutation in outbred populations (such as the gene for the polled phenotype in cow [185]). Linkage disequilibrium marker assisted selection, based on QTL marker in linkage disequilibrium with the causal mutation within the population under selection (such as the calpastatin QTL for carcass quality in pig [186]).

Since the distance of the marker loci from the causal mutation greatly influences the impact of marker information on the response to selection, gene assisted selection always outperforms selection based on linkage equilibrium or linkage disequilibrium markers [187]. One of the reasons for this is that when linkage disequilibrium between the marker and the causal mutation is not complete,. the association is gradually lost over generations due to recombination events [171, 187]. Therefore, the accuracy of the overall breeding objective

due to marker information gradually decreases, depending on the amplitude of the estimated effect and on the distance from the causal mutation. As a matter of fact, the investment in genotyping and detection of molecular markers is justified only when a marker has both large impact on the phenotype and is very close to the causal mutation affecting the phenotype, which is rarely the case [168, 171, 187, 188].

Nevertheless, molecular markers can help in the preselection of candidates to selection, specially for those traits whose phenotypes are expressed late in the life of the individual (such as reproduction traits and milk production) and those whose phenotype cannot be observed directly (such as sex linked traits and carcass traits). Preselecting selection candidates according to molecular marker information can reduce indeed the generation interval, which influences the yearly genetic gain, and to optimize the implementation of breeding schemes. Other benefits of marker information include the possibility to estimate allelic substitution effects across populations, which can be eventually exploited by marker assisted introgression [189], in outbred populations, and the possibility to control inbreeding at molecular level [141].

3 Experimental studies

In the previous chapter it was shown how the efficacy of genetic selection as a control strategy for haemonchosis in small ruminants production systems can be estimated by predicting the expected selection response of an appropriate breeding plan. Some of the genetic parameters required for doing so are the genetic correlations between the traits included in the overall breeding goal and the allelic substitution effects at the available marker loci. The former conveys information on the reciprocal impact of correlated responses among the traits under selection and helps designing the mating schemes in order to maximize the selection responses on all traits simultaneously. The latter helps with exploring what kind of genetic variability is available for selection and, therefore, refining the prediction of the selection response by assessing whether gene/marker assisted selection would be a feasible option for enhancing classical polygenic selection or not [120].

In order to make the first step towards the realization of a model for the selection response prediction, and eventually advise the selection actors to take into account parasitism resistance in selection, the genetic parameters for growth and parasite resistance traits were estimated in sheep (published article) while the heritability estimates of parasite resistance traits and the presence of marker allelic substitution effects significantly affecting them was explored in creole goat (submitted article). Both studies are based on the phenotypes collected by following the same experimental infestation protocol applied on sheep and goat in first and second study, respectively. The protocol implied two subsequent larval challenges with 10000 L3 larvae of *H.contortus*, the animals were managed so as to ensure that they didn't encounter the parasite before the beginning of the experiment. Furthermore, the second larval challenge was performed late enough to allow the animals earlier exposed to the first infestation to develop a specific immune response to the parasite. Therefore, the experimental design allowed to explore the phenotypes expressed during the innate immune response (first infestation), those expressed during the specific immune response (second infestation) and the relationship between the two. The experimental settings also allow to control the ratio between the nutritional level of the diet and the larval challenge, which has been previously shown to influence the host parasite interaction greatly [101, 102]. The experimental settings designed for these experiments featured high larval challenge and high nutritional level diet for all animals, therefore the study of the interaction between these two factors was left

beyond the scope of these experiments.

3.1 Article 1

The first study aimed essentially at addressing the first of the above issues: to estimate the genetic parameters related to correlated responses to selection among growth and parasite resistance traits in sheep. The results obtained from the study feature the heritabilities of and the genetic correlations between growth and parasite resistance traits in sheep. The estimates of the genetic correlation reported in our study add up to the remarkably wide range of values reported previously, which vary from negative to positive numbers [90, 93, 97], and contributes to build a better understanding of the phenomenon. In order to asses the benefit of including molecular information in the estimation model, this study also features a comparison between the genetic parameters' estimates precision obtained by using pedigree information only and those obtained by using pedigree and molecular information together. Furthermore, since growth traits were measured both before the first infestation and during the first infestation, this study also allowed to explore the eventuality of significant genotype by environment interaction affecting growth traits across non-contaminated and contaminated environments. These pieces of information are useful for drawing the guidelines to be followed for optimizing the selection response on all traits simultaneously.

This article has been published on *genetic selection evolution (DOI: 10.1186/1297-9686-46-13).*

Genetic parameters for growth and faecal worm egg count following *Haemonchus contortus* experimental infestations using pedigree and molecular information

Fabrizio Assenza[1§], Jean-Michel Elsen[1], Andrés Legarra[1]; Clement Carré[1], Guillaume Sallé[2, 3], Christele Robert-Granié[1], Carole Moreno[1]

[1] INRA, UR0631 Station d'Amélioration Génétique des Animaux, F-31326 Castanet-Tolosan, France

[2]INRA, UMR1282 Infectiologie et Santé Publique, F-37380 Nouzilly, France

[3]Université François Rabelais de Tours, UMR1282 Infectiologie et Santé Publique, F-37000 Tours, France

[§]Corresponding author

Email addresses:

FA: fabrizio.assenza@toulouse.inra.fr

JME: jean-michel.elsen@toulouse.inra.fr

AL: andres.legarra@toulouse.inra.fr

CC: clement.carre@toulouse.inra.fr

GS: gsalle@tours.inra.fr

CRG: christele.robert-granie@toulouse.inra.fr

CM: carole.moreno@toulouse.inra.fr

3.1.1Abstract

Haemonchosis is a parasitic disease that causes severe economic losses in sheep industry. In recent years, the increasing resistance of the parasite to anthelmintics has raised the need for alternative control strategies. Genetic selection is a promising alternative but its efficacy depends on the availability of genetic variation and on the occurrence of favourable genetic correlations between the traits included in the breeding goal. The objective of this study was twofold. First, to estimate both the heritability of and the genetic correlations between growth traits and parasite resistance traits, using bivariate linear mixed animal models, from the phenotypes and genotypes of 1004 backcross lambs (considered as a single population), which underwent two subsequent experimental infestations protocols with *Haemonchus contortus*. Second, to compare the precision of the estimates when using two different relationship matrices: including pedigree information only or including also SNP (single nucleotide polymorphism) information.

Heritabilities were low for average daily gain before infestation (0.10 to 0.15) and average daily gain during the first infestation (0.11to 0.16), moderate for faecal egg counts during the first infestation (0.21to 0.38) and faecal egg counts during the second infestation (0.48 to 0.55). Genetic correlations between both growth traits and feacal egg count during the naïve infestation were equal to zero but the genetic correlation between faecal egg count during the second infestation and growth was positive in a *Haemonchus contortus* free environment and negative in a contaminated environment. The standard errors of the estimates obtained by including SNP information were smaller than those obtained by including pedigree information only.

The genetic parameters estimates suggest that growth performance can be selected for independently of selection on resistance to naïve infestation. Selection for increased growth in a non contaminated environment could lead to more susceptible animals with long-term exposure to the infestation but it could be possible to select for increased growth in a contaminated environment while also increasing resistance to the long-term exposure to the parasite. The use of molecular information increases the precision of the estimates.

3.1.2 Introduction

Haemonchus contortus (*H. contortus)* is a nematode that feeds on blood through the abomasal mucosa of bovine, ovine and caprine species [6]. The cost of *H. contortus* infection or haemonchosis for the production sectors of sheep farming in terms of anthelmintic treatments that are currently the most popular control strategy, and the resulting economic loss have been estimated in different countries to be in the order of several million dollars per year [2, 8, 9, 11, 17, 190]. Furthermore, anthelmintics tend to select the parasite population under treatment for resistance to the anthelmintic itself [191, 192], which increases the cost of haemonchosis even more.

Both the long-term loss of efficacy and the growing public concern for the use of chemicals in food production fostered the research on alternative control strategies or combinations of them [21], among which genetic selection is one of the most promising approaches [82, 84, 85, 92, 193–203]. Simulation studies based on evolutionary genetics [204] predict a breeding plan's long-term outcomes and also the efficacy of genetic selection as a control strategy. Since genetic improvement depends on the genetic parameters of the traits under selection, the estimates of these parameters must be as precise as possible for reliable long-term predictions. However, consistent estimates of the genetic correlation between production traits and parasite resistance traits have not been reported in the literature [97, 98, 117, 205, 206]. Since most of the estimates found in the literature are computed from observations in natural conditions, where it is not possible to precisely define neither the nutritional level of the diet nor the larval challenge on the pasture, the reason why no consistent estimates are available may be due to the interaction between these two factors [207].

The first objective of this study was to estimate the heritability of average daily gain and faecal egg count from experimental observations, together with the genetic correlations between them. We report the results of an analysis performed on 1004 phenotypic records of growth traits and faecal egg counts collected on genotyped (50k SNPchip) back-cross lambs (25% Martinique black belly and 75% Romane), following two experimental infestation with *H. contortus.* The genetic parameters have been estimated both by using pedigree information only and pedigree and SNP (single nucleotide polymorphism) chip information jointly. Computing the relatedness between individuals using pedigree information only is based on expectation and results in an estimate corresponding to the average number of alleles shared by two individuals, for example: all the individuals belonging to the same full-sibs group would have a coefficient of 0.5 between each other, which meaning means that it does not

take into account the deviation from this average caused by segregation and recombination. However, including molecular information makes it possible to compute the relatedness between individuals by identifying on a relatively dense map the actual number of alleles they share, which provides a more detailed estimate of the relatedness between individuals [208]. Since the observations used in this study were collected on four large groups of half-sibs, the second objective of the study was to test whether including SNP information could help reach more precise estimates than using pedigree information only when the pedigree of the population is poorly informative. Although previous studies have already explored the amount of genetic variability for parasite resistance traits, this study features several novelties: the genetic parameters reported here are estimated from phenotypes collected in experimental settings rather than natural infestation, the growth traits analysed are the average daily gains before infestation and during infestation rather than the body weights and finally the estimates reported here also feature molecular information rather than pedigree information only.

3.1.3 Materials and Methods

3.1.3.1 Experimental design

The population in which the observations were collected from resulted from a back-cross mating scheme between two pure-bred populations: Martinique Black Belly (MBB) and Romane (ROM). MBB is a tropical sheep breed, which is characterized by adaptation to heat-stress, to parasitism and to extensive raising conditions. ; ROM sheep breed features good productive performances (both for meat production and prolificacy) and no selection for resistance to parasites. The pedigree used in the analysis (Figure 1) was three generations deep and counted a total of 3164 animals. Four F1 sires were produced by crossing MBB and ROM individuals. The sires were mated by intra-uterine artificial insemination to 829 pure-bred ROM dams in order to obtain 1265 back-cross offspring (BC), the number of animals used in this study from each group of half-sibs were 282, 251, 247 and 223, respectively.

3.1.3.2 Genotypes

A total of 1044 animals among the population of the back-cross lambs and their four F1 sires were genotyped with the OvineSNP50 Beadchip (Illumina Inc., San Diego, CA). Quality control of the SNPs included the following tests: (1) 50 animals were genotyped twice in order to assess the technical reliability of the genotyping, which resulted in a value of 99.9%, (2) individuals with a call rate below 98% and SNPs with a call rate below 97% or with a

minor allele frequency below 1% or featuring a deviation from expected heterozygosity or showing mendelian inconsistencies were discarded ($p < 10^{-6}$) and (3) quality control of the genotypes resulted in 42 469 SNPs that comply to with all quality checks. More details about the genomic information can be found in Sallé et al.[209].

3.1.3.3 Phenotypes

Phenotypes on growth traits and parasite resistance traits were collected on the BC animals only by performing the following experimental protocol. The lambs were weaned around 64 or 45 days, depending whether their mother was either primiparous or not, and grew in a *H. contortus* free environment until the first experimental infestation was performed. During this period the animals were kept in an *H. contortus* free environment and were weighed twice: at weaning and at the end of the growing period. The growing period ended with the beginning of the following experimental infestation protocol, which also determined when the environment was to be contaminated with *H. contortus* larvae: at around 90 days of age, the lambs received an oral inoculation of 10 000 L3 larvae of *H. contortus* (ENVT strain [210]) and around 41 days after the infestation, they received an anthelmintic treatment (LEVAMISOLE 5%, Vibrac S.A., Carros, France, 7.5 mg/kg live weight). During the infestation, two faecal samples were collected, at 25 and 35 days after infestation, and the animals were weighed on the day of treatment. Then, they entered a recovery period of 8 days, at the end of which they were infested again with the same infestation protocol. During the second infestation, two faecal samples were collected as before but animals were not weighed. During the whole protocol, the animals were fed ad libitum on a diet that covered largely covering their requirements. The faecal egg count in each sample was measured by a modified McMaster procedure [211]. The average of the two faecal egg count observations was computed for each infestation. The latter values were transformed by taking their fourth root in order to bring their distribution closer to normality. A further transformation was applied in order to scale the standard deviation to 1 and avoid zero values. The variables obtained were called: FEC1 (transformed faecal egg counts during the first infestation) and FEC2 (transformed faecal egg counts during the second infestation). The average daily gain from weaning to infestation (ADG0) and average daily gain during the first infestation (ADG1) were computed as follows: $ADG0 = \frac{growth0}{time0}$ and $ADG1 = \frac{growth1}{time1}$, where growth0 and time0 are the weight gain and the days running from the weaning day to the day of beginning

of the first infestation, respectively; growth1 and time1 are the weight gain and the days running from the first day of the first infestation to the day of treatment of the first infestation, respectively.

A transformation for scaling the standard deviation to 1 and avoiding 0s was applied on growth traits as well. The observations outside a range of 2.96 standard deviations around the average of each trait were considered atypical and excluded from the analysis. Finally, only the animals featuring a valid observation both on genotyping and on at least one trait were included in the analysis, which resulted in 40 discarded animals being discarded and 1004 included animals being included.

3.1.3.4 Statistical analysis

Estimation of the genetic parameters was performed by considering the back-cross population as a single breed population. The breed proportions are taken into account by the genetic relationship matrix in the model including SNP information but could not be taken into account in the model including pedigree information only due to convergence failure. The heritability of each trait and both the genetic correlation and phenotypic correlation between each pair of traits were estimated by bivariate animal mixed models, which were solved by the AIREML procedure implemented in AIRemlF90 software [212]. This software features by default the correction for the change in the definition of the base population so that the estimates obtained when using pedigree information only were comparable to those obtained when including molecular information [213, 214]. The bivariate mixed model reads as follows:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} ,$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are the vectors of observations of trait one and two, respectively, $\mathbf{X}_1$ and $\mathbf{X}_2$ are incidence matrices relating each observation to its respective set of fixed effects and $\mathbf{b}_1$ and $\mathbf{b}_2$ are the vectors of the fixed effects: weight at weaning (for ADG0 only) or weight at first infestation (for all the other traits), contemporary group (identified by year, season, weighting lot and infestation lot), sex and feeding mode (breast feeding or bottle feeding). $\mathbf{a}_1$ and $\mathbf{a}_2$ are the vectors of random animals breeding values, with the associated incidence matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$. $\mathbf{e}_1$ and $\mathbf{e}_2$ are the vectors of random residuals. It is assumed that the

random effects are normally distributed and feature the following variance-covariance structure:

$$\mathrm{VAR}\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\\ \mathbf{e_1}\\ \mathbf{e_2}\end{bmatrix}=\begin{bmatrix}\sigma^2_{g11}\mathbf{T} & & \text{symmetric} & \\ \sigma_{g21}\mathbf{T} & \sigma^2_{g22}\mathbf{T} & & \\ 0 & 0 & \sigma^2_{e11}\mathbf{I} & \\ 0 & 0 & \sigma_{e21}\mathbf{I} & \sigma^2_{e22}\mathbf{I}\end{bmatrix}$$

, where σ^2_{g11}, σ^2_{g22} and σ_{g21} are the genetic variances and the genetic covariance between traits 1 and 2, σ^2_{e11}, σ^2_{e22} and σ_{e21} are the residuals variances and the residuals covariance between traits 1 and 2, **I** is an identity matrix and **T** is the genetic relationship matrix between the animals.

The genetic and phenotypic correlations between each couple of traits were computed using two different genetic relationship matrices: one computed by using pedigree information only (PED model) and one computed by including both SNP and pedigree information jointly (SNPED model). The **T** matrix used in the PED model was computed according to Quaas [215] and without using molecular information. The **T** matrix used in the SNPED model was computed according to VanRaden [142], using a three-generation deep pedigree and the following weights: $\mathbf{T}=0.95\mathbf{G}+0.05\mathbf{A_{22}}$; where $\mathbf{A_{22}}$ is the relationship matrix between the genotyped animals computed by using the pedigree information only [215] and **G** is the genomic relationship matrix among genotyped animals. In the software package used for this study [212], the **G** matrix is computed by default as follows: $\mathbf{G}=\mathbf{W}\mathbf{K}^{-1}\mathbf{W}$. [142] where **W** is a rectangular matrix (number of animals by number of SNPs alleles) with elements: $w_{ij}=f_{ij}-2\mathrm{p}_j$, where f_{ij} is a scalar equal to the number of copies of one allele an animal i has at locus j, p_j is the frequency of allele j in the population, **K** is the diagonal matrix of the scaling parameters with elements: $\mathbf{K}_{jj}=2\sum_1^j p_j(1-p_j)$. The weights of **G** and $\mathbf{A_{22}}$ are used for bending the genetic relationship matrix and make it positive definite, as required for its inversion [142], this is similar to the bending procedure occurring in AIREML algorithms for keeping the variance covariance matrix positive definite [215].
The precision of the heritability and correlation estimates was computed by estimating their standard error according to the following formulas [216].

$$SE(h_i^2)=\sqrt{\left(\frac{\sigma_{gii}^2}{\sigma_{pii}^2}\right)^2\left(\frac{VAR(\sigma_{gii}^2)}{(\sigma_{gii}^2)^2}+\frac{VAR(\sigma_{pii}^2)}{(\sigma_{pii}^2)^2}-\frac{2\mathrm{COV}(\sigma_{gii}^2,\sigma_{pii}^2)}{\sigma_{gii}^2\sigma_{pii}^2}\right)}$$

, where $SE(h_i^2)$ is the standard error of the estimate of the heritability of trait i; σ_{gii}^2 and σ_{pii}^2 are the estimates of its genetic and phenotypic variances, respectively ; $VAR(\sigma_{gii}^2)$, $VAR(\sigma_{pii}^2)$ and $COV(\sigma_{gii}^2,\sigma_{pii}^2)$ are the variances of the estimated values and the covariance between the estimated values, respectively, obtained from the information matrix [216].

$$SE(r_g)=\sqrt{r_g^2\left(\frac{VAR(\sigma_{g11}^2)}{4(\sigma_{g11}^2)^2}+\frac{VAR(\sigma_{g22}^2)}{4(\sigma_{g22}^2)^2}+\frac{VAR(\sigma_{g21})}{(\sigma_{g21})^2}\right)}+$$

$$\sqrt{r_g^2\left(\frac{COV(\sigma_{g11}^2,\sigma_{g22}^2)}{2\sigma_{g11}^2\sigma_{g22}^2}-\frac{COV(\sigma_{g11}^2,\sigma_{g21})}{\sigma_{g11}^2\sigma_{g21}}-\frac{COV(\sigma_{g22}^2,\sigma_{g21})}{\sigma_{g22}^2\sigma_{g21}}\right)}$$

,where $SE(r_g)$ is the standard error of the estimate of the genetic correlation, r_g is the estimated value of the genetic correlation; σ_{g11}^2, σ_{g22}^2 and σ_{g21} are the estimates of the genetic variance components described above and VAR(.) and COV(..) are the variance of the estimates between parenthesis and the covariance between the estimates between parenthesis, respectively. The same formula was used to compute the standard error of the phenotypic correlation, but by filling in the entries concerning phenotypic variances and covariances.

