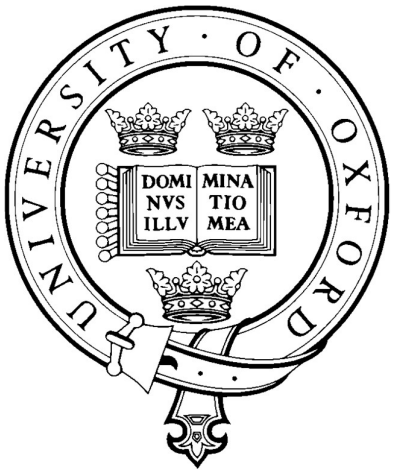
REPRODUCTIVE MICROBIOTAS OF THE RED JUNGLEFOWL, *GALLUS GALLUS*: PATTERNS AND DRIVERS OF VARIATION



ROSIE ECCLESTON

ST PETER'S COLLEGE

UNIVERSITY OF OXFORD

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

ACKNOWLEDGEMENTS

First, I would like to say thank you to my principal supervisor Tom Pizzari for his guidance and encouragement. His knowledge, experience and support have been an invaluable resource throughout. I would also like to thank all those who have so patiently given me their advice, feedback and time; Ben Ashby, Pati Faria-Shayler, Sunetra Gupta, Lucy Larkman, Wook Kim, Harrison Ostridge, Stephen Preston, Kiyono Sekii, Lewis Spurgin, Cedric Tan and David Wilson.

I am also very grateful to all members of the Pizzari Lab; you're all kind, clever and, as would be expected from a group of people who study chicken sex, certainly interesting colleagues. You do you. In particular Grant McDonald has been a huge support as both an infinite source of statistical wisdom, and a friend.

To everyone who has supported me in a variety of ways over the last four years, your wisdom, pep talks, laughter and cups of tea have contributed so much to this thesis: all the Warwick St Laydeez, the Zoology lot, Canalside Mary, The Maggie Crew (Get it done, Handley!), SPCBC, my college family (ILovemyMCRPresident) to name but a few. Two inspirational ladies in particular need to be acknowledged. Ellie Bath and Olivia Ambler. You are the best at everything you do, loyal, kind, hilarious, and, most importantly, as mad as I am.

My family taught me two things, without which I think producing a thesis is impossible. How to work hard, and a good sense of humour. Finally, I wouldn't have started this without my mum, and I wouldn't have finished this without my dad. Thanks for everything. This is for you, Mum.

CONTENTS

AUTHOR CONTRIBUTIONS **4**
THESIS ABSTRACT **6**
CHAPTER 1. INTRODUCTION **7**
CHAPTER 2. THE SEMINAL MICROBIOTA OF THE MALE RED JUNGLEFOWL: METHODOLOGICAL APPROACHES AND PATTERNS OF VARIATION **33**
Abstract ***33***
Introduction ***33***
Material And Methods ***40***
Results ***61***
Discussion ***88***
CHAPTER 3. CHARACTERISING SEMINAL MICROBIOTAS OF RED JUNGLEFOWL LIVING UNDER DIFFERENT ECOLOGICAL CONDITIONS BY COMBINING *IN VITRO* CULTURING WITH GENOMIC SEQUENCING **97**
Abstract ***97***
Introduction ***98***
Materials & Methods ***102***
Results ***111***
Discussion ***128***
Appendix§1 ..... ***136***
Appendix§2 ..... ***138***
Appendix§3 ..... ***144***
CHAPTER 4. DRIVERS OF VARIATION IN SEMINAL MICROBIOTAS WITHIN AND ACROSS MALE RED JUNGLEFOWL **145**
Abstract ***145***
Introduction ***146***
Methods ***154***
Results ***169***
Discussion ***180***
Appendix§1 ..... ***187***
CHAPTER 5. INSEMINATION AS A DRIVER OF VARIATION IN CLOACAL MICROBIOTAS WITHIN AND ACROSS FEMALE RED JUNGLEFOWL **189**
Abstract ***189***
Introduction ***190***
Materials & Methods ***196***
Results ***214***
Discussion ***226***
CHAPTER 6. DISCUSSION **235**
REFERENCES **246**

AUTHOR CONTRIBUTIONS

Chapter 2

Rosie Eccleston, Tom Pizzari and Kiyono Sekii conceived the idea. Sample collection and initial processing was led by Kiyono Sekii (2013) and Rosie Eccleston (2015). Extraction of DNA was performed by Rosie Eccleston (2013 and 2015 samples) and Kiyono Sekki (2013 samples), and the remainder of the laboratory processing was carried out by Rosie Eccleston for all samples, except for the Illumina MiSeq sequencing, which was carried out at the Departmental Sequencing Facility. Initial stages of sequence output processing were performed by Stephen Preston, with all subsequentt process and analysis done by Rosie Eccleston. Rosie Eccleston wrote the chapter.

Chapter 3

Rosie Eccleston, Tom Pizzari and Wook Kim conceived the idea. Sample collection was led by Rosie Eccleston. Culturing of samples and subsequent laboratory processing was carried out by Rosie Eccleston and Harrison Ostridge, except for the Illumina MiSeq sequencing, which was carried out at the Departmental Sequencing Facility. Initial stages of sequence output processing were performed by Stephen Preston, with all subsequent process and analysis done by Rosie Eccleston. Rosie Eccleston wrote the chapter.

Chapter 4

'Rosie Eccleston, Kiyono Sekii and Tom Pizzari conceived the study. Behavioural observations to provide male mating data and blood samples for MHC analysis were collected by Grant McDonald and Tom Pizzari, respectively. MHC data were generated by David Richardson, Lewis Spurgin and Eleanor Fairfield. Microbiota sample collection and initial processing was led by Kiyono Sekii (2013) and Rosie Eccleston (2015). DNA extraction was performed by Rosie Eccleston (2013 and 2015 samples) and Kiyono Sekii (2013 samples), and the remainder of the laboratory processing was carried out by Rosie Eccleston for all samples, except for the Illumina MiSeq sequencing, which was carried out at the Departmental Sequencing Facility. Initial stages

of sequence output processing were performed by Stephen Preston, with all subsequent process and analysis done by Rosie Eccleston. Rosie Eccleston wrote the chapter.'

Chapter 5

Rosie Eccleston, Tom Pizzari and Kiyono Sekii conceived the idea. Artificial Insemination, Sample collection and initial processing was led by Kiyono Sekii and Cedric Tan. Extraction of DNA was performed by Rosie Eccleston (male samples) and Kiyono Sekki (female samples), and the remainder of the laboratory processing was carried out by Rosie Eccleston for all samples, except for the Illumina MiSeq sequencing, which was carried out at the Departmental Sequencing Facility. Initial stages of sequence output processing were performed by Stephen Preston, with all subsequent process and analysis done by Rosie Eccleston. Rosie Eccleston wrote the chapter.

THESIS ABSTRACT

Organisms host diverse and complex bacterial communities (microbiotas), often with fundamental roles in host ecology and evolution. While microbiotas of the gut have received considerable research focus, relatively little is known about reproductive microbiotas, (i.e. those communities living within the reproductive tract and potentially transmitted at mating), despite the potentially critical role they may play in host sexual selection and mating system evolution. This thesis seeks to address this gap by laying the foundations for the systematic study of reproductive microbiotas in bird populations, and the patterns and drivers of intra- and inter-individual variation. I studied a captive population of red junglefowl, *Gallus gallus*, a sexually promiscuous species with well-characterised mating systems. Developing a pipeline based on a next-generation sequencing approach I show that junglefowl ejaculates contain diverse bacterial communities of bacteria (seminal microbiotas), which reflect the gut microbiota and vary in structure across males. I then use *in vitro* culturing to demonstrate that seminal microbiotas are alive and ecologically complex. Next, focusing on patterns of variation in seminal microbiotas, I show an association between inter-male variation in composition and aspects of the male immune-genotype, and find evidence that variation in seminal microbiota structure might be influenced by the sexual dynamics of social groups. Finally, I focus on the female cloacal microbiota and the potential effects that insemination of male semen might have on these communities. Using a series of controlled artificial insemination experiments, I show some evidence that the microbiota of the female cloaca appears to be disrupted by the first inseminations, but might revert back to its initial structure following cessation of inseminations. These effects might be due to the disruption caused by the insemination, rather than by the transmission, and subsequent colonisation of male bacteria. Collectively, these results demonstrate an ecologically complex and dynamic avian reproductive microbiota, providing a platform and direction for future work, by emphasising the need for an understanding of composition, function and variation in reproductive microbiota in order to better understand microbiotas and contextualise the evolutionary ecology of host mating systems.

CHAPTER 1. INTRODUCTION

Since pioneering work such as that by Pasteur and Koch to identify the causative agents of disease in the 19th century (Atlas & Bartha, 1998), it has been widely accepted that microorganisms within a multicellular host can have a negative impact on host health and function. However, more recent advances in the field of microbiology have revealed that there are many bacterial species present within a host's body that do not cause disease, and can even contribute to healthy host function (Cho & Blaser, 2012). The communities of bacterial species living within or on a host are known as microbiotas, and the term can refer to the entire collection of bacteria in a given host species, an individual or a particular anatomical location in the host's body. Various microbiotas have been explored in relation to their role in a range of neurological (Moos *et al.*, 2016), behavioural (Archie & Theis, 2011) and physiological host functions (Sommer & Bäckhed, 2013). As the potential importance of microbiota is revealed by such studies, more research effort has been put into fully characterising and understanding the effects of animal microbiota compositions in order to develop therapeutics (Mimee *et al.*, 2016), explore previously unexplained phenotypic variation (Turnbaugh *et al.*, 2007), and to refine our understanding of evolutionary and ecological processes (Rosenberg & Zilber-Rosenberg, 2016).