3.1.3.5 Significance tests

The parameters under study are the ratio of two normally distributed variables (heritability) and the ratio between a gaussian variable over the square root of the product of two gaussian variables (correlations). The sampling distribution of heritability can be approximated to a gaussian distribution under certain specific conditions only [217]. When these conditions are filled, the significance test for gaussian variables can be applied. However, the significance tests for the correlation coefficient can be developed by deriving its confidence interval according to Fisher's Z-transformation [218]. Otherwise, both parameters can be tested by using a re-sampling procedure such as delete-d jackknife: where d is the number of observations randomly discarded from the dataset and $n^{1/2} < d < n$ (n is the total number of observations in the dataset) [219]. One thousand sub-samples of the whole dataset were created by randomly discarding 20% of the observations. Each parameter computed above

was re-estimated from each sub-sample in order to build its empirical distribution. The empirical distribution of each parameter was used to compute the confidence interval of each parameter by taking its 2.5% quantile as the lower bound and its 97.5% quantile as the upper bound of each estimate. The null hypothesis "the estimate is not different from 0" was tested as follows: if the confidence interval of the estimate included 0, then the null hypothesis was not rejected, otherwise the alternative hypothesis "the estimate is different from 0" was accepted. In order to test whether the SNPED and SNP models converged to the same estimate, the distribution of the difference between the PED and SNPED estimates was built for each estimate as follows: $[\mathbf{d_i}]=[\mathbf{Ped_i} - \mathbf{Snped_i}]$,where $\mathbf{Ped_i}$ is the vector containing the n realizations of the estimate obtained from the PED model, $\mathbf{Snped_i}$ is the vector containing the n realizations of the estimate obtained from the SNPED model, $\mathbf{d_i}$ the vector of the differences between and each element of $\mathbf{Ped_i}$ and $\mathbf{Snped_i}$. The confidence interval of the distribution of the difference was computed as above. The null hypothesis "the difference between the estimate obtained from the PED model and the estimate obtained from the SNPED model is 0" was tested against the alternative hypothesis "the difference between the estimate obtained from the PED model and the estimate obtained from the SNPED model is not 0" as above as well.

3.1.4 Results and discussion

3.1.4.1 Phenotypic variation

Table 1 shows the summary statistics of the variables analysed. The transformations applied to the raw faecal egg counts resulted in the profile of their distribution being closer to normality. The skewness and normalized kurtosis of FEC1 changed from 2.11 to -0.44 and from 6.56 to 0.53, respectively; the skewness and normalized kurtosis of FEC2 changed from 3.87 to 0.14 and from 23.92 to -0.74, respectively. The number of observations on each trait together with the average, standard deviation, minimum and maximum of the raw observations are in Table 1. ADG1 was significantly lower than ADG0 ($p\_value < 0.0001$), indicating that infested animals grew had a slower growth than the parasite-free animals, as expected due to the infestation [6].

Table 2 shows the estimate of the phenotypic correlations (below the diagonal) obtained from the two models for each pair of traits. Although the SNPED and PED models did not always

converge on the same value, according to the significance tests described above, these estimates were not significantly different between the two models. The estimates of the phenotypic correlations between ADG0 and both FEC1 and FEC2 were not significantly different from 0: -0.01 (SE = 0.15) and 0.01 (SE = 0.18) for the PED model and 0.02 (SE = 0.11) and 0.04 (SE = 0.11) for the SNPED model. These results suggest that the phenotype for growth rate in a *H. contortus* free environment was unrelated to the parasite resistance phenotype. However, the estimates of both the phenotypic correlations between ADG1 and FEC1 and between ADG1 and FEC2 were negative: -0.24 (SE = 0.15) and -0.20 (SE = 0.19) for the PED model and -0.23 (SE = 0.11) and -0.19 (SE = 0.11) for the SNPED model. These results suggest an inverse proportionality between the growth rate and the parasite burden, in accordance with the finding that contaminated animals had a slower growth than non contaminated animals. The average faecal egg count during the second infestation was significantly lower than the faecal egg count during the first infestation ($p\_value < 0.0001$), which suggests that the development of a specific immune response was triggered by the first, that enhanced the intrinsic resistance of the animals to subsequent infestations [48]. Furthermore, the positive estimate of the phenotypic correlation between FEC1 and FEC2, 0.46 (SE = 0.43) for the PED model and 0.62 (SE = 0.20) for the SNPED model, shows that the animals featuring higher (or lower) than average FEC1 are likely to express higher (or lower) than average FEC2, and vice versa. This suggests that a repeatable variation in susceptibility occurs within the population.

3.1.4.2 Genetic variation

Table 2 summarizes the estimates of the heritabilities of each trait (block diagonal) and the estimates of the genetic correlations (above the diagonal) between each pair of traits obtained with the PED and SNPED models, together with the standard error of each estimate (between brackets). Due to the pedigree structure that includes only four sires, both the standard errors and the 95% confidence intervals of the estimates were indeed large (in particular, those of the PED model), which led to no significant difference between the estimates obtained from the two models. The results obtained were in general coherent between models, except for the genetic correlation between ADG0 and FEC1 which was positive with the SNPED model and negative with the PED model. However, the latter estimate had a standard error as large as half the parameter space, which resulted in the confidence of the difference between the two estimates to include 0. The reason for this possible inconsistency cannot be defined by the

data available for this study. We can only speculate that it could be the result of the segregation variance captured by the SNP chip, because the phenotypes analysed were collected on the back-cross offspring of only four sires. This pedigree structure is indeed poorly informative if the estimate is computed by using pedigree information only, which is clearly shown by the huge standard error obtained with the PED model. However, this does not explain why such a big difference in the estimates occurs between some pairs of traits only. The heritability of ADG0 was low for both models, 0.10 (SE = 0.08) for the PED model and 0.15 (SE = 0.07) for the SNPED model, and is close to the value 0.17 found in the review of Safari et al. [117]. The heritability of ADG1 was low as well, 0.16 (SE = 0.04) and 0.11 (SE = 0.06) for the PED and the SNPED model respectively and no estimates were found in literature for ADG1. The estimates for the heritabilities of both faecal egg counts were found higher than those obtained in previous studies [97, 98, 205]: FEC1 was 0.21 (SE = 0.06) and 0.38 (SE = 0.04) for the PED and the SNPED model, respectively; FEC2 was 0.55 (SE = 0.09) and 0.48 (SE = 0.06) for the PED and the SNPED model, respectively. The reason for finding a higher value than in previous studies could be twofold. First, it could be due to the fact that most of the estimates found in the literature are computed from observations in natural rather than experimental conditions, the latter of which allows controlling more strictly the environmental conditions and hence could reduce phenotypic variation. The second reason could be the occurrence of breed specific alleles that segregate within the back-cross population, which inflate the genetic variance compared to a pure breed population. These values confirm the availability of a moderate genetic potential in sheep that could be exploited to enhance resistance to parasites.

Concerning previous estimates of the genetic correlation between growth traits and faecal egg counts, no other estimations of these parameters based on experimental infestation were found in the literature. Furthermore, previous studies on similar traits based on natural infestations do not show consistency among them [206], which could be explained by the uncontrolled variation in the larval challenge, in the pathogenicity of the parasite in each population, in the feed intake and the interaction between these three factors [43].

Concerning the genetic correlations between growth traits (ADG0 and ADG1) and parasite resistance during the naïve infestation (FEC1), the following picture can be drawn. In accordance with the estimates of the phenotypic correlation between ADG0 and FEC1, the genetic correlations between these traits were also not significantly different from 0: -0.52 (SE = 1.06) for the PED model and 0.11 (SE = 0.47) for the SNPED model. The same results were

obtained with the estimate of the genetic correlations between ADG1 and FEC1 (although their phenotypic correlations were negative according to both models): -0.19 (SE = 0.80) for the PED model and -0.12 (SE = 0.58) for the SNPED model. These results suggest that during the naïve infestation, the genotype for growth (ADG0 and ADG1) could be expressed independently from the genotype for parasite resistance (FEC1).

The results obtained for the genetic correlations between growth traits and the long-term resistance (FEC2) were on the contrary significantly different from 0. On the one hand, the genetic correlation between ADG0 and FEC2 was positive but not significantly different from 0 according to the PED model (0.25, SE = 0.85) and positive (0.57, SE = 0.38) according to the SNPED model. According to the approximate standard errors, the positive estimate obtained with the SNPED model is more reliable and suggests that if animals were selected for growth in a parasite-free environment, a correlated selection response for lower long-term resistance to gastrointestinal parasites could occur as well. This estimate supports the hypothesis that enhancing growth traits could come to a cost to the sheep's long-term susceptibility to parasite infestations, and vice versa [99, 220]. On the other hand, the correlation between ADG1 and FEC2 was consistently negative between models: -0.48 (SE = 0.67) for the PED model and -0.54 (SE = 0.53) for the SNPED model, which suggests that growth during the naïve infestation can be enhanced together with long-term resistance to the infestation within a single purebred line.

The genetic correlation between FEC1 and FEC2 was 0.46 (SE = 0.43) and 0.62 (SE = 0.20) for the PED and SNPED models, respectively, which suggests that these traits have different determinisms. While FEC1 represents a measure of the parasite resistance expressed by a naïve lamb, FEC2 is a measure of the parasite resistance expressed by an immunized lamb, and indeed the mechanisms by which these types of animals respond to the infestation are different [43].

The estimate of genetic correlation between ADG0 and ADG1 was unstable due to its sensitivity to the starting values used for its estimation and is not reported.

3.1.4.3 Standard errors

Table 3 shows both the ratio of the standard errors of the estimates obtained from the PED model over the standards error obtained from the SNPED model, which ranged from 1.04 to 2.25 and also the ratio between the width of the confidence intervals of the estimates obtained

from the PED model over the width of the confidence intervals obtained from the SNPED model, which spanned an interval between 0.93 and 4.21. According to the ratio of the standard errors, the SNPED model always converged to more precise values, while according to the ratio of the width of the confidence intervals the estimate of the phenotypic correlation between growth traits and FEC1 obtained from the PED model was slightly more precise.

The results obtained show that including SNP information in the computation of the relationship matrix between individuals can increase the precision of the genetic parameter estimates up to twice the precision obtained by using pedigree information only [221]. The increase in precision can be explained by the fact that SNP information allows to compute more precisely than pedigree information what proportion of genome two individuals actually have in common. The pedigree structure in the data available for this study was not ideal to estimate genetic parameters by pedigree information only because all observations are recorded on a population of animals composed of four groups of half-sibs. Such a structure causes the pedigree-based relationship matrix to predict that within each group of half-sibs all animals share one quarter of the sire's genome. Whereas, the marker-based relationship matrix allows capturing the segregation variance, which means capturing the random deviation of the proportion of genes shared by two individuals around the expected proportion of shared genes according to the pedigree [208].

3.1.5 Conclusions

According to the results obtained by the model including both pedigree and molecular information, the genotypes for growth and for resistance to naïve infestation can be selected for independently. However, the genetic correlations between long-term parasite resistance traits and growth traits were different from 0 and suggest that increasing growth performance in a *H. contortus* free environment could result in more susceptible animals, whereas growth performance in a contaminated environment can be increased while enhancing long-term resistance to *H. contortus*. The two results taken together can also be interpreted as an indication of genotype by environment interaction affecting growth expressed across the two environments [116]. The model that includes pedigree information only converged to similar results, except for the genetic correlation between growth before infestation and faecal egg count during the first infestation which was affected by a very large standard error. The reason for this inconsistency needs further investigation.

This study shows that, when the pedigree is poorly informative using molecular information

and pedigree information jointly result in more precise genetic parameters than using pedigree only.

Figure 1

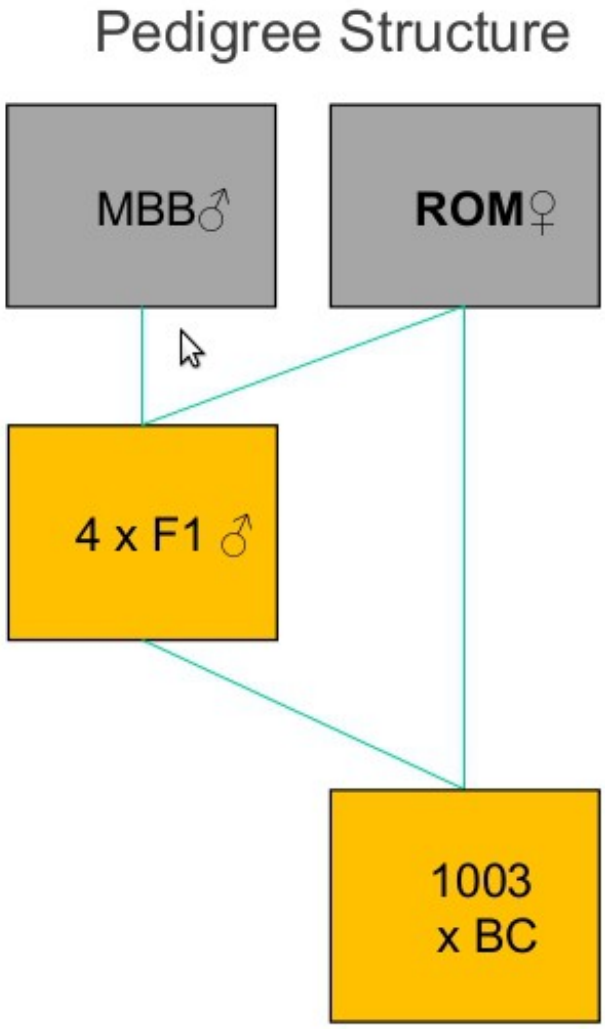


Figure 1 - Schematic representation of the pedigree structure

MBB is the Martinik black Belly population, ROM is the Romane population, FI are the 4 F1 sires resulting from the crossing of MBB sires with ROM dams, BC is the back cross population obtained by mating the F1 sires to ROM dams.

Table 1 - Descriptive statistics of the raw observations

Raw Observations	Number of observations	Average	Standard deviation	Min	Max
$ADG0_{RAW}$ (g/day)	997	293,3	69,4	93.0	504,5
$ADG1_{RAW}$ (g/day)	963	102,8	43,3	-27.0	230.0
$FEC1_{RAW}$ (eggs/g)	987	10494	9827	0	75898
$FEC2_{RAW}$ (eggs/g)	967	2724	4259	0	42667

$ADG0_{RAW}$ is the average daily gain before infestation, **$ADG1_{RAW}$** is the average daily gain during the first infestation, **$FEC1_{RAW}$** is the faecal egg count during the first infestation and **$FEC2_{RAW}$** is the faecal egg count during the second infestation.

Table 2 - Heritabilities, phenotypic correlations and genetic correlations obtained by the SNPED model and the PED model

TRAITS	MODEL	ADG0	ADG1	FEC1	FEC2
ADG0	**SNPED**	0.15 (0.07)*	-	0.11 (0.47)	0.57 (0.38)*
	PED	0.10 (0.08)*	-	-0.52 (1.06)	0.25 (0.85)
ADG1	**SNPED**	-	0.11 (0.06)*	-0.12 (0.58)	-0.54 (0.53)*
	PED	-	0.16 (0.04)*	-0.19 (0.80)	-0.48 (0.67)*
FEC1	**SNPED**	0.02 (0.11)	-0.23 (0.11)*	0.38 (0.04)*	0.62 (0.20)*
	PED	-0.01 (0.15)	-0.24 (0.15)*	0.21 (0.06)*	0.46 (0.43)*
FEC2	**SNPED**	0.04 (0.11)	-0.19 (0.11)*	0.31 (0.08)*	0.48 (0.06)*
	PED	0.01 (0.18)	-0.20 (0.19)*	0.29 (0.14)*	0.55 (0.09)*

ADG0 is the average daily gain before infestation, **ADG1 is** the average daily gain during the first infestation, **FEC1** is the faecal egg count during the first infestation and **FEC2** is the faecal egg count during the second infestation. **SNPED** refers to estimates obtained by using the joint pedigree and molecular information relationship matrix.(SNPED model). **PED** refers to the pedigree-only relationship matrix (PED model). The correlations between ADG0 and ADG1 were much more sensitive than the others to the starting values used for the estimation and to the resampling, and are not presented. **Heritabilities** are on the block diagonal, **genetic correlations** are above the diagonal and **phenotypic correlations** are below the diagonal. The **standard errors** of the estimates are between parenthesis. The superscript * marks the estimates which were significantly different from 0.

Table 3 - Ratios between the precisions estimators obtained by the PED model over those obtained by the SNPED model

TRAITS	ADG0	ADG1	FEC1	FEC2
ADG0	^{SE}1.12 ^{CI}1.42	-	^{SE}2.24 ^{CI}1.26	^{SE}2.25 ^{CI}4.21
ADG1	-	^{SE}1.23 ^{CI}2.37	^{SE}1.38 ^{CI}1.30	^{SE}1.26 ^{CI}1.57
FEC1	^{SE}1.37 ^{CI}0.93	^{SE}1.33 ^{CI}0.98	^{SE}1.72 ^{CI}1.5	^{SE}2.22 ^{CI}2.2
FEC2	^{SE}1.73 ^{CI}1.41	^{SE}1.68 ^{CI}1.30	^{SE}1.76 ^{CI}1.25	^{SE}1.67 ^{CI}1.9

ADG0 is the average daily gain before infestation, **ADG1** is the average daily gain during the first infestation, **FEC1** is the faecal egg count during the first infestation and **FEC2** is the faecal egg count during the second infestation. The table shows the both the values of the ratio between the approximate standard errors (marked with superscript SE) and the ratios of the width of the confidence intervals obtained from the empirical distribution (marked with superscript CI) of the parameters estimates obtained by using either pedigree relationship matrix or the joint pedigree and molecular information relationship matrix. The ratios on the diagonal refer to the **heritability** estimates. The ratios above diagonal refer to the **genetic correlation** estimates. The ratios below the diagonal refer to the **phenotypic correlation** estimates. The correlations between ADG0 and ADG1 were much more sensitive than the others to the starting values used for the estimation and to the resampling, and are not presented.

3.2 Article 2

The aim of the second study was to explore the eventual occurrence of QTLs affecting parasite resistance traits in creole goat; therefore, to address the issue of whether to include the assistance of molecular markers in predicting the selection response or not. The literature concerning QTL detection for parasite resistance in goat is quite limited, compared to what has been published in the same topic on sheep.

The estimated genetic parameters were limited to the heritabilities of the traits because of the small number of observations available. The family structure and the depth of the pedigree limited the inventory of statistical tools to running a multiple testing procedure for each SNP locus – in order to test the significance of each marker loci allelic substitution effect. Each test was performed separately by applying a linear mixed animal model featuring the gene content at the locus of interest as a fixed effect and the genetic relationship matrix to account for the population structure. The significance thresholds for multiple testing were computed chromosome-wise and following a method which allows for taking into account the linkage disequilibrium between adjacent marker loci, instead of assuming their tests to be independent [179].

The results obtained from this study convey information on whether the selection response prediction should consider marker assisted selection as a feasible option.

QTLs for faecal egg counts and packed cell volume detected during two subsequent experimental infestations with *Haemonchus contortus* on Creole goats.

F. Assenza [1,a], C. Moreno[1], J.M. Elsen[1], C. Barbier[2], C. Anaïs.[2], J.-C., Bambou[2], N.Mandonnet[3].

[1] *INRA UMR1388, GenPhySe, 24 chemin de Borde-Rouge, 31326, Castanet-Tolosan, France*

[2] *INRA, UE1294, Plateforme Tropicale d'Expérimentation sur l'Animal, F-97170 Petit-Bourg, France*

[3]*INRA, UR0143, Unité de Recherches Zootechniques, F-97170 Petit-Bourg, France* .

Corresponding author: Fabrizio Assenza. Email: carole.moreno@toulouse.inra.fr

Short title: genetics of resistance to *H.contortus* in creole goat

3.2.1Abstract

Haemonchosis is a gastrointestinal parasitosis greatly impacting small ruminant production worldwide. Due to the recent aggravation of anthelmintic resistance, alternative strategies to the anthelmintic treatments are being studied in order to reduce the economic impact of *H.contortus.* Genetic selection being one of the most promising, its efficacy might be enhanced by using molecular markers. In this study, we report a genome-wide association study for detecting QTL affecting faecal egg count and packed cell volume. The analysis was performed on the basis of observations collected from two subsequent experimental infestations on creole goat drenched twice with 10000 L3 larvae of *H.contortus.* The heritabilities of faecal egg counts resulted moderate, those of packed cell volumes resulted moderate-high. A QTL on chromosome 6 would affect faecal egg count during both the unprimed and the primed infestation. Faecal egg count expressed during the immune response would also be affected by QTLs on chromosomes 10 and 25 - the latter being highly significant. Faecal egg count expressed during the primed immune response would also be affected by another QTL on chromosome 8, as reported previously. Packed cell volume during the primed infestation would be affected by two QTLs located on chromosomes 22 and 25, respectively.

Keywords: Faecal egg count, Packed cell volume, QTL, heritability, Goat

3.2.2 Introduction

H.contortus is one of the most problematic gastrointestinal nematodes in small ruminant industry [2], due to its remarkable adaptability to different environments [7] and to its heavy haematophagy [6]. Indeed the estimates of the economic loss it causes to small ruminant production systems worldwide reaches values on the order of hundreds of millions of dollars per year [2]. Gastrointestinal nematodes are particularly problematic in the tropical areas [222] for two main reasons. First, the warm and humid climate favours the accomplishment of the external life cycle stages of these parasites [223]. Second, the production systems in tropical climates are typically outdoors and feature higher exposure to harsh environmental conditions (for example heat stress) than the indoors systems found mostly in the temperate regions. Gastrointestinal nematodes have been normally controlled by using anthelmintic treatments, however this strategy is doomed to lose efficacy in the long term [191]. The occurrence of anthelmintic resistance have indeed been reported for many compounds and in many countries [191, 222]. Other control strategies (such as vaccines, biological control, farm management strategies [82, 84, 202]) have been investigated but none of them was shown to be effective in practice, except for farm management. However, the efficacy of these control strategies can be enhanced by coupling it with genetic selection, which features some attractive characteristics specially for low input enterprises. Indeed, genetic selection allows to enhance the resistance (or resilience) to gastrointestinal nematodes in a livestock population within a given production context. It complies with low input production systems better than anthelmintic treatments because the modifications of the genetic makeup of the population under selection naturally persist in time by genetic inheritance, whereas the latter strategy requires new investment each and every time the treatment is required. Indeed, enhancing the natural resistance to helminths of the animals makes it possible to reduce, at least, the yearly number of anthelmintic treatments required to support a certain production level. Despite applying a selective pressure for parasite resistance on the host results inevitably in applying a selection pressure on the parasite population as well, the selection intensity on the parasite derived from genetic selection for parasite resistance on the host seems likely insufficient for allowing the parasites' adaptation to the sheep selected for parasite resistance [198]. New opportunities to enhance genetic selection by marker assisted design have recently spawned from the recent developments in genotyping technology, which allow relatively inexpensive SNP chips for genotyping individuals on increasingly dense marker maps. This has made QTL detection more affordable and hence marker assisted selection to be possibly applied in

practice. Previous studies on small ruminants have reported several QTLs for resistance to *H.contortus* both in sheep and goat affecting traits such as faecal egg count, packed cell volume and eosinophils count. The results of de la Chevrotière et al. [224] on QTL affecting Creole goat in tropical climate, some QTL are reported to affect faecal egg count on chromosomes: 8, 22 and 26. In the review by Bishop & Morris [90] and Dominik [93] they report some QTL for FEC to be located on the following chromosomes of sheep: 1, 6, 19 and 20. A more recent study on sheep by Riggio *et al.* [225] confirm a QTL for parasite resistance on chromosome 6. The literature review reports essentially more studies conducted on sheep, with results obtained from different populations of sheep sometimes overlapping. The literature about goat is rather limited, de la Chevrotière [224] being the only one we found reporting results from observations on Creole goat and Bolormaa et al. [226] reporting a QTL on chromosome 23 for faecal egg counts and eosinophil count on Angora goat. Both of these studies are based on micrositellites, whereas in the present study we report the QTLs affecting faecal egg count and packd cell volume detected by a GWAS based on a high-density SNPchip. The phenotypes are collected on a herd of creole goat in Guadeloupe after two subsequent experimental infestations with *H.contortus. T*he estimates of both the heritiability of each trait and of the phenotypic correlations between them are also reported.

3.2.3 Material and methods

3.2.3.1Phenotypic information

Creole goat is an indigenous breed of the Antilles islands. It is mostly exploited for meat production and has been classified among the goat breeds featuring genetic variability on resistance to parasites [227]. In the production flock of INRA-PTEA in Guadeloupe F.W.I. (16° 15' 0" N / 61° 34' 59" W) Creole goats are routinely indexed for resistance to gastro-intestinal nematodes after mixed infestation at pasture. Two extreme groups of creole goat were selected in order to maximize their difference on faecal egg count. Six resistant bucks and 6 susceptible ones were mated to 55 resistant and 51 susceptible does respectively, according to their breeding value on FEC at 11 months of age. The divergence between the 2 groups of parents reached 1.1 genetic standard deviation between sires and 0.9 between does. The matings and the raising of the kids under selection took place at the Iexperimental flock of INRA-PTEA in Guadeloupe F.W.I., during five cohorts.. In order to keep track of the genealogy the does were kept in cages during the periparturient period. Kids were weaned at 3 months of age and were reared indoors in parasite free environment in order to ensure that all

animals did not encounter any gastrointestinal strongyle before the infestation protocol began. They underwent two experimental infestations with *H.contortus* larvae. They were given a first experimental infestation of 10000 L3 larvae of *H.contortus*. Faecal samples for recording faecal egg count and blood samples for recording packed cell volume were collected weekly until day 42 after infection. Then the animals were treated by anthelmintic and entered a second control period, during which faecal egg counts were recorded again and lasted until all animals featured zero faecal egg count and recovered from the parasitosis to a normal body score. Then the second period of infestation began, following the same protocol as the first infestation period, the same larval dose was used and observations on the same traits were collected weekly. After six weeks the animals were treated with anthelmintics (ORAMEC®). The observations on each trait on each animal were averaged over each period, starting from the second observation because the prepatent period of haemonchosis is normally 15-21 days [6]. Both the averages of faecal egg counts during the first and during the second infestation were transformed by taking their square root in order to approach their distribution closer to normality. The resulting variables were trimmed by removing the outliers which were over 3 standard deviations far from the average. The variables obtained after averaging, transforming and trimming from the observations on faecal egg count and packed cell volume during the first infestation period were called: FEC1 and PCV1, respectively. Those obtained from the observations collected during the second infestation period were called accordingly: FEC2 and PCV2.

3.2.3.2 Genetic information

The genealogical information on the pedigree of the animals featuring phenotypic observation did not go deeper than one generation, i.e. only the parents of the phenotyped kids were included.

Individual blood samples were collected after weaning. DNA extraction and genotyping have been performed at LABOGENA facility (WWW.labogena.fr). The samples were genotyped with the Illumina goat SNP50 Beadchip (Illumina, Inc., San Diego, CA). The marker loci used for the QTL detection were only those which complied with having minor allele frequency (MAF) higher than 0.2 and call rate higher than 0.99. Such stringent thresholds were chosen in order to compensate for small size of the population, which might increase the likelihood of detecting false positives on low MAF alleles.

3.2.3.3 Heritabilities

Heritabilities were computed by applying a multivariate linear mixed model including both pedigree and SNP information which read as follows [149]:

$$\begin{bmatrix} FEC1 \\ PVC1 \\ FEC2 \\ PVC2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ 0 & 0 & X_3 & 0 \\ 0 & 0 & 0 & X_4 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 & 0 & 0 \\ 0 & Z_2 & 0 & 0 \\ 0 & 0 & Z_3 & 0 \\ 0 & 0 & 0 & Z_4 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

Where the matrix on the left side of the equation contains the vectors of the variables described above, the $\mathbf{X_i}$ matrices are the incidence matrices connecting the observations of each vector to its fixed effects pattern, $\mathbf{b_i}$ are the vectors of the estimated fixed effects at each trait. The fixed effects for this model were the contemporary group, the sex, the age at the beginning of the infestation and the factor identifying which one of the genetic selection line each animal belonged to (either resistant or susceptible). The $\mathbf{Z_i}$ matrices are the incidence matrices connecting the observations of each vectors to the relative random effect, $\mathbf{a_i}$ are the vectors of the estimated random effects for each trait and finally $\mathbf{e_i}$ are the vectors of residuals for each observation of each vector. The variance structure of the model was: $\mathrm{VAR}\begin{bmatrix} \mathbf{A} \\ \mathbf{E} \end{bmatrix} = \begin{bmatrix} \mathbf{G}\otimes\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}\otimes\mathbf{I} \end{bmatrix}$, where **A** and **E** are the vectors of random effects and the residuals described above, **I** is an identity matrix and **T** is the genetic relationship matrix computed according to VanRaden 2008 [142]: $\mathbf{T}=0.95\mathbf{S}+0.05\mathbf{P_{22}}$. $\mathbf{P_{22}}$ is the a submatrix of the classic relationship matrix computed by using the pedigree information [215]. **S** is the genomic relationship matrix between the genotyped animals, it is computed by following one of the methods proposed by VanRaden [142]. Let **M** be a matrix with number of rows equal the number of individuals and number of column equals the number of SNP loci. The elements of **M** codes the genotype of a given animal has at a given marker locus as follows: -1 if homozygous for one allele, 0 if heterozygous and 1 if homozygous for the other allele. The elements of **M** are then centred, in order to having 0 average allele substitution effect, by subtracting to each column of **M** its respective column of matrix **P**. The latter features the element of each column vector equal to $2(p_i - 0.5)$, where p_i is the allelic frequency of the

second allele. Let **Z** be the centred **M** matrix, then the **S** matrix is finally computed as: $\mathbf{S}=\frac{\mathbf{ZZ'}}{2\sum p_1(1-p_1)}$. **G** and **R** are the variance/covariance matrices of the vectors **A** and **B**:

$$\mathbf{G}=\begin{bmatrix} \sigma^2_{a1} & \text{symmetric} & & \\ \sigma_{a21} & \sigma^2_{a2} & & \\ \sigma_{a31} & \sigma_{a32} & \sigma^2_{a3} & \\ \sigma_{a41} & \sigma_{a42} & \sigma_{a43} & \sigma^2_{a4} \end{bmatrix};\quad \mathbf{R}=\begin{bmatrix} \sigma^2_{e1} & \text{symmetric} & & \\ \sigma_{e21} & \sigma^2_{e2} & & \\ \sigma_{e31} & \sigma_{e32} & \sigma^2_{e3} & \\ \sigma_{e41} & \sigma_{e42} & \sigma_{e43} & \sigma^2_{e4} \end{bmatrix};$$

σ^2_{ai} and σ^2_{ei} are the genetic and residual variances of each trait, σ_{aij} and σ_{eij} are the genetic and residual covariances between traits. Heritabilities were computed as $h^2_i=\frac{\sigma^2_{ai}}{\sigma^2_{ai}+\sigma^2_{ei}}$ [204], with standard error computed according to the general asymptotic approximation for the variance of a ratio between two variance estimates [139]:

$$SE\left(\frac{\sigma^2_n}{\sigma^2_d}\right)=\sqrt{\left(\frac{\sigma^2_n}{\sigma^2_d}\right)^2\left(\frac{VAR(\sigma^2_n)}{\sigma^4_n}+\frac{VAR(\sigma^2_d)}{\sigma^4_d}-\frac{2\text{COV}(\sigma^2_n,\sigma^2_d)}{\sigma^2_n\sigma^2_d}\right)}$$

, where σ^2_n and σ^2_d are the estimates at the numerator and at denominator, respectively; VAR(.) and COV(.,.) are the variance and the covariance between the estimates between parenthesis. The phenotypic correlations was computed as $r_{pij}=\frac{(\sigma_{aij}+\sigma_{eij})}{\sqrt{(\sigma^2_{ai}+\sigma^2_{ei})(\sigma^2_{aj}+\sigma^2_{ej})}}$ [204], with standard error computed according to the general asymptotic approximation for the variance of a correlation estimate between two normally distributed random variables i and j:

$$SE(r_{ij})=\sqrt{r^2_{ij}\left(\frac{VAR(\sigma^2_i)}{4\sigma^4_i}+\frac{VAR(\sigma^2_j)}{4\sigma^4_j}+\frac{VAR(\sigma_{ij})}{\sigma^2_{ij}}\right)}$$

$$+\sqrt{r^2_{ij}\left(\frac{2\text{COV}(\sigma^2_i,\sigma^2_j)}{4\sigma^2_i\sigma^2_j}-\frac{2\text{COV}(\sigma^2_i,\sigma_{ij})}{2\sigma^2_i\sigma_{ij}}-\frac{2\text{COV}(\sigma^2_j,\sigma_{ij})}{2\sigma^2_j\sigma_{ij}}\right)}$$

The computation was performed by ASReml.3 [139]

3.2.3.4 QTL detection

The model applied for QTL detection is based on the hypothesis that, among the genes influencing the phenotypes under study, some of them might have a major impact on it. QTL detection was performed following the method of Kang *et Al.* 2010 [177], which allows the estimation of the SNP effect by taking into account the population structure of the data set used for its estimation. The method applies the following linear mixed animal model:

$$\mathbf{y}_i = \mu_i + \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{Z}_i\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i$$

y_i is the vector of observations on one of the i trait described above and μ_i is its overall mean. $\mathbf{X_i}$ is the incidence matrix of factors and co-variables featuring the fixed environmental effects, the age and the SNP genotype at a given locus. The SNP genotype is coded as the number of copies of one of the alleles at the locus under study hence it features values 0 1 and 2 if the animal is homozygous for the allele 1, heterozygous or homozygous for allele 2, respectively. $\boldsymbol{\beta_i}$ is the vector of the estimated fixed effects. $\mathbf{Z_i}$ is the incidence matrix of the random animal effects and $\boldsymbol{\alpha_i}$ is its vector of estimated breeding values. The variance structure reads $\mathrm{VAR}\begin{bmatrix}\boldsymbol{\alpha}_i \\ \boldsymbol{\varepsilon}_i\end{bmatrix} = \begin{bmatrix}\mathbf{A}\,\sigma^2_{ai} & 0 \\ 0 & \mathbf{I}\,\sigma^2_{ei}\end{bmatrix}$, where σ^2_{ai} is the estimated genetic variance of the vector $\boldsymbol{\alpha_i}$, σ^2_{ei} is the estimated residual variance of the vector $\boldsymbol{\varepsilon_i}$, $\mathbf{I}$ is an identity matrix and $\mathbf{A}$ is the genetic relationship matrix computed by using the pedigree information. The effects of the SNPs were estimated one by one, by testing the null hypothesis: "the estimated effect of the SNP at locus i on the response variable is not different from zero." This test was performed by fitting one different SNP locus at a time for each trait, therefore the significance of each test must be interpreted by methods fitting a multiple testing scenario. The threshold for the multiple testing was obtained according to Muller *et al* 2011 [179], which allows to take into account the linkage disequilibrium between markers rather than the more conservative Bonferroni correction as in Kang *et al* 2010 [177], which treats the marker loci in the same chromosome to be independent.

Three chromosome-wise thresholds, computed by taking into account the linkage disequilibrium between the markers on the same chromosome [179] are reported in the result section: the low significance threshold, featuring a type one error probability equal to 0.1; the

significance threshold, featuring a type one error probability equal to 0.05 and finally a strong significance threshold, featuring a type one error probability equal to 0.01. These computations were performed by the Muller package released by the department of Animal Genetics of INRA [228]. The confidence intervals were computed according to Li 2011 [229].