Microbiota Function

A widely studied topic is the relationship between the microbiota and host metabolic function. Early studies focused on leguminous plants, which contain nitrogen-fixing rhizobium bacteria in their root nodules (e.g. Klein, 1894) and then ruminants harbouring *Lactobacillus* sp. that enable them to break down the cellulose within their diet (e.g Jensen *et al.*, 1956). These traditional examples describe mutualistic relationships between hosts and bacteria, which fulfil a specific metabolic requirement of the host not met by the host's own physiology. Bacterial species within the host can therefore function as an extended phenotype (Dawkins, 1989), allowing the host to occupy new ecological niches. As research progresses, these studies of consistent, dyadic relationships between a single host function and a single group of bacteria have

gradually been replaced by a more holistic view of dynamic microbial communities within the host, affecting function through a number of interconnected mechanisms and pathways, and the value of studying the microbiota as a whole ecological community is now recognised (Costello *et al.*, 2012; Antwis *et al.*, 2017). Therefore, in more recent studies the focus is on microbial community composition as a whole, and function is demonstrated through differing communities associated with different phenotypes or modes of life as well as through microbial imbalance associated with dysfunction (e.g. review by Lozupone *et al.*, 2012).

Filippo *et al.* (2010) found evidence that the gut microbiota of human children consuming a traditional, rural diet is likely to have evolved to maximise energy intake from fibres in a polysaccharide-rich diet as their faecal microbiota had a greater abundance of bacteria that digest cellulose and perform xylan hydrolysis than children on a more westernised diet with higher consumption of sugar and animal fats. Similarly, (Duncan *et al*., 2004) demonstrated the presence of butyrate-producing bacteria, some of which utilize lactate, suggesting a role of the gut microbiota in preventing lactate accumulation. Moreover, a microbial imbalance of the gut microbiota has been shown to result in metabolic disorders. Turnbaugh *et al.* (2006) identified that the gut microbiota contributes to pathological obesity, with an obese microbiome in mice, *Mus musculus*, having an increased capacity to harvest energy from food intake, and Ley *et al.*, (2006) showed that the proportion of *Bacteroidetes* is decreased in obese people by comparison with lean people, and increases within an individual with weight loss. Differences in the microbial composition of faecal samples attributable to several specific genera was also detected between humans with and without Irritable Bowel Syndrome (Shankar *et al.* 2013); and human infants with eczema had different abundances of *Bifidobacterium* and *Enterobacteriaceae* compared to those without (Hong *et al.*, 2010). Whilst related to the gut microbiota however, these disorders are not necessarily causally determined by the ability of the gut microbiota to metabolise nutrients. A less obvious role of the gut microbiota is in modulating healthy host immune function.

Round & Mazmanian (2010) demonstrated the role of the microbiota in immunomodulation, and early gut microbiota has been shown to shape maturation of the immune response and subsequently pathogen resistance in, for example, mice, (Chung *et al*., 2012) and domestic chickens, *Gallus (gallus) domesticus* (Crhanova *et al*., 2011). As well as direct modulation of immune development and response, the risk of pathogenic infection can also be affected through competitive exclusion and interference by the existing microbiota. Castro *et al*. (2013) suggest that the differential interference of different *Lactobacillus* strains with the adherence of bacterial vaginosis-associated bacteria, *Gardnerella vaginalis,* to epithelial cells in the human vagina contributes to the association of *Lactobacillus iners* colonisation with the likelihood of developing bacterial vaginosis. Further, a disrupted human vaginal microbiota, i.e. one that is not dominated by *Lactobacillus* species, is associated with an increased risk of contraction and transmission of HIV, even when the dysbiosis itself is asymptomatic (Borgdorff *et al*., 2014). The psychological health of the host has also been linked to the gut microbiota, with response to maternal separation in mice modulated by the composition of the gut microbiota (De Palma *et al*., 2015). Given the existence of the gut-brain axis – the connection of gut function with the emotional and cognitive centres of the brain through a number of immune, neurological and endocrinological pathways - it has been posited that the gut microbiota could modulate communication via these pathways and thus affect host psychological functioning (Carabotti *et al*., 2015). Indeed, the numerous studies in rodent model systems demonstrating a link between the manipulation of gut microbiota and anxiety and depressive like behaviours is reviewed by (Foster & Neufeld, 2013), although little is currently understood of the exact mechanism of such associations. These studies suggest that, in addition to influencing its physiology, microbiota can also have far-reaching implications for the behaviour of the host.