3.2.4 Results and discussion

3.2.4.1 Descriptive statistics

After trimming the dataset, the number of animals featuring observations on both the traits under study resulted in the following figure: 174 animals had observations on the FEC1 and PVC1, 124 of which also had observations on FEC2 and PVC2. All these animals also featured genotype information and the number of SNP markers left for the analysis after the quality checks was equal to 34336. Table 1 summarizes the descriptive statistics of the traits under study.

The overall average of FEC2 was lower than the overall average of FEC1 ($p<0.0001$) whereas the average PVC2 was higher than PVC1 ($p<0.0001$). The reduction of the overall egg excretion observed during the second infestation can be explained by assuming that the animals have had enough time to develop a specific immune response to the parasite. The acquired immune response is indeed more effective than the innate immune response in limiting the parasite infestation and symptoms [31]. This assumption is also supported by the result observed on packed cell volume, which appears to be higher during the second infestation. FEC is a measure of the resistance of the individual to gastrointestinal nematodes infestation and low FEC is by definition referred to as an indication of resistance to gastrointestinal parasites. On the other hand, PCV is a measure of resilience and high values of PCV indicate that the individual is able to compensate for the blood loss caused by the parasite [110]. Despite this compensation, called erythropoiesis, can buffer the blood loss quite readily, it can be exhausted. The length and quality of erythropoiesis are a symptom of how much resources are available and how much of them are allocated to it. Resources can be available either from a high nutritional level diet (the effect of which is not investigated in this study because all the animals were fed the same diet) or from the ability of the goat to limit the blood loss by interfering with the parasite's activities [230]. The latter is measured by the FEC observations, which have lower average contemporary to the higher average of PCV.

The estimates of their phenotypic correlations (Table 2) which equal -0.21 (SE 0.09) for FEC1 and PCV1 during the first infestation and -0.41 (SE 0.08) for FEC2 and PCV2 during

the second suggests that the two phenotypes could be related. Previous studies have already reported the same phenotypic correlation [115], which can be easily explained by the marked haematophagia typical of *H.contortus*. The availability of nutrients, protein specially, has been shown to influence strongly the symptoms of the disease [102] which suggests that the availability of resources plays a major role in the response of the host to the parasite. To mount an immune response is indeed a process demanding large nutrients availability, both of energy and of amino acids. In this perspective it is possible to understand the role that packed cell volume plays in this correlation as the transporter of the amount of oxygen necessary to catalyse the oxidative biochemical reactions for producing and using energy [231]. Energy is used by any synthesis process in the organism, including erythropoiesis and the mounting of an immune response. The latter could explain how the availability of oxygen, PCV, would be negatively correlated to the number of eggs found in the faeces, FEC, because faecal egg count is a measure of both how many adult worms have escaped the immune response [39] and to what extent their fitness is impaired by the immune response [232].

A negative phenotypic correlation is also estimated between periods of infestation, as both the correlation estimates of PVC1 with FEC2 and of PVC2 with FEC1 are negative: -0.24 (SE 0.1) and -0.25 (SE 0.09), respectively. Moreover both the phenotypic correlations of FEC1 with FEC2 and of PVC1 with PVC2 are positive: 0.40 (SE 0.09) and 0.75 (SE 0.04), respectively. The estimated obtained for these correlation suggest that some repeatability of the phenotype can occur. This means in practice that the animals which featured higher (lower) phenotypes during the first infestation are likely to show the same feature during the second infestation. The estimate obtained between FEC1 and FEC2 indicates a moderate correlation, which can be explained by the fact that non-specific immune response and the specific immune response are two different mechanisms [39]. The estimate of the phenotypic correlation between PCV1 and PCV2 indicates instead a high correlation. Considering the more general role of the packed cell volume in the response to the parasite, it makes sense to imagine that the influence of PCV in sustaining the non-specific immune response and its role in sustaining the specific immune response do not differ enormously.

The contrast between the resistant group's and the susceptible group's measurements on FEC1 and on FEC2 results in the resistant group featuring significantly lower ($p<0.05$) faecal egg count during both infestation periods, suggesting that the selection process which the two groups of animals were obtained from, had a significant impact on the average feacal egg counts both during the unprimed and during the primed infestations. No significant difference

was found on the expression of packed cell volume instead. In fact, the selection protocol included faecal egg count only.

3.2.4.2 Heritabilities

The heritability estimates obtained from our data-set are in the same range as those reported from previous studies [206, 227], confirming that the estimates for the traits under study are from moderate to high (Table 2). This result is also in accordance with previous studies exploring the genetic variability for faecal egg count which reported the existence of breed differences [233].

We could indeed observe some difference in the overall averages of both FEC1 and FEC2 between the animals descending from the susceptible line and those descending from the resistant line. During both the first and the second infestation, the former group expressed higher average faecal egg count than the latter. This difference can be interpreted as an indication that genetic selection can actually influence the observed FEC phenotype both for unprimed animals and primed animals.

This trend was not observed for the packed cell volume. Nevertheless, both its heritability estimate obtained in this study and previous estimates [206] indicate that a moderate genetic variation for selecting on this trait as available as well. Selecting on PCV would have the advantage that its measurement is more easy to be automated than FEC, however the former trait is probably affected by more factors than the latter and this could make it somewhat less related to resistance than a direct measure of faecal egg count [115]. The estimates obtained from the observations during the second infestation specially, are slightly higher than what is found in the literature [115]. Three factors might have had an impact on the magnitude of the heritabilities: the phenotypic variation of the sample might have been reduced both by the fact that we computed the heritabilities on the averaged observations and by the reduction of the environmental variance due to the experimental settings. Also age might have played a role, as reported on previous studies [227].

3.2.4.3 QTL detection

A total of 13 signals across the four traits were strong enough to pass the low significance threshold at alpha=0.1. These signals came from chromosomes 6, 8, 10, 16 and 25 for the faecal egg counts and from chromosomes 11, 22 and 25 for packed cell volumes. One of these SNP locus has greater estimated effect than the high significance threshold at alpha=0.01; seven SNP loci had greater estimated effect than the significance threshold at alpha=0.05 and five SNP loci had greater estimated effect than the low significance threshold at alpha=0.1. Table 3 reports the SNP loci with greater estimated effect than the threshold at alpha=0.05 and at alpha=0.01.

The analysis on faecal egg counts resulted in the following picture. Both FEC1 and FEC2 would be affected by three SNP loci located on chromosome 6 whose estimated effect resulted significant at alpha=0.05. As suggested by the close positions of these SNP loci affecting FEC1 on chromosome 6, it is more likely that both SNP detect the signal coming from a single QTL, due to the linkage disequilibrium between them. The confidence interval for the position of the QTL affecting FEC1 and the confidence interval of the position of the QTL affecting FEC2 on chromosome 6 suggest that the same QTL might affect both traits, because their confidence intervals overlap. The other SNP loci affecting significantly FEC1 and FEC2 were located in different chromosomes (Figure 1). FEC1 featured one more SNP locus on chromosome 10 at alpha=0.05, close by an other SNP locus which results significant if the threshold is pushed down to alpha=0.1. At the low significance threshold we also found one SNP locus affecting FEC1 on chromosome 16 and one on chromosome 25. The latter chromosome also bears the only SNP which passed the high significance threshold at alpha=0.01, close to the SNP locus found at alpha=0.1 (figure 2). By taking into account the positions and relative confidence intervals of the SNPs significantly affecting FEC1 on chromosome 6, 10 and 25, we suppose the number of putative QTL affecting FEC1 to equal one on chromosome 6, one on chromosome 10 and one on chromosome 25. The last putative QTL affecting faecal egg counts would be located on chromosome 8 and affects FEC2. It is detected as a couple of SNP loci next to each other and associated to allelic substitution effects which result significant at alpha=0.05 and alpha=0.1, respectively. These results are partially in accordance with de la Chevrotiere *et Al* [224], which reports a significant QTL effecting faecal egg counts on chromosome 8 in a separate group of the same population of goats. Despite this study has been conducted on the same creole breed and in similar climate conditions, our results do not indicate the presence of any SNP with estimated effect

significantly different from zero corresponding to those they report on chromosome 22 and 26. Furthermore, they did not report any of the signals we detected from chromosomes 6, 10, 16 and 25. It is worth noticing that the literature review on QTL detection for parasite resistance in goat is not as rich as that on the same subject concerning sheep. Comparing our results to the QTLs for resistance to *H.contortus* reported in sheep reveals that also chromosome 6 was already resulted as bearing significant SNP loci for faecal egg count [226]. To the best of our knowledge, the putative QTLs for resistance to *H.contortus* we found on chromosomes 10 and 25 were not previously reported in sheep studies either.

Despite the results obtained from the phenotypic analysis showed no significant difference between the overall average PCV of the resistant group and of the susceptible group for neither of the infestation periods, according to our results, chromosome 25 would bear a QTL for the trait PCV2 at alpha=0.05 (figure 2), whose confidence interval do not overlap with the confidence interval for the position of the QTL for FEC1. At the same threshold a significant SNP on chromosome 22 was also found for PCV2. Pushing down the threshold to alpha=0.1 reveals a QTL on chromosome 11 for PCV1. None of these QTLs were reported on previous studies, neither in goat or sheep, which report instead signals of QTLs affecting packed cell volume from chromosome 5 in goat and chromosome 1 in sheep.

Given the small size of this data set, it can be considered an encouraging result as the limited number of observations would only allow for QTL with large effect to be found. Indeed, by summing up for each trait the percentages of phenotypic variance explained by the putative QTLs, obtained by averaging the estimated effects of multiple SNPs on the same chromosome and summing up across chromosomes, results equal to: 12.43% for FEC1, 14.16% for FEC2 and 10.05% for PVC2.

Finally, in order to better interpret any result on QTL detection for both traits related to immunity, such as Faecal egg count, and on traits related to erythropoiesis, such as packed cell volume, it is important to keep in mind that normally the QTL dectected by any study is strictly related to the alleles segregating within the finite population under study. Furthermore, such traits are more likely determined by the complex interplay of a large number of genes expressed differently by various cellular populations, because what is known about the biochemistry of the possible pathways involved suggests so [39, 230]. Nevertheless some of the QTLs detected in this study overlap with the chromosomes reported on previous studies, indeed on the same breed of goat [224].

3.2.5 Conclusions

The results obtained from the infinitesimal model on the estimates of the heritabilities of faecal egg counts and packed cell volume suggest that a moderate to high potential for genetic selection is available on these traits. The design of the experiment resulted in slightly higher estimates than what is found in literature.

The putative QTLs detected in this study suggest that marker assisted selection could be considered for enhancing an eventual breeding scheme for increasing parasite resistance. Chromosome 6 would bear a QTL affecting both faecal egg count during the unprimed infestation and the primed infestation. Faeal egg count during the unprimed infestation would also be affected by two more QTLs: one on chromosome 10 and one on chromosome 25. The analysis on packed cell volume during the unprimed infestation resulted in a weak signal on chromosome 11 which passed the low significance threshold only. During the primed infestation we found another putative QTL located on chromosome 8 affecting faecal egg count, which was previously reported by another study on creole goat. Packed cell volume during the primed infestation would be affected by a QTL on chromosome 22 and one on chromosome 25, the latter is likely not to be the same as the QTL affecting faecal egg count during the unprimed infestation.

Table 1 Descriptive statistics of the variables under study. FEC1_raw and FEC2_raw are the crude avrages of faecal egg counts during the first and the second infestation, respectively. FEC1 and FEC2 are their log-transforms. PCV1 and PCV2 are the averages of packed cell volume during the first and second infestation, respectively.

Trait	N	Average	RSD	Min – Max
FEC1_raw	174	2941	2736	6.447 – 13075
FEC2_raw	124	1117	1605	15.00 – 8253
FEC1[1]	174	48.03	25.27	2.539 – 114.3
FEC2[1]	124	27.37	19.29	3.873 – 90.85
PCV1[2]	174	26.05	3.571	17.42 – 35.43
PCV2[2]	124	27.89	4.151	16.00 – 40.25

Trait = name of the variable; N = number of observations; Average = Mean of the variabe; RSD= residuals standard error; MIN – MAX= minimum value – maximum value

[1] The resistant group featured significantly lower ($p < 0.05$) average faecal egg count than the susceptible group within both the unprimed (FEC1) and the primed infestation (FEC2). The average faecal egg count during the primed infestation (FEC2) was significantly lower ($p < 0.01$) than the one measured during the unprimed infestation (FEC1).

[2] The contrast of average packed cell volume between the resistant and the susceptible genetic line did not differ significantly in neither of the infestation periods. The average packed cell volume during the primed infestation (PCV2) was significantly lower ($p < 0.01$) than the one measured during the unprimed infestation (PCV1).

Table 2 Heritabilities (diagonal) and phenotypic correlations (below diagonal) between the phenotypes under study, the standard errors of the estimates are between parenthesis.

Trait	FEC1	PCV1	FEC2	PCV2
FEC1	0.30 (0.21)			
PCV1	-0.21 (0.09)	0.47 (0.21)		
FEC2	0.40 (0.09)	-0.24 (0.10)	0.25 (0.24)	
PCV2	-0.25 (0.09)	0.75 (0.04)	-0.41 (0.08)	0.35 (0.25)

Trait = name of the variable; FEC1 = faecal egg count during the first infestation; PCV1 = packed cell volume during the first infestation; FEC2 = faecal egg count during the second infestation; PCV2 = packed cell volume during the second infestation.

Table 3 Marker loci with significant ($p < 0.05$) estimated allelic substitution effect. FEC1 and FEC2 are the faecal egg counts during the first and second infestation, respectively. PCV2 is the packed cell volume during the second infestation.

Trait	Chr	LCI	Position	UCI	MAF	alpha	Variance of y (%)
FEC1	**6, 10, 25**						**TOT=12.43**
	6	36793909	108380000	110575626	0.43	5%	3.46
	6	36793909	103570000	110575626	0.22	5%	
	10	4696745	19807678	94495372	0.39	5%	3.91
	25	4320451	4379659	4535433	0.44	1%	5.07
FEC2	**6, 8**						**TOT=14.16**
	6	6020308	68025618	100710196	0.27	5%	5.60
	8	26547948	104110000	107240140	0.40	5%	8.56
PCV2	**25, 22**						**TOT=10.06**
	22	19447368	51248980	51333833	0.27	5%	5.14
	25	4580345	26853182	28844607	0.29	5%	4.92

Trait = name of the variable; Chr = Chromosome; LCI = lower confidential interval of the position of the putative QTL; Position = position of the putative QTL; UCI = upper confidential interval of the position of the putative QTL; MAF = minor allele frequency of the marker locus; alpha = most stringent significance threshold at which the estimated allelic substitution effect resulted significant; Variance of y(%) = percentage of trait's variance explained.

Figure 1 Manhattan plot of chromosome 6 for faecal egg counts.

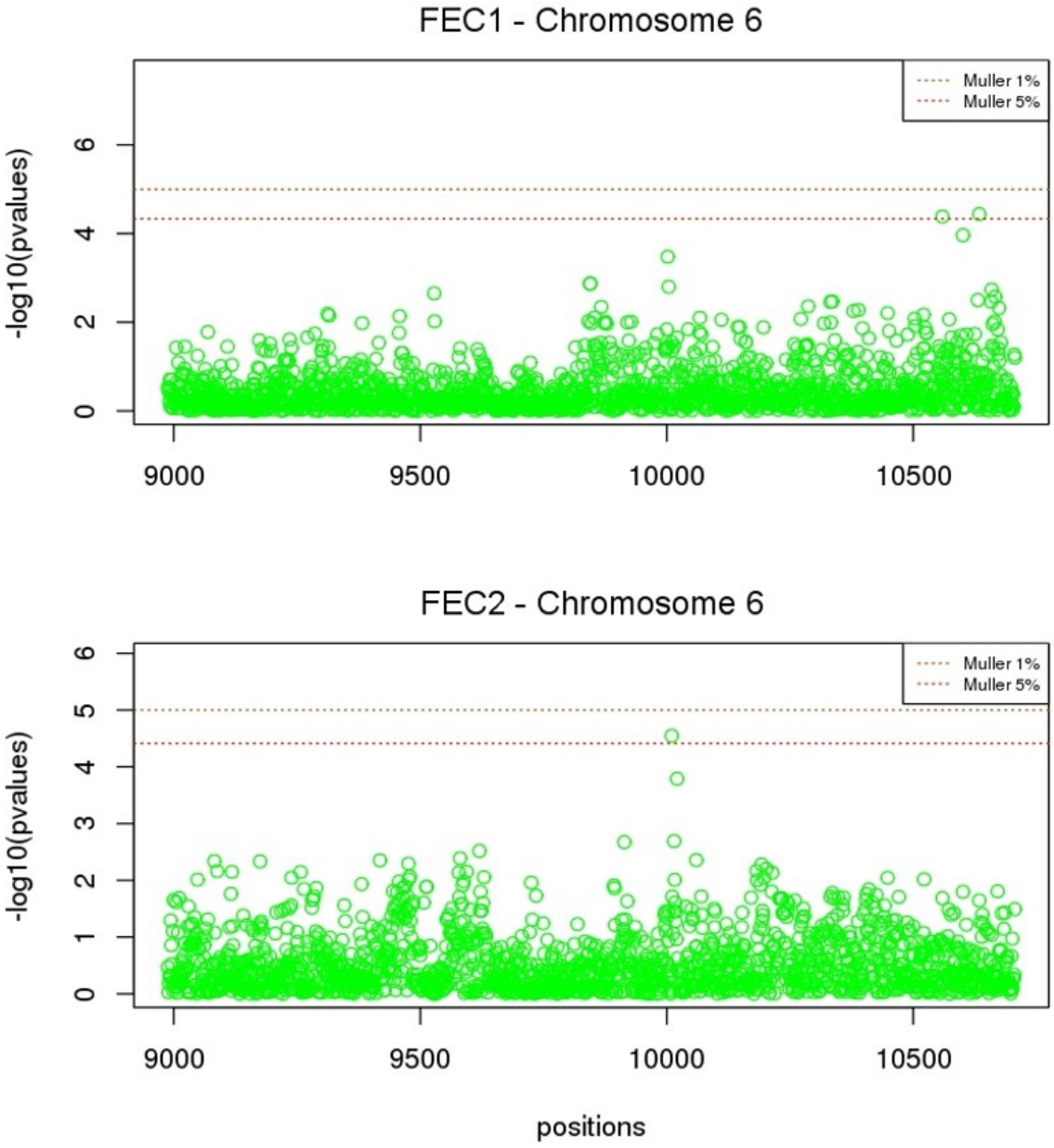


Figure 1 QTLs detected on chromosome 6 for faecal egg count during the first and second infestation, FEC1 (above) and FEC2 (below) respectively. Position = position relative to the SNP rank on the chromosome; -log10(pvalues) = likelihood ration test for the significance of the estimated effect; Muller 1% = significance threshol at alpha = 0.01; Muller 5% = significance threshold at alpha = 5%.