A clear example of microbiota-modulated host behaviour is observed in a range of social interactions. For instance, scent signals in several mammals are produced from activity of microbiota species, and provide a mechanism for social interactions, e.g. mediating kin and group recognition, and the identification of potential mates. The scent gland microbiota of spotted

hyenas, *Crocuta crocuta*, is dominated by fermentative bacteria which correlate with volatile fatty acid profiles in their scent secretions, both varying with sex and reproductive state (Theis *et al*., 2013). Similarly, the composition of the anal scent secretions were more similar within than between different groups of meerkats, *Suricata suricatta*, (Leclaire *et al*., 2014), and female meerkats were able to discriminate between anal scent secretions of kin and non-kin (Leclaire *et al*., 2013). In the fruit fly, *Drosophila melanogaster*, changes in levels of *Lactobacillus* species modulated mate choice, potentially through changes in cuticular hydrocarbons (Sharon *et al*., 2010) and pheromones in faecal pellets of desert locusts, *Schistocerca gregaria*, that promote aggregation are produced by their gut bacteria (Dillon *et al*., 2000).

Microbiota Variation

The empirical studies discussed above indicate that host health and function can be affected by variation in microbiota composition at multiple levels: across con-specific groups or populations, amongst individuals within groups, and within an individual host over time. This variation could represent deviation from a 'norm' resulting in host dysfunction, adaption of a particular species, group or individual to a habitat or life history, or there could be several different community compositions that achieve healthy host function within the same niche. Further, variation can be determined by the presence, absence or abundance of a single species, or reflect the properties of the community as a whole, such as the number or diversity of species, or the extent to which a community is dominated by a given taxonomic grouping. For example, variation in composition of the human vaginal microbiota composition could be summarised using five microbiota 'types', with four dominated by *Lactobacillus* species, and the fifth with a higher species diversity and not dominated by any given taxonomic group (Ravel *et al*., 2011). This higher diversity and lack of *Lactobacillus* domination was associated with microbiotas of the fifth type being less stable (i.e. more susceptible to change and disruption) (Gajer *et al*., 2012) and with a higher risk of vaginosis, which is often associated with dysbiosis (Ravel *et al*., 2011). Similarly, Arumugam *et al*. (2011) found that gut samples from several studies of different human populations across the world vary in a more stratified than continuous manner, and can be

grouped into three 'enterotypes' (i.e. distinct clusters with similar community composition). These enterotypes differ in their bacterial composition but all can achieve the same healthy host gut function, demonstrating that variation between bacterial communities does not necessarily lead to differences in function if the species present in each community collectively provide the same functional capabilities for their host.

This is not to say that a single microbiota member cannot impact health and host function. *Helicobacter pylori* as a single species has been linked to protection against the complications of gastroesophageal reflux disease (Vicari *et al*., 1998) and is associated with reduced risks of asthma and allergy (Chen & Blaser, 2007). However, presence of *H. pylori* can also be a risk factor for subsequent gastric cancer development (Forman *et al*., 1991). Thus, whilst individual species can be linked to host function, it is insufficient to consider a species in isolation of the rest of a bacterial community or other physiological conditions, or classify a species singly as either a consistently beneficial or detrimental microbiota member.

These examples reinforce the holistic concept that variation in microbiotas as a whole, as well as their component species in isolation, contributes to host function, or disruption of function, and thus the microbiota needs to be studied as such. As well as the range in the nature and potential effects of microbiota variation, there is also a range of potential sources of such variation. Diet is an obvious driver of variation in the microbiota of the gut, and indeed Muegge *et al*., (2011) found clustering of the microbiota according to diet across several mammalian lineages, and also found that variation in bacterial assemblages was associated with variation in functional genes present in the microbiota, including those producing carbohydrate active enzymes and proteases. Additionally, Moeller *et al*., (2014) conducted a phylogenetic analysis to demonstrate divergence of the primate gut microbiota with each evolutionary split in the lineage. They found an accelerated change in the human gut microbiota since the divergence of modern humans, probably associated with a move away from a plant-based diet and towards a diet rich in animal-derived fats and proteins; there was an increase in *Bacterioides*, associated with the

digestion of animal products, and a decrease in *Methanobrevibacter* species, which promote degradation of polysaccharides found in plants. Such variation due to diet can explain variation between groups of the same species occupying different habitats. Indeed, variation in diet associated with different habitats contributed to variation in microbial composition of wild savannah baboons, *Papio cynocephalus*, faecal samples (Ren *et al*., 2015). Social behaviour may also introduce a source of variation in microbiota either through cryptic social effects on diet or through independent effects. For example, Tung *et al*., (2015) found variation in faecal microbiota between groups of savannah baboons existing in very similar savannah habitats, with social group membership being the strongest single predictor of gut microbiome composition even after controlling for genetic relatedness. This is strong evidence for a specific effect of social interactions themselves on the microbiota, not just the sharing of the same space and eating the same food. Specifically, in this study, grooming partners had more similar communities of gut bacteria to each other than to other group members. Similarly, social group membership predicted gut microbiota similarity in red-bellied lemurs, *Eulemur rubriventer* (Raulo *et al*., 2017). Pairs of lemurs more closely associated according to indices of grooming and huddling behaviours had more similar gut microbial composition and an individual's sociality indices were negatively associated with the diversity of their faecal samples.

Additional to external sources of variation such as environment, diet and social affiliations, consistent differences in microbiota between individual hosts can be explained by individual sources of variation, including phenotypic and genetic factors. Genetically determined, species-specific antimicrobial peptides are associated with species-specific bacterial associations in four species of the cnidarian Hydra (Franzenburg *et al*., 2013), and the faecal flora of the Adelie Penguin, *Pygoscelis adeliae*, was more similar between individuals separated by a smaller genetic distance (Banks *et al*., 2009). In vertebrates, the major histocompatibility complex (MHC), the set of genes encoding immune recognition and response, is a promising genetic determinant of microbiota composition given its function in the identification and response to non-host cells, and evidence has linked variation at these genes with inter-individual variation in microbiotas; the

MHC genotype of laboratory mice contributed to variation in the profiles of bacterium-derived fatty acids (Toivanen *et al*., 2001). Similarly, diversity of the three-spined stickleback, *Gasterosteus aculeatus*, MHC Class II genes is correlated with microbial diversity in the gut, and the presence of certain MHC allele combinations is associated with variation in the relative abundance of certain microbial taxa (Bolnick *et al*., 2014). In human infants, the MHC genotype commonly associated with high risk of coeliac disease was associated with differing proportions of bacterial groups in faecal microbiota (Palma *et al*., 2010).

Microbiota composition can also change within individual hosts over time, e.g. throughout the host's lifetime as microbiotas change and develop with host age. The complexity of microbial communities in the gut of the chicken increases in the first few weeks of life, with individual communities becoming more diverse, but more similar to those of other chickens, as they approach maturity (Crhanova *et al*., 2011; Juricova *et al*., 2013), suggesting that the initial microbiota is a result of environmental inoculation with subsequent convergence of all members of the flock as the colonisation of their gut progresses. Similarly, adult black-legged kittiwakes, *Rissa tridactyla*, have a greater average number of bacterial OTUs (Operational Taxonomic Units, equivalent to species) per individual than chicks, although more OTUs were found in chicks overall, and there were pronounced compositional differences between chicks and adults (van Dongen *et al*., 2013). In humans, young children had a less diverse microbiota than adults, with the two groups separated by several differences in taxon abundances at the phylum and genus level (Ringel-Kulka *et al*., 2013), and humans over sixty-five years of age had a microbiota distinct from younger subjects, with higher variation between individuals in this older age group (Claesson *et al*., 2012).

Variation in an individual's microbiota over time can also be as a result of more acutely disruptive events than the gradual aging process. The gut microbiota of human females changes dramatically from the first to the third trimester of pregnancy, with an increase in *Proteobacteria* and *Actinobacteria*, increased diversity between mothers and reduced richness of individual

microbiotas, a shift, which may be related to pregnancy-induced immune or hormonal changes (Koren *et al*., 2012). There is also evidence of stress disrupting the gut microbiota, with 'stressed' mice deprived of food, water and bedding for 48 hours having a different microbial community in their gastrointestinal tract than control mice (Tannock & Savage, 1974). Similarly, stress induced through maternal separation leading to a decrease of culturable microflora in the gut of rhesus macaques, *Macaca mulatta*, correlated with the display of stress-indicative behaviours (Bailey & Coe, 1999). Pathogenic infection is another potentially disruptive event, and has been associated with changes in the gut microbiota of humans. Faecal samples from HIV-infected individuals have a greater diversity and richness of bacterial species and differ in the relative abundance of certain taxa compared with uninfected individuals (Lozupone *et al*., 2013). In addition, *Salmonella enterica* infection delayed the development of the gut microbiota of newly hatched chickens up to four weeks old, although the greatest effects were seen in the youngest chicks (Juricova *et al*., 2013; Crhanova *et al*., 2011). Interestingly, Deriu *et al*., (2016) found that *Enterobacteriacae* made up a significantly larger proportion of the intestinal microbiota of mice infected with pulmonary influenza even though the viral infection is normally restricted to the lung. This suggests a systemic immune mechanism of microbiota modulation, which may account for symptoms such as vomiting and diarrhoea, despite the infection being pulmonary.

The treatment of infection with antibiotics has also been demonstrated to affect microbiota composition; ciproflaxin administration affects the abundances of at least a third of the bacterial taxa in the human gut, reducing both richness and diversity, although normal intestinal function is maintained and the microbiota returns to a state similar to the pre-treatment community within four weeks (Dethlefsen *et al*., 2008). Similarly, tetracycline and streptomycin therapies resulted in changes in the relative abundances of taxa in the chicken gut microbiota, with the microbiota of the treated chickens becoming once again indistinguishable from the non-treated birds twelve days after cessation of therapy (Videnska *et al*., 2013). However, the human gut microbiota did not fully return to its pre-treatment state after the second in a series of repeated ciproflaxin therapies (Dethlefsen & Relman, 2011). Jakobsson *et al*. (2010)

demonstrated that whilst the pre-disruption state of the gut and throat microbiota prior to calrithromyocin and metronidazole treatment was generally restored post-treatment, in some individuals the changes persisted for up to four years.