Figure 2 Manhattan plot of chromosome 25 for faecal egg count and packed cell volume.

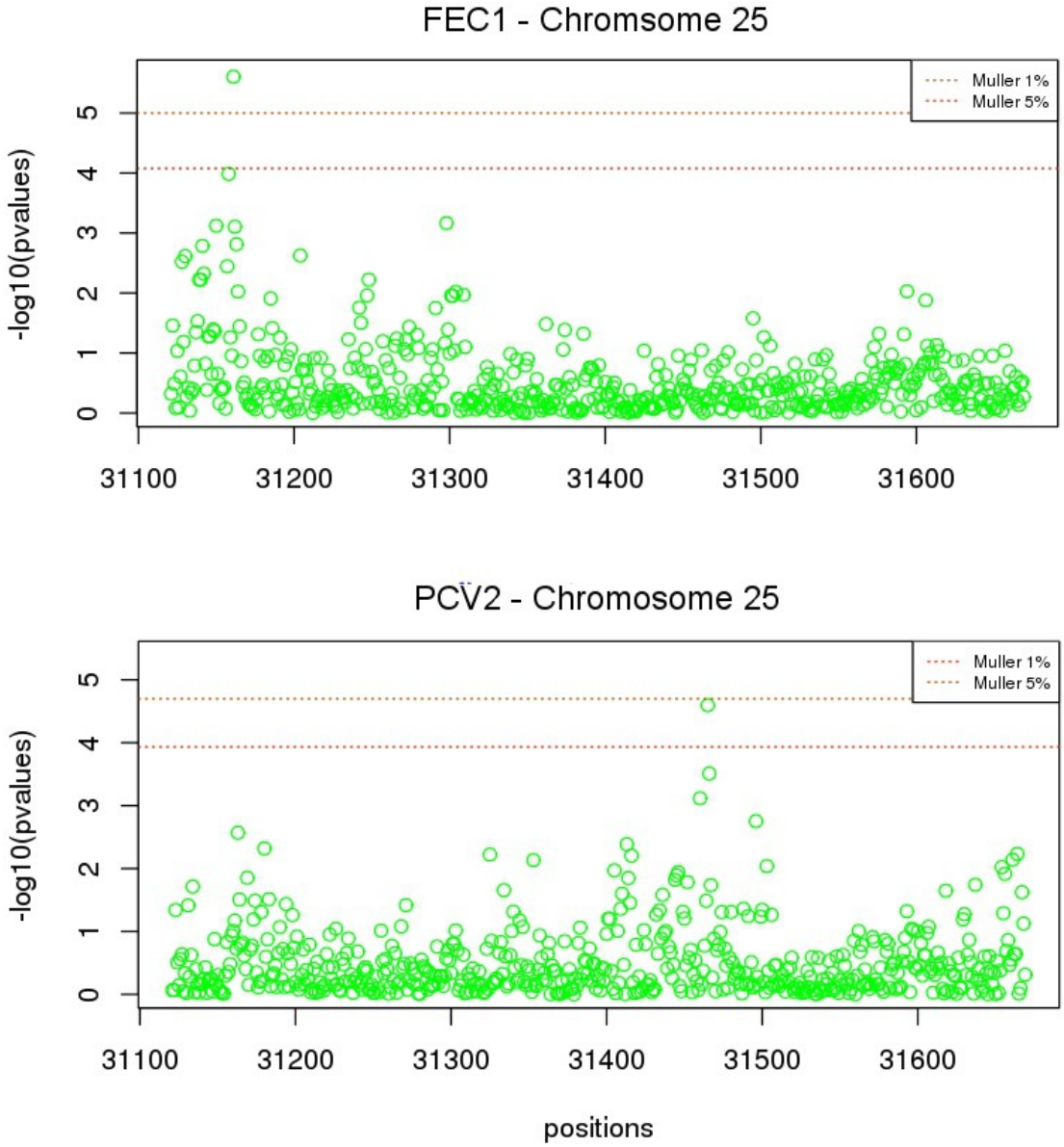


Figure 2 QTLs detected on chromosome 25 for faecal egg count during the first infestation, FEC1 (above), and packed cell volume during the second infestation, PCV2 (below). Position = position relative to the SNP rank on the chromosome; -log10(pvalues) = likelihood ration test for the significance of the estimated effect; Muller 1% = significance threshol at alpha = 0.01; Muller 5% = significance threshold at alpha = 5%.

4 Discussion and perspectives

4.1 Experimental design: weak and strong points

Genetic analysis normally requires a large number of observations in order to obtain accurate and precise estimates. This was indeed a limiting factor for both the study on sheep and the study on goats; whose impact can be noticed on the width of the estimates' standard errors. However, the small number of observations was partly compensated by the experimental settings of both studies, which reduce the environmental variance by allowing to control influential factors such as the larval challenge and the nutritional level. Also, the contribution of the studies performed during this PhD project to the literature of genetic parameters concerning parasite resistance in small ruminants is twofold. First, we provided estimates of the genetic parameters for both parasite resistance and growth traits in sheep featuring phenotypes from experimental settings and molecular information; which might hopefully bare useful information for better understanding the inconsistency of the values previously reported in the literature of such genetic parameters in sheep. Second, the GWAS on parasite resistance traits on creole goat is the only one featuring high density SNPchip, to the best of our knowledge.

Although the study on sheep was performed on more animals than the study on goat, the design of the matings was not ideal for the estimation of the genetic parameters, because the phenotypes were available only on four groups of half-sibs derived from four F1 sires. Furthermore, the population was derived from matings between two extremely different breeds (Romane and Martinik Black Belly) which would require a separate estimation of the genetic parameters for each of the two base populations [234]. Unfortunately, applying a model model for estimating the variance components for each breed separately caused the software to fail in converging on the estimates of the genetic and environmental variances, which is probably due to the fact that only 4 F1 sires were used and the genealogy on the Martinik Black Belly was not very informative. As a matter of fact, having used a back-cross with four sires only, which was done in order to collect observations for a previous study on QTL detection [209], resulted in having a poorly informative pedigree. According to the information present in the pedigree, we have: all animals within each sire's offspring having the same relationship coefficient between them; from the paternal side; only four individuals with a lot of progeny information, compared to all other animals; a single base population that is actually a mixture of two breeds. It may be these features of the genealogy of the

population under study in the first article that actually caused a consistent difference between the estimates' precision obtained from the tested models. The standard error of the estimates obtained by including molecular information was always smaller than the standard error of the relative estimate obtained by including pedigree information only. Indeed, molecular information is particularly helpful in estimating the relationship coefficients among the individuals within a group of full-sibs and/or half-sibs. As a matter of fact, the information conveyed by molecular markers includes the Mendelian sampling term, which is neglected in the case of estimation of relatedness coefficients based on pedigree information only [208]. The pedigree structure of the population studied in the first article aimed, in fact, at maximizing the linkage disequilibrium within families and at simplifying the traceability of parental phases, as required by QTL detection studies [235]. Having used F1 sires ensures high heterozygosity of their genome overall, including the putative loci of the QTLs and at marker loci; which results in having also a large proportion of informative markers in the SNPchip. The genotyping of both the sires and of their offspring allow to track recombination events and finally to increase the power of the experiment [174].

The same could have been done with the population of goats but the divergent lines are still under selection. Simply having a deeper pedigree, hence more than one generation of selection, and/or more observations would also have helped to increase the power to detect putative QTLs. Such a short timespan of selection was actually due to the fact that both the divergent lines populations raised in the experimental unit of Guadeloupe were recently lost in a fire and the selection had to be restarted all over again. Nevertheless, this study resulted in the detection of some QTL, which is encouraging for continuing the selection plan on the divergent lines and eventually setting up a mating design for increasing the power to detect QTLs, such as a backcross.

Furthermore, the precision of the genetic parameres estimates could have been increased, in both studies, by having more observations available and also by having used a bigger number of sires.

4.2 The biology underlying the observed phenotypic variation

The results obtained from the analyses of the phenotypes draw the following picture. For both species under study, the average faecal egg counts measured during the first infestation was higher than the average faecal egg count measured during the second infestation. In both experiments, the animals were kept so as to avoid the contact with any gastrointestinal nematodes and their faecal egg count was controlled for being null before they underwent the

infestation protocol. On the other hand, the second infestation was performed for both experiments at least two months later than the first contact with L3 larvae of *H.contortus* occurred, which is long enough for the primed T cells to be operative for mounting a specific immune response. This can lead us to interpret the results obtained from the analyses of the phenotypes collected during the first infestation as a measurement of the innate immune response, whereas those obtained from the second infestation - as a measurement of the specific immune response. As discussed in the literature review on the immune response to the gastrointestinal parasites, the specific immune response proved to be the most effective of the two. In fact, the immune system features cellular populations fulfilling the function of antigen presenting cells, among which dendritic cells appear to be the most effective against parasite infestations [236]. When an antigen is captured for the first time by these cells, it is processed by the intracellular vacuoles and the resulting epitope is then exposed on the cellular surface, bound to its specific MHCII receptor and depending on its affinity to it. Then the dentritic cell accidentally comes into contact with Th2 cell, which is activated through different signal transduction pathways [231]. The latter determine its differentiation to a Th2 or Th1 cell. When the resulting cell is a Th2 (CD4+) cell, it secretes cytokines, such as IL-4 [43], which, in turn, activate the B plasma cell to initiate an humoral immune response and to express much more numerous antigen receptors on its surface. Furthermore, the B cell proliferates to a differentiated state which makes its next activation more efficient. However, this cascade of events also includes migration to the lymph nodes and prolonged contact between the two successive cell types of each step. Therefore, it can take up to two months to obtain the resulting primed B cells.

Upon phagocytosis of an antigen, the antigen presenting cells and other cellular populations are also activated to a secreting phase, which results in the release of the cytotoxic compounds and of the cytokines determining the typical state of inflammation. The latter is known as innate immune response, which also initiates the cascade of events resulting in the specific immune response two months later, *circa*. That explains how the faecal egg counts during the specific immune response were always lower than those during the innate one.

Furthermore, the article two also shown that the packed cell volume follows an opposite trend than that of faecal egg counts. The packed cell volume is lower when the faecal egg count is higher (during the first infestation) and vice versa - the packed cell volume was higher when the faecal egg count was lower (during the second infestation). This confirms the higher efficacy of the specific immune response compared to the innate one because of two reasons. The indicator of the worm burden (faecal egg count) is lower when the immune

response is more efficient; while the lower level of anemia (packed cell volume) indicates a mildening of the symptoms. As a further confirmation, the phenotypic correlation between faecal egg counts and packed cell volume is always negative.

The first study also reported the phenotypic correlation between faecal egg count and average daily gain. The estimates of the phenotypic correlations between the average daily gain measured before the infestation and both faecal egg counts of the first and second infestation were not significantly different from zero, which means that they are independent. The correlations between the average daily gain during the first infestation and both faecal egg counts during the same period and during the second infestation were negative, instead. The latter correlation estimates can be also interpreted, alike the genetic correlations between the faecal egg counts and the packed cell volume, as being a consequence of the higher efficiency of the specific immune response. Indeed, some of the causes of the average daily gain reduction are: the blood loss, resulting in anaemia, and the immunity related anorexia caused by the release of IL-1 [63]. All of these factors contribute to the reduction of nutrients available for growth. Furthermore, the phenotypic correlations between each trait and its respective trait measured during the next infestation period (faecal egg count during the first infestation with faecal egg count during the successive infestation, and so on) were all positive. The only exception was the correlation between the average daily gain before the infestation and the average daily gain during the first infestation, for the estimation of which the software did not converge. However, this normally happens when the correlation between two traits is very high.

These globally positive estimates suggest the existence of some repeatable variation within these phenotypes. Some evidence of repeatable variation was also observed in the contrast between the faecal egg counts of the two divergent genetic lines of goat, which resulted in a significantly lower average faecal egg count within the resistant group compared to the average of the susceptible group during both infestation periods, consistently. The latter also suggests that the genetic selection which targeted the two lines did result in a significant effect on the phenotypes under selection (faecal egg counts). However, the same contrast for packed cell volumes did nor result in a significant difference between the two groups. This might be explained both by the limited number of observations and by the fact that the breeding goal featured higher weight on faecal egg count then on packed cell volume, but a positively correlated response on packed cell volume could be expected [115].

4.3 Genetic parameters

The heritability of these traits was between moderate to high, for all traits. The heritabilities obtained from article 2 were the only ones showing a slightly higher estimate than what was found in the literature [115]. The reason for this can be twofold. The estimated genetic variance can be higher, which is possible because the phenotypes were collected on divergent lines; and/or the phenotypic variance is smaller, because the phenotypes were averaged over each period of the data-collection protocols and because the experimental settings reduce the environmental. Nevertheless, the standard errors of these estimates were large enough to include zero into their confidence interval, which is not supported by previous literature neither for faecal egg counts, nor for packed cell volume [115].

The number of observations available for the first study was sufficient to compute genetic correlations between average daily gain and faecal egg count in sheep. However, the design of the study was not ideal and, therefore, the standard errors of some of the estimated genetic parameters were quite large. The estimation software failed to converge on an estimate of the correlation between average daily gain before infestation and the one during the first infestation. The global picture of the obtained genetic correlations displays the following features. The correlations between the faecal egg count during the first infestation and both the average daily gains before and during the first infestations were not significantly different from zero, following their respective estimated phenotypic correlation. While the estimates of the genetic correlations between faecal egg count during the first infestation and both the average daily gains during the first and second infestations were negative, some inconsistency was present in the results obtained from the two models applied. This can be interpreted as a suggestion that selecting for growth in a non-contaminated environment might result in long term susceptibility to the parasite, while selecting for growth in a contaminated environment would allow a simultaneous improvement of both growth and resistance. Despite we could not obtain a consistent estimate of the genetic correlation between the average daily gain before the infestation and the average daily gain during the first infestation, previous study already report the occurrence of a significant genotype by environment interaction influencing the expression of growth traits when measured across non contaminated and contaminated environments. Further proof of this possibility can be deduced from the results reported by Coop and Kyriazakis [101, 116], which show that the protein level of the diet has a big impact on the severity of the symptoms.

4.4 Impact of molecular information

The first study featured a comparison between the precision of the estimates obtained from

pedigree information only with those obtained from the same model but including molecular information. The precision of the estimates was measured by two methods: by the asymptotic approximation of their standard errors and by computing a 95% confidence interval by a bootstrap procedure. The latter was added as a follow up on the results of the asymptotic approximation because the distribution law of the correlation estimate follows a Gaussian one only in very special cases, making it difficult to test hypothesis on it when its distribution deviates from normality [218, 237]. Parallel computing allowed to build up a flexible framework which reduced the computational burden of bootstrap estimation to satisfactory time lapses. Both precision estimators indicated that the molecular information increased the precision of genetic parameters. This is possibly due to the particular pedigree structure of the study, as discussed above.

Molecular markers also allowed to compute the allelic substitution effect of their marker loci, in order to detect significant QTLs affecting faecal egg count and packed cell volume in goat. The results obtained from the second article suggest the presence of a QTL on chromosome 6 affecting both faecal egg counts during the first and second infestation. Two more QTLs affecting faecal egg count during the first infestation on chromosomes 10 and 25. While only another one on chromosome 8 would affect faecal egg count during the primed immune response. Chromosome 25 also would bear a QTL affecting packed cell volume during the primed immune response. The latter trait would also be affected by another QTL on chromosome 22. The most significant allelic substitution effect for packed cell volume during the first infestation did only pass the low significance threshold at alpha = 0.1. Some of the results we obtained for creole goat are supported by another study on the same breed [224], where they report a QTL for faecal egg count on chromosome 8 in the same breed. Some QTLs for parasite resistance have also been described in sheep, often close to the regions where genes associated to the Th2 immune response were discovered, such as: the region of MHCII class genes [226], containing the code for the proteic backbone of the receptor expressed on the surface of the antigen presenting cells – which, in turn, triggers the development of an humoral immune response specific to the parasite. However, considering the elevated number of biochemical pathways involved (immune response, neuroendocrin communication, signal transduction, erythropoieis, etc.) and the heavily epistatic transcription of the genes involved in the immune response, as well as the increasing evidence of the high level of epistatsis occurring all along the whole genome (which begins to define a new definition of gene [170]), it should be reasonable not to expect to find any high impact allelic substitution effects [43, 52, 231].

4.5 Practical implications

The practical implications of the QTLs detected on faecal egg count are humbled by the lack of a robust estimate of the economic value of this trait, which makes it difficult to define whether an allelic substitution effect can have a significant economic impact. Further research is required for the application of marker-assisted selection on parasite resistance traits in order to apply them in the production context. Furthermore, goat breeding is normally to be contextualized in low-input production systems, which represent 80% of the whole goat production sector [154]. Nevertheless, the molecular information can enhance genetic selection by increasing the accuracy and precision of genetic parameters and breeding values. Furthermore, it can be helpful during the preselection of selection candidates and in monitoring inbreeding more precisely. Also, it can help gathering information that might prove useful for deciding whether the most appropriate genetic model for estimating breeding values for some trait should include the occurrence of QTLs or whether an infinitesimal model would be sufficiently accurate.

On the other hand selection based on polygenic effects only has already been proven effective for reducing faecal egg counts, for example by the results obtained in Australia and in new Zealand with sheep resistant to *H.contortus* and *Trychostrongylus* spp. [91]. Ideally, breeding plans would be based upon a global breeding goal featuring an economic weight to each trait under selection; however, when the economic weight is not available, it is possible to derive a weight based on the desired yearly genetic gain. What makes the derivation of an economic weight for increased resistance to gastrointestinal nematode infestation particularly puzzling are several factors. First of all, the markets of small ruminant production define very different breeding goals, depending on the country; which results in the need to define different economic weights according to the target market [83]. The need for adapting breeding plans to local context, featuring specific stress factors, is also supported as a solution to possible genotype by environment interaction affecting production traits [116]. Second, it is not trivial to estimate the reduction of production loss due to increased resistance to the parasite; the estimates of the economic impact of gastrointestinal nematodes on small ruminant production are, in fact, based on the cost of anthelmintic treatments rather than on the production loss they cause. Third, since it appears that genetic resistance to these parasites is mainly determined by genes underlying the immune response [39, 233], it is sensible to expect an interaction between genetic selection for parasite resistance and the responsiveness to vaccines [238]. Fourth, nutritional factors have also been proven to play a key role in the

interaction between the host and the parasite [101, 102], which could also interact with the genetic make-up of the host [239].

Another issue which requires further investigation is whether selecting for resistance to gastrointestinal parasites would result in a correlated response towards susceptibility to other diseases. What can be speculated from immunology theory is that selection for parasite resistance can result in animals featuring an enhanced Th2 response, which is indeed more effective than the Th1 response for protecting the individual from gastrointestinal parasites. However, this could also make the Th1 response less efficient, which would result in higher susceptibility to other pathogens, for example intracellular parasites [43, 88].

Furthermore, the genetic make-up resulting from a possible breeding plan would depend on the traits included in the breeding goal and their economic weights. The traits typically used for measuring resistance to gastrointestinal nematodes in sheep span from faecal egg count, to packed cell volume, to IgA in saliva or blood; which are all related to resistance, as their genetic correlations prove [110], but still they are quite different traits from a biological point of view [60, 62, 66, 105]. For example we could speculate that selection on IgA might be more likely to interact with the reaction to vaccination; while selection on resilience might be more likely to do so with nutritional factors. Finally, even when genetic selection would become common practice for controlling gastrointestinal nematodes, it would still be coupled with other control strategies, such as those described briefly in the literature review.

Nevertheless, the estimates of the genetic correlations between growth traits and parasite resistance traits can provide some hint for the design of an eventual breeding plan aiming at improving parasite resistance in sheep. The results obtained suggest the following guidelines. Growth and parasite resistance during the innate immunity should not influence reciprocally their selection responses. On the contrary, selection on growth could negatively affect parasite resistance, when growth was selected in a non-contaminated environment; whereas selecting for growth in a contaminated environment can allow to improve both parasite resistance and growth traits, simultaneously. Genotype by environment interaction on average daily gain might indeed be an issue.

4.6 Modelling

The guidelines illustrated in the paragraph above can also be used for predicting the response to genetic selection in silico, by software simulations. There have already been proposed some epidemio-genetic models [240], which include genetic parameters among the adjustable

parameters for predicting the impact of the genetic make-up of a sheep population both on production traits and on parasite resistance traits, such as egg excretion. These models allow also to extend the prediction to the larval contamination of the pasture, according to both the estimated egg excretion and to the development of the free-life stages of the parasite. Furthermore, it also features many user-defined parameters which allow to simulate the impact of genetic selection on the interaction between the host and the parasite. This approach also allows to account for the influence of genetic selection by simply ccoding the parameters related to the genetic value of the animals as a function that follows the predicted selection response resulting from a particular breeding plan under study.

However, for the outcome of a model to feature some predictory value in practice, the model's behaviour must be analysed thoroughly by a number of good practices, such as: calibration, sensitivity analysis, validation and extrapolation [241]. The calibration process consists of estimating the model's parameters which minimize the difference between the observations collected on the phenomenon the model aim at describing and the model's outcome. This results in obtaining a set of reference parameters which simply ensures that the model can describe the phenomenon of interest, within the conditions under which the real observations were collected. Once a reference set of parameters have been estimated it is possible to proceed with validation. Sensitivity analysis can be considered as a procedure for validating the behaviour of the model in response to the variation of the value of its parameters. Indeed, this procedure allows to test whether the influence each parameter on the model's output actually reflects the theory which the modeller translated into machine language, including non linear interaction between parameters [242]. This results in a map of the model's behaviour against the values of its parameters. The next validation step is to test the resemblance between the model's outcome and a set of observations collected in similar conditions to those used for its calibration, but independent from the latter. This can be done by collecting two different sets of observations under similar conditions, by randomly splitting a single set observations into a calibration set and a validation set. The validation process defines what is the range of values the model's parameters can assume while still resulting in realistic outcomes [243]. The procedures only ensure that the model can describe the phenomenon of interest within certain observed conditions and also define the limits for the models parameters values which result in realistic outcomes. However, if the purpose of a model is to describe the phenomenon of interest within non-observed conditions, the model must be further tested for extrapolation. This problem is treated within the theory of risk analysis; because the only “quantity” that can be measured to assess the reliability of a

model's prediction outside the parameter space used for its validation, is the risk associated with using the model's outcome for decision making in the case that the prediction is wrong [244]. Unfortunately, good practices for model calibration, validation and extrapolation are normally neglected and replaced by common practices [245]. It is also worth noticing that these practices were mostly developed in the context of models used in engineering, physics and other disciplines which are based on much more exact laws than what is achievable in biology.

That said and considering all the above unexplored interaction factors possibly influencing the outcome of genetic selection for parasite resistance, a valid data-set for supporting meaningful calibration, validation and extrapolation procedures for an eventual epidemio-genetic model for predicting the efficiency of genetic selection as a control strategy for gastrointestinal parasites should be very carefully designed and would be very time-consuming to collect. For example, the dataset should feature observations from experiments exploring the interaction between different nutritional levels and larval challenge. In order to make it possible to extrapolate the outcome obtained in silico to the practice of small ruminant production industry, such dataset should also feature observations exploring the interaction between the above factors and anthelmintic treatments, at least. It is very unlikely that anthelmintic treatments will be withdrawn entirely, but they might hopefully be performed with more synergy between farmers, veterinarians and researchers.

Given the hyperdimensional nature of the calibration, validation and extrapolation problems to be solved for making such a model useful in practice, parallel computing would be a must. Indeed, we could already appreciate its power in some statistical procedures, such as bootstrap [219], which rendered fairly feasible testing hypotheses on possibly non-Gaussian variables. It can also support, more effectively than linear code, both an eventual breakthrough in the estimation of epistatic effects and the increasing computational burden required for analyzing more and more dense genetic maps.

5 Conclusions

The study on sheep suggests that it is feasible to select for parasite resistance and growth traits simultaneously, by taking into account the possibility of genotype by environment interaction affecting growth traits. Moreover, including molecular information in the estimation models proved to increase the precision of the genetic parameters estimates.

The study on goat shows that some genetic variation is available to select upon for parasite resistance traits. Furthermore, it suggests that some QTLs might be segregating within the population under study, which could justify the inclusion of marker-assisted selection in the estimation of the selection to response, as a feasible option.

Considering all the issues discussed, it is not reasonable to expect that a unique solution for the control of gastrointestinal parasites by genetic selection. This is a consequence of the fact that breeding plans require a ranking of individuals according to a specific breeding goal. The breeding goal defines what are the features of the best individual relative to the economic weights included in the breeding goal itself [120]. Therefore, the features of the "best-individual" will change according to the production system within which its offspring will be farmed. Breeding for resilience, i.e. breeding for production traits within a parasitized environment, might be a more robust breeding goal than parasite resistance.

Despite the fact that breeding for resistance to gastrointestinal parasites is biologically feasible, its application in practice need to comply with the economics of small ruminant production also. In order to obtain reliable estimates of the efficiency of genetic selection as a control strategy for gastrointestinal parasites in practice it is necessary to develop a model that includes both the biology, genetics and epidemiology, and the economics, economic weights for parasite resistance and/or resilience traits, of the problem. Then, this model needs to be carefully calibrated, properly validated and its predictive power needs to be tested by using a dedicated set of real observations. Once an appropriate dataset has been collected, applying the correct methodology [244, 245] by parallel computing can reduce the time lapse necessary for the development and testing of such a predictive model both to comply with good scientific practice and to satisfy the demand for scientific publications.

6 References

1. Kemper KE, Elwin RL, Bishop SC, Goddard ME, Woolaston RR: **Haemonchus contortus and Trichostrongylus colubriformis did not adapt to long-term exposure to sheep that were genetically resistant or susceptible to nematode infections**. *Int J Parasitol* 2009, **39**:607–614.

2. Waller P, Chandrawathani P: **Haemonchus contortus: Parasite problem No. 1 from Tropics-Polar Circle. Problems and prospects for control based on epidemiology**. *Trop Biomed* 2005, **22**:131–137.

3. May RM, Anderson RM: **Parasite—host coevolution**. *Parasitology* 1990, **100**:S89.

4. Morgan E, Charlier J, Hendrickx G, Biggeri A, Catalan D, von Samson-Himmelstjerna G, Demeler J, Müller E, van Dijk J, Kenyon F, Skuce P, Höglund J, O'Kiely P, van Ranst B, de Waal T, Rinaldi L, Cringoli G, Hertzberg H, Torgerson P, Wolstenholme A, Vercruysse J: **Global Change and Helminth Infections in Grazing Ruminants in Europe: Impacts, Trends and Sustainable Solutions**. *Agriculture* 2013, **3**:484–502.

5. Perry BD: *Investing in Animal Health Research to Alleviate Poverty*. ILRI (aka ILCA and ILRAD); 2002.

6. Urquhart GM: *Veterinary Parasitology*. Oxford, UK; Ames, Iowa: Blackwell Science, ;; 1996.

7. Jacquiet P, Cabaret J, Thiam E, Cheikh D: **Host range and the maintenance of< i> Haemonchus</i> spp. in an adverse arid climate**. *International journal for parasitology* 1998, **28**:253–261.

8. McLeod RS: **Costs of major parasites to the Australian livestock industries**. *Int J Parasitol* 1995, **25**:1363–1367.

9. Vlassoff A, McKenna PB: **Nematode parasites of economic importance in sheep in New Zealand**. *New Zealand Journal of Zoology* 1994, **21**:1–8.

10. Le Jambre LF, Knox MR, Australian Centre for International Agricultural Research (Eds): *Sustainable Parasite Control in Small Ruminants: An International Workshop Sponsored by ACIAR and Held in Bogor, Indonesia 22-25 April 1996*. Canberra, ACT: Australian Centre for International Agricultural Research; 1996.

11. Sani RA, Gray GD, Baker RL: *Worm Control for Small Ruminants in Tropical Asia*. ACIAR; 2004.

12. Waller P., Rudby-Martin L, Ljungström B., Rydzik A: **The epidemiology of abomasal nematodes of sheep in Sweden, with particular reference to over-winter survival strategies**. *Veterinary Parasitology* 2004, **122**:207–220.

13. Hoste H, Chartier C, Le Frileux Y: **Control of gastrointestinal parasitism with nematodes in dairy goats by treating the host category at risk**. *Veterinary Research* 2002,

33:531–545.

14. Jackson F, Coop RL: **The development of anthelmintic resistance in sheep nematodes**. *Parasitology* 2000, **120 Suppl**:S95–107.

15. Morgan ER, Wall R: **Climate change and parasitic disease: farmer mitigation?** *Trends in Parasitology* 2009, **25**:308–313.

16. Van Dijk J, Sargison ND, Kenyon F, Skuce PJ: **Climate change and infectious disease: helminthological challenges to farmed ruminants in temperate regions**. *animal* 2010, **4**:377.

17. Waller P, Dash K, Barger I, Le Jambre L, Plant J: **Anthelmintic resistance in nematode parasites of sheep: learning from the Australian experience**. *Veterinary Record* 1995, **136**:411–413.

18. Papadopoulos E: **Anthelmintic resistance in sheep nematodes**. *Small Ruminant Research* 2008, **76**:99–103.

19. Pomroy W: **Anthelmintic resistance in New Zealand: A perspective on recent findings and options for the future**. *New Zealand Veterinary Journal* 2006, **54**:265–270.

20. Sutherland IA, Leathwick DM: **Anthelmintic resistance in nematode parasites of cattle: a global issue?** *Trends in parasitology* 2010.

21. Bernstein J: *The European Union's Common Agricultural Policy: Pressures for Change*. US Dept. of Agriculture, Economic Research Service; 1999.

22. Kocher DK, Ahuja SP, Sood ML: **Haemonchus contortus: Characterization and Purification of Excretory/Secretory Protease(s)**. *Journal of Veterinary Medicine, Series B* 1998, **45**:567–572.

23. Banks D, Singh R, Barger I, Pratap B, Le Jambre L: **Development and survival of infective larvae of Haemonchus contortus and Trichostrongylus colubriformis on pasture in a tropical environment**. *International Journal for Parasitology* 1990, **20**:155–160.

24. Besier RB, Dunsmore JD: **The ecology of Haemonchus contortus in a winter rainfall region in Australia: the development of eggs to infective larvae**. *Vet Parasitol* 1993, **45**:275–292.

25. Miller JE, Bahirathan M, Lemarie SL, Hembry FG, Kearney MT, Barras SR: **Epidemiology of gastrointestinal nematode parasitism in Suffolk and Gulf Coast Native sheep with special emphasis on relative susceptibility to Haemonchus contortus infection**. *Vet Parasitol* 1998, **74**:55–74.

26. O'Connor LJ, Walkden-Brown SW, Kahn LP: **Ecology of the free-living stages of major trichostrongylid parasites of sheep**. *Veterinary Parasitology* 2006, **142**:1–15.

27. Gruner L, Mauleon H, SAUVE C: **Migrations of trichostrongyle infective larvae experiments with ovine parasites in soil**. In *Annales de recherches veterinaires*. *Volume 13*; 1982:51–59.

28. Blitz NM, Gibbs HC: **Studies on the arrested development of Haemonchus contortus in sheep—II. Termination of arrested development and the spring rise phenomenon**. *International Journal for Parasitology* 1972, **2**:13–22.

29. Blitz NM, Gibbs HC: **Studies on the arrested development of Haemonchus contortus in sheep—I. The induction of arrested development**. *International Journal for Parasitology* 1972, **2**:5–12.

30. Gibbs HC: **Mechanisms of survival of nematode parasites with emphasis on hypobiosis**. *Veterinary Parasitology* 1982, **11**:25–48.

31. Alba-Hurtado F, Muñoz-Guzmán MA: **Immune Responses Associated with Resistance to Haemonchosis in Sheep**. *BioMed Research International* 2013, **2013**:1–11.

32. Maizels RM, Yazdanbakhsh M: **Immune Regulation by helminth parasites: cellular and molecular mechanisms**. *Nature Reviews Immunology* 2003, **3**:733–744.

33. Douch PGC, Morum PE, Rabel B: **Secretion of anti-parasite substances and leukotrienes from ovine gastrointestinal tissues and isolated mucosal mast cells**. *International Journal for Parasitology* 1996, **26**:205–211.

34. Lamont JT: **Mucus: The Front Line of Intestinal Mucosal Defense**. *Annals of the New York Academy of Sciences* 1992, **664**(1 Neuro-immuno-):190–201.

35. Hohenhaus MA, Outteridge PM: **The immunogenetics of resistance to Trichostrongylus colubriformis and Haemonchus contortus parasites in sheep**. *Br Vet J* 1995, **151**:119–140.

36. Rainbird MA, Macmillan D, Meeusen ET: **Eosinophil-mediated killing of Haemonchus contortus larvar: effect of eosinophil activation and role of antobody, complement and interleukin-5**. *Parasite immunology* 1998, **20**:93–103.

37. Meeusen ENT, Balic A: **Do eosinophils have a role in the killing of helminth parasites?** *Parasitology Today* 2000, **16**:95–101.

38. Balic A, Cunningham CP, Meeusen ENT: **Eosinophil interactions with Haemonchus contortus larvae in the ovine gastrointestinal tract**. *Parasite Immunology* 2006, **28**:107–115.

39. Karrow NA, Goliboski K, Stonos N, Schenkel F, Peregrine A: **Review: Genetics of helminth resistance in sheep**. *Canadian Journal of Animal Science* 2014, **94**:1–9.

40. Saluja R, Metz M, Maurer M: **Role and Relevance of Mast Cells in Fungal Infections**. *Frontiers in Immunology* 2012, **3**.

41. Inagaki-Ohara K, Sakamoto Y, Dohi T, Smith AL: **γδ T cells play a protective role during infection with Nippostrongylus brasiliensis by promoting goblet cell function in the small intestine: Importance of γδ T cells in protection against intestinal nematode infection**. *Immunology* 2011, **134**:448–458.

42. Taylor MD, van der Werf N, Maizels RM: **T cells in helminth infection: the regulators and the regulated**. *Trends in Immunology* 2012, **33**:181–189.

43. Tizard IR: *Veterinary Immunology : An Introduction*. St. Louis, Mo.: Saunders; 2009.

44. Amarante AFT, Bricarello PA, Huntley JF, Mazzolin LP, Gomes JC: **Relationship of abomasal histology and parasite-specific immunoglobulin A with the resistance to Haemonchus contortus infection in three breeds of sheep**. *Veterinary Parasitology* 2005, **128**:99–107.

45. Bungiro RD, Sun T, Harrison LM, Shoemaker CB, Cappello M: **Mucosal antibody responses in experimental hookworm infection**. *Parasite Immunology* 2008, **30**:293–303.

46. Harrison GBL, Pulford HD, Doolin EE, Pernthaner A, Shoemaker CB, Hein WR: **Antibodies to surface epitopes of the carbohydrate larval antigen CarLA are associated with passive protection in strongylid nematode challenge infections.** *Parasite Immunology* 2008.

47. Strain SAJ, Bishop SC, Henderson NG, Kerr A, McKellar QA, Mitchell S, Stear MJ: **The genetic control of IgA activity against Teladorsagia circumcincta and its association with parasite resistance in naturally infected sheep**. *Parasitology* 2002, **124**(Pt 5):545–552.

48. Murphy L, Eckersall PD, Bishop SC, Pettit JJ, Huntley JF, Burchmore R, Stear MJ: **Genetic variation among lambs in peripheral IgE activity against the larval stages of Teladorsagia circumcincta**. *Parasitology* 2010, **137**:1249–1260.

49. Pernthaner A, Cole S-A, Morrison L, Green R, Shaw RJ, Hein WR: **Cytokine and antibody subclass responses in the intestinal lymph of sheep during repeated experimental infections with the nematode parasite Trichostrongylus colubriformis**. *Veterinary Immunology and Immunopathology* 2006, **114**:135–148.

50. Carranza F, Falcón CR, Nuñez N, Knubel C, Correa SG, Bianco I, Maccioni M, Fretes R, Triquell MF, Motrán CC, Cervi L: **Helminth Antigens Enable CpG-Activated Dendritic Cells to Inhibit the Symptoms of Collagen-induced Arthritis through Foxp3+ Regulatory T Cells**. *PLoS ONE* 2012, **7**:e40356.

51. Barone R: *Anatomie comparée des mammifères domestiques. appareil digestif, appareil respiratoire Tome troisième, Tome troisième,*. Paris: Vigot; 2009.

52. Schubert ML, Peura DA: **Control of Gastric Acid Secretion in Health and Disease**. *Gastroenterology* 2008, **134**:1842–1860.

53. Lichtenberg: **gastroduodenal mucosal defense**. *Current opinions in gastroenterology* 1999, **15**:463–472.

54. Vuyyuru L, Harrington L, Arimura A, Schubert ML: **Reciprocal inhibitory paracrine pathways link histamine and somatostatin secretion in the fundus of the stomach**. *Am J Physiol* 1997, **273**(1 Pt 1):G106–111.