The Reproductive Microbiota

Most of the work on host microbiotas has focused on the communities hosted by the gut and to a lesser degree by specific glands (e.g. mammalian scent glands). Relatively little is known about the bacterial communities living within the reproductive tract and genitalia of the host. As with much of the work investigating microbiotas, reproductive microbiotas have largely been explored with a human, clinical focus; e.g. the effect of bacteria in ejaculates on a male's fertility, or the potential role of vaginal microbiotas in protecting against the contraction of sexually transmitted pathogens.

Although varying in their conclusions relating to these clinical questions, studies of the seminal microbiota – the bacterial community of the ejaculate - have revealed the ubiquitous presence of bacteria in human ejaculates. Hou *et al*. (2013) found highly diverse bacterial communities, with the majority of samples in their study containing bacteria at a higher concentration than sperm cells, and (Weng et al., 2014a) found an average of 570 bacterial species per sample. Similarly, Javurek *et al*. (2016) found 878 species overall in a study of seminal microbiota in mice, and established that this was a unique seminal microbiota (i.e. distinct from the microbiotas of other parts of the host) as, whilst there was overlap with bacteria isolated from faecal samples, 593 of these OTUs were unique to the seminal samples. This may differ in species such as birds, where the presence of a cloaca, the common opening for the urinary, digestive and reproductive tract, may create scope for the incorporation of the gut microbiota into the avian reproductive microbiota Studies of the avian seminal microbiota usually focus on commercially relevant poultry species in the context of biosecurity (e.g. scope for sexually transmitted infections and zoonoses). Turkey, *Meleagris gallopavo*, semen samples cultured and assayed for bacterial presence contained *Campylobacter coli*, *Salmonella enteritidis*, *Pseudomona aeruginosa*,

Escherchia coli, *Escherichia hermanni*, *Staphylococcus lentus*, *Staphylococcus sciuri* and *Staphylococcus capitis* (Donoghue *et al*., 2004). Domestic fowl semen cultures have also tested positive for bacterial presence with *Staphylococcus* and *Micrococcus* the predominant genera, followed by *Escherichia*, *Enterococcus*, and *Salmonella* (Reiber *et al*., 1995). Cox *et al*. (2002) cultured *Camplylobacter* species from the semen of this species. In an earlier study, Smith (1949) found Coliform bacilli, *Staphylococcus* species, and diphtheroid bacilli in ejaculates and cloacal swab samples, but samples collected from dissection of the same birds found that the male vas deferens was sterile, suggesting contamination of ejaculates originates from its passage through the cloaca.

Early investigations of ejaculate bacteria in wild birds have relied largely on *in vitro* culturing and therefore have been limited by culturing conditions in the range of species captured. Exposure of female red-winged blackbirds, *Agelaius phoeniceus*, to bacteria in the male ejaculate during mating was suggested by the growth of bacteria in 11 of 18 aerobically cultured semen samples (Westneat & Rambo, 2000) and 42% of 233 samples in further work in the same species by Hupton *et al*. (2003). Similarly, Lombardo & Thorpe (2000) found bacteria in 19 of 30 cultured tree swallow, *Tachycineta bicolor*, ejaculate samples, some known pathogens and some potentially beneficial bacteria, with large variation between males. As well as the ejaculate communities, there has been investigation of the reproductive microbiota of males through sampling of the cloaca, which is the site of mating and, as was suggested by Smith (1949), the bacteria from which can be incorporated into the seminal microbiota. Bacterial colonisation of the avian male cloaca has been demonstrated in red-winged blackbirds (Hupton *et al*., 2003), European barn swallows, *Hirundo rustica*, (Kreisinger *et al*., 2015), some species of parrots (Xenoulis *et al*., 2010), black-legged kittiwakes (White *et al*., 2010) and house sparrows, *Passer domesticus* (Stewart & Rambo, 2000). In house sparrows, bacteria were also detected in the testes of some, but not all, males, suggesting that the cloaca is not the only potential source of seminal microbiota members.