55. Schubert ML, Jong MJ, Makhlouf GM: **Bombesin/GRP-stimulated somatostatin secretion is mediated by gastrin in the antrum and intrinsic neurons in the fundus**. *Am J Physiol* 1991, **261**(5 Pt 1):G885–889.

56. Schubert ML, Makhlouf GM: **Gastrin secretion induced by distention is mediated by gastric cholinergic and vasoactive intestinal peptide neurons in rats**. *Gastroenterology*

1993, **104**:834–839.

57. Tari A, Kamiyasu T, Yonei Y, Hamada M, Sumii M, Sumii K, Kajiyama G, Dimaline R: **Role of gastrin/CCK-B receptor in the regulation of gastric acid secretion in rat**. *Dig Dis Sci* 1997, **42**:1901–1907.

58. Zaki M, Koduru S, McCuen R, Vuyyuru L, Schubert ML: **Amylin, released from the gastric fundus, stimulates somatostatin and thus inhibits histamine and acid secretion in mice**. *Gastroenterology* 2002, **123**:247–255.

59. Dargie JD, Allonby EW: **Pathophysiology of single and challenge infections of< i> Haemonchus contortus</i> in Merino sheep: Studies on red cell kinetics and the "self-cure" phenomenon**. *International journal for parasitology* 1975, **5**:147–157.

60. Fox MT: **Pathophysiology of infection with gastrointestinal nematodes in domestic ruminants: recent developments**. *Veterinary Parasitology* 1997, **72**:285–308.

61. Kyriazakis I, Tolkamp BJ, Hutchings MR: **Towards a functional explanation for the occurrence of anorexia during parasitic infections**. *Animal Behaviour* 1998, **56**:265–274.

62. Simpson H.: **Pathophysiology of Abomasal Parasitism: Is the Host or Parasite Responsible?** *The Veterinary Journal* 2000, **160**:177–191.

63. Daun JM, McCarthy DO: **The role of cholecystokinin in interleukin-1-induced anorexia**. *Physiology & Behavior* 1993, **54**:237–241.

64. Holden RJ, Pakula IS: **The role of tumor necrosis factor-alpha in the pathogenesis of anorexia and bulimia nervosa, cancer cachexia and obesity**. *Med Hypotheses* 1996, **47**:423–438.

65. Grovum WL: **Factors affecting the voluntary intake of food by sheep. 3. The effect of intravenous infusions of gastrin, cholecystokinin and secretin on motility of the reticulo-rumen and intake**. *Br J Nutr* 1981, **45**:183–201.

66. Charleston WAG: **Comparison of Blood and Bone Marrow Changes in Lambs with Haemonchosis and in Lambs Repeatedly Bled**. *Journal of Comparative Pathology and Therapeutics* 1964, **74**:223–IN33.

67. Dash KM: **Distribution of trichostrongylid nematodes in the abomasum of sheep**. *International journal for parasitology* 1985, **15**:505–510.

68. Rahman WA, Collins GH: **The establishment and development of Haemonchus contortus in goats**. *Vet Parasitol* 1990, **35**:189–193.

69. Dockray GJ, Varro A, Dimaline R, Wang T: **The gastrins: their production and biological activities**. *Annual Review of Physiology* 2001, **63**:119–139.

70. McKellar Q, Duncan JL, Armour J, Lindsay FE, McWilliam P: **Further studies on the response to transplanted adult Ostertagia ostertagi in calves**. *Res Vet Sci* 1987, **42**:29–34.

71. Lehmann FS, Golodner EH, Wang J, Chen MC, Avedian D, Calam J, Walsh JH, Dubinett S, Soll AH: **Mononuclear cells and cytokines stimulate gastrin release from canine antral cells in primary culture**. *Am J Physiol* 1996, **270**(5 Pt 1):G783–788.

72. Weigert N, Schaffer K, Schusdziarra V, Classen M, Schepp W: **Gastrin secretion from primary cultures of rabbit antral G cells: stimulation by inflammatory cytokines**. *Gastroenterology* 1996, **110**:147–154.

73. Bado A, Moizo L, Laigneau JP, Delwaide J, Lewin MJ: **H3-receptor regulation of vascular gastrin and somatostatin releases by the isolated rat stomach**. *Yale J Biol Med* 1994, **67**:113–121.

74. Simcock DC, Joblin KN, Scott I, Burgess DM, Rogers CW, Pomroy WE, Simpson HV: **Hypergastrinaemia, abomasal bacterial population densities and pH in sheep infected with Ostertagia circumcincta**. *Int J Parasitol* 1999, **29**:1053–1063.

75. Murray M: **Structural Changes in Bovine Ostertagiasis Associated with Increased Permeability of the Bowel Wall to Macromolecules**. *Gastroenterology* 1969, **56**:763–772.

76. Jennings J, F. W., Lawon, D. D., Armour, J. J: **Experimental Ostertagia ostertagi infections in calves: studies with abomasal cannulas.** *American Journal of Veterinery Researh* 1966, **27**:1249 – 1257.

77. Serrano MT, Lanas AI, Lorente S, Sáinz R: **Cytokine effects on pepsinogen secretion from human peptic cells**. *Gut* 1997, **40**:42–48.

78. Prichard R: **Genetic variability following selection of Haemonchus contortus with anthelmintics**. *Trends in Parasitology* 2001, **17**:445–453.

79. Dobson RJ, Lejambre L, Gill JH: **Management of anthelmintic resistance: Inheritance of resistance and selection with persistent drugs**. *International Journal for Parasitology* 1996, **26**:993–1000.

80. Van Wyk JA, Bath GF: **The FAMACHA system for managing haemonchosis in sheep and goats by clinically identifying individual animals for treatment**. *Veterinary Research* 2002, **33**:509–529.

81. Newton SE, Meeusen ENT: **Progress and new technologies for developing vaccines against gastrointestinal nematode parasites of sheep**. *Parasite Immunol* 2003, **25**:283–296.

82. Barger IA, Siale K, Banks DJ, Le Jambre LF: **Rotational grazing for control of gastrointestinal nematodes of goats in a wet tropical environment**. *Vet Parasitol* 1994, **53**:109–116.

83. Kosgey IS, Baker RL, Udo HMJ, Van Arendonk JAM: **Successes and failures of small ruminant breeding programmes in the tropics: a review**. *Small Ruminant Research* 2006, **61**:13–28.

84. Terrill TH, Larsen M, Samples O, Husted S, Miller JE, Kaplan RM, Gelaye S: **Capability of the nematode-trapping fungus Duddingtonia flagrans to reduce infective larvae of gastrointestinal nematodes in goat feces in the southeastern United States: dose titration and dose time interval studies**. *Vet Parasitol* 2004, **120**:285–296.

85. Paraud C, Hoste H, Lefrileux Y, Pommaret A, Paolini V, Pors I, Chartier C: **Administration of Duddingtonia flagrans chlamydospores to goats to control gastro-intestinal nematodes: dose trials**. *Vet Res* 2005, **36**:157–166.

86. Bisset SA, Morris CA: **Feasibility and implications of breeding sheep for resilience to nematode challenge**. *International journal for parasitology* 1996, **26**:857–868.

87. Albers GAA, Gray GD: **Breeding for worm resistance: a perspective**. *International journal for parasitology* 1987, **17**:559–566.

88. Stear M., Bishop S., Mallard B., Raadsma H: **The sustainability, feasibility and desirability of breeding livestock for disease resistance**. *Research in Veterinary Science* 2001, **71**:1–7.

89. Fisher, R. A.: *The Genetical Theory of Natural Selection*. Oxford University Press. Oxford; 1930.

90. Bishop S, Morris C: **Genetics of disease resistance in sheep and goats**. *Small ruminant research* 2007, **70**:48–59.

91. Gray GD, Woolaston RR, Eaton BT: *Breeding for Resistance to Infectious Diseases in Small Ruminants*. ACIAR Canberra; 1995.

92. Woolaston RR, Baker RL: **Prospects of breeding small ruminants for resistance to internal parasites**. *International Journal for Parasitology* 1996, **26**:845–855.

93. Dominik S: **Quantitative trait loci for internal nematode resistance in sheep: a review**. *Genetics Selection Evolution* 2005, **37**(Suppl 1):S83.

94. Ridley M: *Evolution*. Oxford; New York: Oxford University Press; 2004.

95. Roy BA, Kirchner JW: **Evolutionary dynamics of pathogen resistance and tolerance**. *Evolution* 2000, **54**:51–63.

96. Pier GB, Grout M, Zaidi T, Meluleni G, Mueschenborn SS, Banting G, Ratcliff R, Evans MJ, Colledge WH: **Salmonella typhi uses CFTR to enter intestinal epithelial cells**. *Nature* 1998, **393**:79–82.

97. J. C. McEwan, K. G. DODDS, G. J. GREER, W. E. BAIN, S. J. DUNCAN: **Genetic estimates for parasite resistance traits in sheep and their correlations with production traits**. *New Zealand Journal of Zoology* 1995, **22**:177.

98. Bisset SA, Vlassoff A, Morris CA, Southey BR, Baker RL, Parker AGH: **Heritability of and genetic correlations among faecal egg counts and productivity traits in Romney sheep**. *New Zealand Journal of Agricultural Research* 1992, **35**:51–58.

99. Greer AW: **Trade-offs and benefits: implications of promoting a strong immunity to gastrointestinal parasites in sheep**. *Parasite Immunol* 2008, **30**:123–132.

100. de Jong, G., van Noordwijk, A. J.: **Acquisition and Allocation of Resources: Genetic (CO) Variances, Selection, and Life Histories**. *The American Naturalist* 1992, **139**:749–770.

101. Coop RL, Holmes PH: **Nutrition and parasite interaction**. *International journal for parasitology* 1996, **26**:951–962.

102. Coop RL, Kyriazakis I: **Influence of host nutrition on the development and consequences of nematode parasitism in ruminants**. *TRENDS in Parasitology* 2001,

17:325–330.

103. Kyriazakis I, Houdijk J: **Immunonutrition: Nutritional control of parasites**. *Small Ruminant Research* 2006, **62**:79–82.

104. Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits*. Sunderland: Sinauer Associates; 1998.

105. Hein WR, Pernthaner A, Piedrafita D, Meeusen EN: **Immune mechanisms of resistance to gastrointestinal nematode infections in sheep.** *Parasite Immunology* 2010.

106. Meeusen ENT, Balic A, Bowles V: **Cells, cytokines and other molecules associated with rejection of gastrointestinal nematode parasites**. *Veterinary Immunology and Immunopathology* 2005, **108**:121–125.

107. McEwan, J. C. JC, Mason, P., Baker, R. L., Clarke, J. N., Hickey, S. M., Turner, K.: **Effect of selection for productive traits on internal parasite resistance in sheep.** *Proceedings of the New Zealand Society of Animal Production* 1992, **52**:53–56.

108. Baker, R. L., Mwamachi, D. M., Audho, J. O., Thorpe, W.: **Genetic resistance to gastro-intestinal nematode parasites in red Maasai, Dorper and Red Maasai X Dorper ewes in the sub-humid tropics.** *Animal Science* 1999, **69**:335–244.

109. Woolaston RR, Elwin RL, Barger IA: **No adaptation of Haemonchus contortus to genetically resistant sheep**. *Int J Parasitol* 1992, **22**:377–380.

110. Douch P, Green R, Morris C, McEwan J, Windon R: **Phenotypic markers for selection of nematode-resistant sheep**. *International Journal for Parasitology* 1996, **26**:899–911.

111. Stear MJ, Bishop SC, Duncan JL, McKellar QA, Murray M: **The repeatability of faecal egg counts, peripheral eosinophil counts, and plasma pepsinogen concentrations during deliberate infections with Ostertagia circumcincta**. *Int J Parasitol* 1995, **25**:375–380.

112. Douch PG, Green RS, Morris CA, Hickey SM: **Genetic factors affecting antibody responses to four species of nematode parasite in Romney ewe lambs**. *Int J Parasitol* 1995, **25**:823–828.

113. McEwan, Dodds, Bray, Scales, knight, Green, Douch: **Parasite resistance of terminal sire breeds mated with Merino dams.** In *Proceedings of the 23rd Annual Meeting of the New Zealand Society for Parasitology*. Wallaceville; 1994:49.

114. Stear MJ, Bairden K, Bishop SC, Buitkamp J, Duncan JL, Gettinby G, McKellar QA, Park M, Parkins JJ, Reid SWJ, Strain S, Murray M: **The genetic basis of resistance to Ostertagia circumcincta inlambs**. *The Veterinary Journal* 1997, **154**:111–119.

115. Vanimisetti H, Andrew S, Zajac A, Notter D: **Inheritance of fecal egg count and packed cell volume and their relationship with production traits in sheep infected with Haemonchus contortus**. *Journal of Animal Science* 2004, **82**:1602.

116. Pollott GE, Greeff JC: **Genotype x environment interactions and genetic parameters for fecal egg count and production traits of Merino sheep**. *J Anim Sci* 2004, **82**:2840–2851.

117. Safari E, Fogarty NM, Gilmour AR: **A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep**. *Livestock Production Science* 2005, **92**:271–289.

118. Prache S, Benoit M, Boutonnet J-P, François D, Sagot L: **La production d'ovins-viande en France. 1ère partie**. *Viandes et Produits Carnés, VPC-2013-29-5-2* 2013, **4**.

119. Fisher, R. A.: **The Correlation between Relatives on the Supposition of Mendelian Inheritance.** *Philosophical Transactions of the Royal Society of Edinburgh* 1918, **52**:399–433.

120. Falconer DS, Mackay TFC: *Introduction to Quantitative Genetics*. Essex, England: Longman; 1996.

121. Chen ZJ: **Molecular mechanisms of polyploidy and hybrid vigor**. *Trends in Plant Science* 2010, **15**:57–71.

122. Eisenhart, C.: **The assumptions underlying the analisis of variance.** *Biometrics* 1947, **3**:1–21.

123. Henderson, C. R.: **Estimation of changes in herd environment.** *Journal of Dairy Science* 1949, **32**:709.

124. Patterson, H.D., Thompson, R.: **Recovery of inter-block information when block sizes are unequal.** *Biometrika* 1971, **58**:545–554.

125. Henderson, C. R.: **Estimation of Variance and Covariance Components.** *Biometrics* 1953, **9**:226–252.

126. Gianola D, Hammond K (Eds): *Advances in Statistical Methods for Genetic Improvement of Livestock*. *Volume 18*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1990. [Yaron B, Bresler E, Thomas GW, Vleck LD (Series editors): *Advanced Series in Agricultural Sciences*]

127. Sorensen, D. A., Kennedy, B. W.: **Estimation of genetic variances from unselected and selected populations.** *Journal of Animal Science* 1984, **59**:1213–1233.

128. Gianola D: **Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures**. *Genetics* 2006, **173**:1761–1776.

129. Meuwissen T, Hayes B, Goddard M: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**:1819.

130. Gianola D, van Kaam JBCHM: **Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits**. *Genetics* 2008, **178**:2289–2303.

131. Wang J: **Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods**. *Journal of Evolutionary Biology* 2014, **27**:518–530.

132. Wright, S.: **Coefficients of inbreeding and relationship.** *American naturalist* 1922, **56**:330–339.

133. Malecot, G.: *Les Mathématiques de L'hérédité.* Masson. Paris; 1948.

134. Jacquard, A.: *The Genetic Structure of Populations.* Springer. New York; 1974.

135. Metzker ML: **Sequencing technologies — the next generation**. *Nature Reviews Genetics* 2010, **11**:31–46.

136. Powell JE, Visscher PM, Goddard ME: **Reconciling the analysis of IBD and IBS in complex trait studies**. *Nature Reviews Genetics* 2010, **11**:800–805.

137. Toro MÁ, García-Cortés LA, Legarra A: **A note on the rationale for estimating genealogical coancestry from molecular markers**. *Genet Sel Evol* 2011, **43**:27.

138. Henderson CR: **Rapid method for computing the inverse of a relationship matrix**. *Journal of Dairy Science* 1975, **58**:1727–1730.

139. Gilmour AR, Gogel BJ, Cullis BR, Thompson R, Butler D: **ASReml user guide release 3.0**. *VSN International Ltd, Hemel Hempstead, UK* 2009.

140. Misztal I: **Reliable computing in estimation of variance components**. *Journal of Animal Breeding and Genetics* 2008, **125**:363–370.

141. Oliehoek PA: **Estimating Relatedness Between Individuals in General Populations With a Focus on Their Use in Conservation Programs**. *Genetics* 2006, **173**:483–496.

142. VanRaden P: **Efficient methods to compute genomic predictions**. *Journal of dairy science* 2008, **91**:4414–4423.

143. Amin, N., van Duijn, C. M., Aluchenko, Y. S.: **A genomic background based method for association analysis in related individuals.** *PLoS ONE* 2007, **2**.

144. Vitezica ZG, Aguilar I, Misztal I, Legarra A: **Bias in genomic predictions for populations under selection**. *Genetics Research* 2011, **93**:357–366.

145. Bömcke E, Soyeurt H, Szydlowski M, Gengler N: **New method to combine molecular and pedigree relationships**. *Journal of Animal Science* 2011, **89**:972.

146. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS: **Invited Review: Reliability of genomic predictions for North American Holstein bulls**. *Journal of Dairy Science* 2009, **92**:16–24.

147. Legarra A, Aguilar I, Misztal I: **A relationship matrix including full pedigree and genomic information**. *Journal of dairy science* 2009, **92**:4656–4663.

148. Henderson CR: **Best Linear Unbiased Estimation and Prediction under a Selection Model**. *Biometrics* 1975, **31**:423.

149. Mrode RA: *Linear Models for the Prediction of Animal Breeding Values*. 2nd ed. Wallingford, UK ; Cambridge, MA: CABI Pub; 2005.

150. Sorensen D, Gianola D: *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. New York: Springer-Verlag; 2002.

151. Henderson, C. R.: **Estimation of genetic parameters.** *Annals of mathematical statistics* 1950, **21**:309.

152. *Animal Breeding: Selected Articles from the World Animal Review*. Rome: Food and Agriculture Organization of the United Nations; 1977. [*FAO Animal Production and Health Paper ; 1*]

153. Womack JE: **Advances in livestock genomics: Opening the barn door**. *Genome Research* 2005, **15**:1699–1705.

154. Guimarães, E.: *Marker-Assisted Selection: Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish*. Rome: Food and Agriculture Organization of the United Nations; 2007.

155. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ: **Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score**. *Journal of Dairy Science* 2010, **93**:743–752.

156. De los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL: **Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding**. *Genetics* 2013, **193**:327–345.