Studies of the female reproductive microbiota have similarly been dominated by a clinical, human focus. As previously described, four of five stable community types of the human vaginal microbiota are dominated by *Lactobacillus* species (Ravel *et al*., 2011) and the other by a mixture of anaerobic bacteria. This domination of the female reproductive microbiota by *Lactobacillus* may be peculiar to humans. A study comparing the vaginal microbiotas of nine primate species found that all the non-human primate species sampled had significantly lower levels of *Lactobacillus* species than humans (Yildirim *et al*., 2014). In fact, several studies of non-human primates have shown the vaginal microbiota not to be dominated by *Lactobacillus* species at all but instead to be highly diverse and dominated by anaerobic bacteria. For instance, in species such as rhesus macaques (Spear *et al*., 2010), pig-tailed macaques, *Macaca nemestrina*, (Spear *et al*., 2012) and olive baboons, *Papio anubis* (Hashway *et al*., 2014; Uchihashi *et al*., 2015), the vaginal microbiota most closely resembles the microbiota of human women associated with adverse symptoms as a result of vaginal dysbiosis. Similarly, studies of other mammals have found that *Lactobacillus* species are often present but rarely dominate; the domestic cow, *Bos taurus*, and sheep, *Ovis aries*, vaginal microbiota are dominated by the phyla *Bacteriodetes*, *Fusobacteria* and *Proteobacteria* (Swartz *et al*., 2014), the guinea pig, *Cavia porcellus*, vaginal microbiota is majority obligate and facultative anaerobes (Neuendorf *et al*., 2015) and the minipig, *Sus scrofa domesticus*, vaginal microbiota is polymicrobial and not dominated by *Lactobacillus* species (Lorenzen *et al*., 2015). These studies are in the context of using such animals as a model system for human female reproductive health, and thus conclude that these differences merit caution, and further work, in the proper interpretation and exploration from animal models to humans.

Female reproductive microbiotas in birds may differ substantially from mammalian microbiotas, e.g. because the avian cloaca introduces more scope for incorporation of the gut microbiota into the reproductive microbiota than in mammals. A comprehensive study of the barn swallow cloacal microbiome found that females harboured similar cloacal microbes to males and that the community was dominated by the same kind of organisms one finds in the avian

digestive tract (Kreisinger *et al*., 2015). Ruiz-de-Castañeda *et al*. (2011) found several species of bacteria in the cloaca of female pied flycatchers, *Ficedula hypoleuca*, and interestingly none of these were anaerobic, but all either aerobic or facultative anaerobic. However, this study focused on culturable bacteria that could contaminate eggshells and this narrow focus could have biased the successful identification of the whole community of species. Studies have also confirmed the presence of bacteria in cloacal swabs from two species of lizards the common lizards, *Lacerta vivipara* (White *et al*., 2011), and striped plateau lizards, *Sceloporus virgatus* (Martin *et al*., 2010) and female birds including the black-legged kittiwake (White *et al*., 2010), red-winged blackbirds (Hupton *et al*., 2003), tree swallows (Lombardo *et al*., 1996) and house sparrows (Stewart & Rambo, 2000). There have also been a number of studies on bacteria found in the cloaca and reproductive tract of broiler hens (i.e. domestic chicken lines selected for meat production), with the primary focus of such work on food-borne pathogens and antibiotic resistance of livestock. These studies demonstrate the presence of *Campylobacter* (Buhr *et al*., 2002; Camarda *et al*., 2000; Hiett *et al*., 2002), *Salmonella* (Reiber & Conner, 1995) and *Lactobacillus* (Cauwerts *et al*., 2006) species in the reproductive tract of female broiler chickens, from the cloaca to the oviduct, but do not fully explore and characterise the bacterial community present.

Functional Roles and Impact of the Reproductive Microbiota

The potential role of reproductive bacterial communities in various aspects of reproductive function has largely been explored in humans and human model animals. A much-studied example for the female reproductive microbiota is its role in protection against, and transmission of, disease. Human vaginal bacterial communities dominated by *Lactobacillus crispatus* or *L. iners* were associated with a lower prevalence of HIV and other sexually transmitted infections, as well as a lower likelihood of genital shedding of HIV (Borgdorff *et al*., 2014). Additionally, Castro *et al*. (2013) demonstrated that *L. crispatus* could prevent the adherence of *Gardnerella vaginalis* to cervical epithelial cells, providing a potential mechanism for the protective function of the female reproductive microbiota. In dairy cows, disruption of the normal microbiota by parturition manifests in a bloom of pathogenic bacteria, which likely

contributes to the development of metritis (Wang *et al*., 2013). The human female reproductive microbiota has also been associated with health in terms of fertility and pregnancy; a decreased likelihood of conception and an increased risk of early pregnancy loss have both been associated with an increase in bacterial abundance, symptoms associated with dysbiosis of the vaginal microbiota, and a decrease in the abundance of *Lactobacillus* species (Eckert *et al*., 2003; Nelson *et al*., 2015). In fact, it has also been demonstrated that a mother's vaginal microbiota can affect the health of her infant after birth. Vaginally delivered babies, as opposed to those delivered by caesarean, share a significantly higher proportion of gut microbiota species with their own mother compared with other mothers for the first two years of their life. The caesarean babies had a lower gut microbial diversity, were colonised later by usual members of the human gut microbiota and had differing immune response pathways (Jakobsson *et al*., 2010).