157. Hoerl AE, Kennard RW: **Ridge Regression: Biased Estimation for Nonorthogonal Problems**. *Technometrics* 1970, **12**:55–67.

158. Tibshirani, R.: **Regression Shrinkage and Selection via the Lasso.** *Journal of the Royal Statistical Society* 1996, **58**:267–288.

159. Zou H, Hastie T: **Regularization and variable selection via the elastic net**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005, **67**:301–320.

160. Endelman JB: **Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP**. *The Plant Genome Journal* 2011, **4**:250.

161. Park T, Casella G: **The Bayesian Lasso**. *Journal of the American Statistical Association* 2008, **103**:681–686.

162. Calus MPL, Meuwissen THE, de Roos APW, Veerkamp RF: **Accuracy of Genomic Selection Using Different Methods to Define Haplotypes**. *Genetics* 2008, **178**:553–561.

163. Yu X, Meuwissen TH: **Using the Pareto principle in genome-wide breeding value estimation**. *Genetics Selection Evolution* 2011, **43**:35.

164. Zou F, Huang H, Lee S, Hoeschele I: **Nonparametric Bayesian Variable Selection With Applications to Multiple Quantitative Trait Loci Mapping With Epistasis and Gene-Environment Interaction**. *Genetics* 2010, **186**:385–394.

165. Habier D, Fernando RL, Kizilkaya K, Garrick DJ: **Extension of the Bayesian alphabet for genomic selection**. *BMC bioinformatics* 2011, **12**:186.

166. Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R: **Additive Genetic Variability and the Bayesian Alphabet**. *Genetics* 2009, **183**:347–363.

167. De los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM: **Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree**. *Genetics* 2009, **182**:375–385.

168. Neuner S, Emmerling R, Thaller G, Götz K-U: **Strategies for Estimating Genetic Parameters in Marker-Assisted Best Linear Unbiased Predictor Models in Dairy Cattle**. *Journal of Dairy Science* 2008, **91**:4344–4354.

169. Gianola D, Okut H, Weigel KA, Rosa GJ: **Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat**. *BMC Genetics* 2011, **12**:87.

170. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M: **What is a gene, post-ENCODE? History and updated definition**. *Genome Research* 2007, **17**:669–681.

171. Dekkers JC: **Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons**. *Journal of Animal Science* 2004, **82**(13 suppl):E313–E328.

172. Hayes B, Goddard ME: **The distribution of the effects of genes affecting quantitative traits in livestock**. *Genetics Selection Evolution* 2001, **33**:209–230.

173. Griffiths AJF (Ed): *An Introduction to Genetic Analysis*. 7th ed. New York: W.H. Freeman; 2000.

174. Van der Werf JHJ, Marshall K, Lee S: **Methods and experimental designs for detection of QTL in sheep and goats**. *Small Ruminant Research* 2007, **70**:21–31.

175. Hill WG, Robertson A: **Linkage disequilibrium in finite populations**. *Theoretical and Applied Genetics* 1968, **38**:226–231.

176. Zhao H, Speed TP: **On genetic map functions**. *Genetics* 1996, **142**:1369–1377.

177. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, Sabatti C, Eskin E: **Variance component model to account for sample structure in genome-wide association studies**. *Nature Genetics* 2010, **42**:348–354.

178. Ott L: *An Introduction to Statistical Methods and Data Analysis*. 6th ed. Australia ; United States: Brooks/Cole Cengage Learning; 2010.

179. Müller BU, Stich B, Piepho HP: **A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants**. *Heredity* 2011, **106**:825–831.

180. Lander, E. S., Botstein, D.: **Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps**. *Genetics* 1989, **121**:185–199.

181. Gimelfarb A: **Offspring-parent genotypic regression: how linear is it?** *Biometrics* 1986, **42**:67–71.

182. Visscher PM, Hill WG, Wray NR: **Heritability in the genomics era — concepts and misconceptions**. *Nature Reviews Genetics* 2008, **9**:255–266.

183. Calus MPL, Veerkamp RF: **Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM**. *J Anim Breed Genet* 2007, **124**:362–368.

184. Grobet L, Poncelet D, Royo LJ, Brouwers B, Pirottin D, Michaux C, Ménissier F, Zanotti M, Dunner S, Georges M: **Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle**. *Mammalian genome* 1998, **9**:210–213.

185. Schmutz SM, Marquess FLS, Berryere TG, Moker JS: **DNA marker-assisted selection of the polled condition in Charolais cattle**. *Mammalian Genome* 1995, **6**:710–713.

186. Ciobanu DC, Bastiaansen JWM, Lonergan SM, Thomsen H, Dekkers JC, Plastow GS, Rothschild MF: **New alleles in calpastatin gene are associated with meat quality traits in pigs**. *Journal of animal science* 2004, **82**:2829–2839.

187. Villanueva B, Pong-Wong R, Woolliams JA: **Marker assisted selection with optimised contributions of the candidates to selection**. *Genetics Selection Evolution* 2002, **34**:679.

188. Lande R, Thompson R: **Efficiency of marker-assisted selection in the improvement of quantitative traits.** *Genetics* 1990, **124**:743–756.

189. Visscher PM, Haley CS, Thompson R: **Marker-assisted introgression in backcross breeding programs**. *Genetics* 1996, **144**:1923–1932.

190. Nieuwhof G, Bishop S: **Costs of the major endemic diseases of sheep in Great Britain and the potential benefits of reduction in disease impact**. *ANIMAL SCIENCE-GLASGOW THEN PENICUIK-* 2005, **81**:23.

191. Kaplan RM: **Drug resistance in nematodes of veterinary importance: a status report**. *Trends Parasitol* 2004, **20**:477–481.

192. Wolstenholme AJ, Fairweather I, Prichard R, von Samson-Himmelstjerna G, Sangster NC: **Drug resistance in veterinary helminths**. *Trends in Parasitology* 2004, **20**:469–476.

193. Bisset S, Morris C, McEwan J, Vlassof A: **Breeding sheep in New Zealand that are less reliant on anthelmintics to maintain health and productivity**. *New Zealand Veterinary Journal* 2001, **49**:236–246.

194. Eysker M, Bakker N, Kooyman FNJ, van der Linden D, Schrama C, Ploeger HW: **Consequences of the unusually warm and dry summer of 2003 in The Netherlands: poor development of free living stages, normal survival of infective larvae and long survival of adult gastrointestinal nematodes of sheep**. *Vet Parasitol* 2005, **133**:313–321.

195. Getachew T, Dorchies P, Jacquiet P: **Trends and challenges in the effective and sustainable control of Haemonchus contortus infection in sheep. Review**. *Parasite* 2007, **14**:3–14.

196. González Garduño R, Cordero Ortega JC, Torres Hernández G, Arece García J, Mendoza de Gives P: **Efecto del hipoclorito de sodio y extracto de cítricos en la reducción de la infestación con nematodos gastrointestinales resistentes a antihelmínticos en ovinos de pelo**. *Revista mexicana de ciencias pecuarias* 2010, **1**:179–187.

197. Hoste H, Torres-Acosta JFJ: **Non chemical control of helminths in ruminants: Adapting solutions for changing worms in a changing world**. *Veterinary Parasitology* 2011, **180**:144–154.

198. Kemper K, Elwin R, Bishop S, Goddard M, Woolaston R: **Haemonchus contortus and Trichostrongylus colubriformis did not adapt to long-term exposure to sheep that were genetically resistant or susceptible to nematode infections**. *International journal for parasitology* 2009, **39**:607–614.

199. Kloosterman A, Parmentier HK, Ploeger HW: **Breeding cattle and sheep for resistance to gastrointestinal nematodes**. *Parasitology Today* 1992, **8**:330–335.

200. Larsen M: **Biological control of nematode parasites in sheep**. *Journal of Animal Science* 2006, **84**(13 suppl):E133.

201. SAUL G: **Effects of two pasture systems on faecal nematode egg counts in breeding ewes**. *Australian Veterinary Journal* 1996, **74**:154–155.

202. Smith W, Zarlenga D: **Developments and hurdles in generating vaccines for controlling helminth parasites of grazing ruminants**. *Veterinary parasitology* 2006, **139**:347–359.

203. Torres-Acosta J, Hoste H: **Alternative or improved methods to limit gastro-intestinal parasitism in grazing sheep and goats**. *Small Ruminant Research* 2008, **77**:159–173.

204. Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits*. Sunderland: Sinauer Associates; 1998.

205. Eady S, Woolaston R, Mortimer S, Lewer R, Raadsma H, Swan A, Ponzoni R: **Resistance to nematode parasites in Merino sheep: sources of genetic variation**. *Australian Journal of Agricultural Research* 1996, **47**:895.

206. S. C. Bishop, K. Bairden, Q. A. McKellar, M. Park, M. J. Stear: **Genetic parameters for faecal egg count following mixed, natural, predominantlly Ostertagia circumcinta infection and relationships with live weight in young lambs**. *British Society of Animal Science* 1996, **63**:423–428.

207. Walkden-Brown SW, Eady SJ: **Nutritional influences on the expression of genotypic resistance to gastrointestinal nematode infection in sheep**. *Australian Journal of Experimental Agriculture* 2003, **43**:1445.

208. Lee SH, Goddard ME, Visscher PM, Van Der Werf JHJ: **Research Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits**. 2010.

209. Salle G, Jacquiet P, Gruner L, Cortet J, Sauve C, Prevot F, Grisez C, Bergeaud JP, Schibler L, Tircazes A, Francois D, Pery C, Bouvier F, Thouly JC, Brunel JC, Legarra A, Elsen JM, Bouix J, Rupp R, Moreno CR: **A genome scan for QTL affecting resistance to Haemonchus contortus in sheep**. *Journal of Animal Science* 2012.

210. Lacroux C, Nguyen THC, Andreoletti O, Prevot F, Grisez C, Bergeaud J-P, Gruner L, Brunel J-C, Francois D, Dorchies P, Jacquiet P: ***Haemonchus contortus*** **(Nematoda:**

Trichostrongylidae) infection in lambs elicits an unequivocal Th2 immune response. *Veterinary Research* 2006, **37**:607–622.

211. Raynaud JP: **[Study of the efficiency of a quantitative coproscopic technic for the routine diagnosis and control of parasitic infestations of cattle, sheep, horses and swine]**. *Ann Parasitol Hum Comp* 1970, **45**:321–342.

212. **Complex models, more data: simpler programming?** [https://journal.interbull.org/index.php/ib/article/download/470/468]

213. Christensen OF: **Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation**. *Genet Sel Evol* 2012, **44**:37.

214. Vitezica ZG, Aguilar I, Misztal I, Legarra A: **Bias in genomic predictions for populations under selection**. *Genet Res (Camb)* 2011, **93**:357–366.

215. Quaas RL: **Computing the Diagonal Elements and Inverse of a Large Numerator Relationship Matrix**. *Biometrics* 1976, **32**:949–953.

216. Klei B, Tsuruta S: *Approximate Variance for Heritability Estimates*. 2008.

217. Díaz-Francés E, Rubio FJ: **On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables**. *Statistical Papers* 2013, **54**:309–323.

218. Fisher RA: **Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population**. *Biometrika* 1915, **10**:507–521.

219. Efron B: *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993. [*Monographs on Statistics and Applied Probability*, vol. 57]

220. Liu SM, Smith TL, Briegel J, Murray A, Masters DG, Karlsson LJE, Palmer DG, Greeff JC, Besier RB, Gao SB: **Comparing productive performance of nematode resistant Merino sheep with non-selected control**. *Livestock Production Science* 2005, **97**:117–129.

221. Veerkamp R, Mulder H, Thompson R, Calus M: **Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped**. *Journal of dairy science* 2011, **94**:4189–4197.

222. Over HJ: *Distribution and Impact of Helminth Diseases of Livestock in Developing Countries*. Rome: Food and Agriculture Organization of the United Nations; 1992. [*FAO Animal Production and Health Paper*, vol. 96]

223. Kao RR, Leathwick DM, Roberts MG, Sutherland IA: **Nematode parasites of sheep: a survey of epidemiological parameters and their application in a simple model**. *Parasitology* 2000, **121 (Pt 1)**:85–103.

224. De la Chevrotière C, C. Bishop S, Arquet R, Bambou JC, Schibler L, Amigues Y, Moreno C, Mandonnet N: **Detection of quantitative trait loci for resistance to gastrointestinal nematode infections in Creole goats**. *Animal Genetics* 2012, **43**:768–775.

225. Riggio, V., Matika, O., Pong-Wong, R., Stear, M.J., Bishop, S.C.: **Genome-wide association and regional heritability mapping to identify loci underlying variation in**

nematode resistance and body weight in Scottish Blackface lambs. *Heredity* 2013, **110**:420–429.

226. Bolormaa S, Van Der Werf JHJ, Walkden-Brown SW, Marshall K, Ruvinsky A: **A quantitative trait locus for faecal worm egg and blood eosinophil counts on chromosome 23 in Australian goats: A QTL for faecal worm egg & blood eosinophil counts**. *Journal of Animal Breeding and Genetics* 2009, **127**:207–214.

227. Mandonnet N, Aumont G, Fleury J, Arquet R, Varo H, Gruner L, Bouix J, Khang JV: **Assessment of genetic variability of resistance to gastrointestinal nematode parasites in Creole goats in the humid tropics.** *Journal of Animal Science* 2001, **79**:1706–1712.

228. **Muller 0.0.3** [https://forge-dga.jouy.inra.fr/documents/354]

229. Li H: **A quick method to calculate QTL confidence interval**. *Journal of genetics* 2011, **90**:355.

230. Dukes HH: *Dukes' Physiology of Domestic Animals*. 12th ed. Ithaca: Comstock Pub. Associates; 2004.

231. Berg JM, Tymoczko JL, Stryer L, Clarke ND: *Biochemistry*. New York: Freeman and Company; 2002.

232. Stear MJ, Bishop SC, Doligalska M, Duncan JL, Holmes PH, Irvine J, McCRIRIE L, McKELLAR QA, Sinski E, Murray M: **Regulation of egg production, worm burden, worm length and worm fecundity by host responses in sheep infected with Ostertagia circumcincta**. *Parasite Immunology* 1995, **17**:643–652.

233. Stear MJ, Murray M: **Genetic resistance to parasitic disease: particularly of resistance in ruminants to gastrointestinal nematodes**. *Veterinary Parasitology* 1994, **54**:161–176.

234. García-Cortés LA, Toro MÁ: **Multibreed analysis by splitting the breeding values**. *Genetics Selection Evolution* 2006, **38**:601–615.

235. Van der Werf JHJ, Marshall K, Lee S: **Methods and experimental designs for detection of QTL in sheep and goats**. *Small Ruminant Research* 2007, **70**:21–31.

236. Maizels RM, Balic A, Gomez-Escobar N, Nair M, Taylor MD, Allen JE: **Helminth parasites - masters of regulation**. *Immunological Reviews* 2004, **201**:89–116.

237. Liu BH, Knapp SJ, Birkes D: **Sampling distributions, biases, variances, and confidence intervals for genetic correlations**. *TAG Theoretical and Applied Genetics* 1997, **94**:8–19.

238. Colditz IG, Watson DL, Gray GD, Eady SJ: **Some relationships between age, immune responsiveness and resistance to parasites in ruminants**. *Int J Parasitol* 1996, **26**:869–877.

239. Gray GD: **The use of genetically resistant sheep to control nematode parasitism**. *Veterinary parasitology* 1997, **72**:345–366.

240. Doeschl-Wilson AB, Vagenas D, Kyriazakis I, Bishop SC: **Exploring the assumptions underlying genetic variation in host nematode resistance**. *Genetics Selection Evolution*

2008, **40**:241–264.

241. Trucano TG, Swiler LP, Igusa T, Oberkampf WL, Pilch M: **Calibration, validation, and sensitivity analysis: What's what**. *Reliability Engineering & System Safety* 2006, **91**:1331–1357.

242. Campolongo F, Saltelli A, Cariboni J: **From screening to quantitative sensitivity analysis. A unified approach**. *Computer Physics Communications* 2011, **182**:978–988.

243. Arlot S, Celisse A: **A survey of cross-validation procedures for model selection**. *Statistics Surveys* 2010, **4**:40–79.

244. Maranzano CJ, Krzysztofowicz R: **Bayesian Reanalysis of the Challenger O-Ring Data**. *Risk Analysis* 2008, **28**:1053–1067.

245. Saltelli A, Annoni P: **How to avoid a perfunctory sensitivity analysis**. *Environmental Modelling & Software* 2010, **25**:1508–1517.

AUTHOR: Assenza Fabrizio

TITLE: Modelling genetic selection for the resistance to gastrointestinal parasites in small ruminants

SUPERVISOR: Moreno Carole

PLACE AND DATE: Toulouse 30/09/2014

Abomasal nematodes are a major constraint to small ruminants industry worldwide. Recently their economic impact has increased due to the recrudescence of anthelmintic resistance among many parasite populations. Genetic selection might be a valid strategy for enhancing the efficacy of anthelmintics. We explored the genetic variability, in both sheep and goat, possibly available for a breeding plan featuring parasite resistance as its breeding goal. The results obtained in terms of heritabilities, genetic correlations and QTLs, suggest that the variation in the genetic pool of the population under study might comply with the requirements of a breeding goal including both parasite resistance and production traits. Furthermore, marker assisted selection could be a feasible option to enhance the selection response.

La modélisation de la résistance aux parasites gastro-intestinaux chez les petites ruminants

Les nématodes gastro-intestinaux sont des parasites de la caillette des petits ruminants qui posent des contraintes majeures pour l'élevage de ces animaux dans le monde. Récemment leur impact économique a augmenté notamment à cause de l'apparition de nématodes résistants aux anthelminthiques. La sélection génétique pourrait être une stratégie complémentaire des traitements chimiques. Dans cette thèse, nous avons exploré la variabilité génétique disponible qui permettrait une sélection sur la résistance aux nématodes. Les résultats obtenus en termes de héritabilites, corrélations génétiques et QTLs, suggèrent que la variation génétique des populations étudiées pourrait satisfaire les requis d'un objectif de sélection permettant à la fois d'améliorer la résistance aux nématodes et la croissance des animaux. En outre, l'identification de loci SNP associés à la variation observée sur les caractères de résistance aux nématodes pourrait nous permettre d'améliorer la réponse à la sélection.

KEY-WORDS : Gastrointestinal nematodes, Small ruminants, Genetic parameters, QTLs, robustness, growth.

DISCIPLINE ADMINISTRATIVE : Sciences Ecologiques, Vétérinaires, génétiques et bioingégnieries

---

ISTITUTIONAL NAME AND ADRESS : UMR3188 GenPhySe, INRA -Auzeville, 24 chemin de borde-rouge, 31326, Castanet-Tolosan, France

---