Studies of the male seminal microbiota have generally focused on the role of the microbiota in fertility, chronic prostatitis and the load and shedding of pathogens often with reference to leukospermia i.e. the presence of white blood cells in the ejaculate. Investigations into the role of the seminal microbiota in human fertility have been either contradictory or inconclusive. Some culturing studies have found no association between bacterial growth from ejaculate samples and sperm quality (Vilvanathan *et al*., 2016; Willeén *et al*., 1996) or fertility (Shalika *et al*., 1996), whilst others have found significantly less growth and diversity in fertile groups (Toth & Lesser, 1981; Rehewy *et al.,*1979), or associations between sperm performance and certain microbial organisms (Rodin *et al.*, 2003). However, there is some *in vitro* evidence in non-human organisms of an effect of bacteria on sperm performance. Reduced sperm motility was observed when ejaculates were inoculated with cultured intestinal microbiota members in both domestic chickens (Haines *et al.*, 2013) and turkeys (Triplett *et al.*, 2016) and Vizzier-Thaxton *et al.*, (2006) observed bacteria attached to sperm in domestic chickens using electron microscopy, which could result in reduced motility and performance. Further, increased contamination of boar, *Suf scrofus,* semen with *Escherichia coli* was associated with reduced litter size (Martin *et al.*, 2010). An effect of the seminal microbiota on sperm performance could have

important consequences for sexual selection in species where polyandry results in sperm competition i.e. competition between the sperm of different males within a female over fertilisation of the ova. The seminal microbiota would mediate the outcome of sperm competition, and microbiota members that inhibited the ability of the sperm of that ejaculate to outcompete other sperm and achieve fertilisation would be selected against.

The inconsistencies of results from *in vivo* studies of fertility in humans could be due to culture-based approaches being unable to fully capture organisms of relevance, and the fact that different studies use differing media and culturing conditions, and thus are likely to produce different bacterial communities simply through selection on the different growth requirements of the community members. However, studies employing culture-independent techniques have also produced contradictory conclusions. Hou *et al*., (2013) used a culture independent method to characterise the seminal microbiota of fertile and infertile men and found that similarity of communities was not predicted by fertility status, and only an *Anaerococcus* species was negatively associated with infertility. (Weng et al., 2014a) also used a culture-independent approach, but found that species of the *Lactobacillus* and *Gardnerella* genera were higher in abundance in samples with 'normal' sperm performance, whilst *Prevotella* species were higher in abundance in low quality samples. Communities clustered into one of three groups dominated either by *Lactobacillus*, *Prevotella* and *Pseudomonas* species with samples with low quality sperm more likely to be in the *Prevotella* and *Pseudomonas* groups than that dominated by *Lactobacillus*.

There is evidence of anti-bactericidal activity and immune cell presence in ejaculates of a range of organisms, but once again their effect on both the microbiota and fertility is contradictory. For example, fruit fly ejaculates contain peptides with anti-bactericidal activity (Lung *et al*., 2001), and the ejaculates of the superb fairy wren, *Malurus cyaneus*, exhibits lysozyme-associated bactericidal activity (Rowe *et al*., 2013). Such immune activity suggests some negative effect of the reproductive microbiota upon reproductive health or success of a male, as it could act to protect the sperm from the host's own microbiota or that of the female, or

to reduce bacterial load and thus avoid attack by the female's immune response following insemination. Indeed, Otti *et al*. (2013) found reduced sperm mortality of bed bugs upon exposure to bacteria with lysozyme *in vitro*, though such an effect was not detected in the fairy wren (Rowe *et al*., 2013). Additionally, a reduction in the proportion of live sperm in an ejaculate in response to wounding was interpreted as an effect of immune activation upon fertility in honey bees, *Apis mellifera* (Stürup *et al*., 2013). However, evidence of an effect of white blood cell presence in the ejaculate is lacking in human *in vivo* studies and Eggert-Kruse *et al*. (1998) find no association between the microbiota found in the ejaculate and anti-sperm antibodies and neither Virecoulon *et al*. (2005) nor Rodin *et al*. (2003) found associations between bacteria cultured from ejaculates and leukospermia.

The effect of microbiota on reproductive health in terms of the male's own health has also been explored. For example, different bacterial species were isolated from men with chronic prostatitis compared to those from healthy men (Ivanov *et al*., 2009), and a positive association between semen bacterial load, pro-inflammatory cytokines and semen viral load found by Liu *et al*. (2014) was interpreted as evidence of a localised induction of immune inflammatory responses by the seminal bacteria leading to increased HIV viral load and shedding. If the microbiota is indeed associated with the health or virility of a male and thus his quality as a mate, it could also provide a cue to females of this quality and there is tantalising evidence of this possibility. In the review of avian microparasites (Poiani, 2010) found that species with sexual dichromatism had lower parasite richness which the authors posit as evidence for using sexual dichromatism as a signal to avoid mating with an infected individual, and male mallards, *Anas platyrhynchos,* with more colourful bills had superior anti-bactericidal activity in their ejaculates, potentially signalling that has male has less damaged sperm and fewer STIs (Rowe et al., 2011). In the red junglefowl, pathogenic infection affected male secondary sexual characteristics, and subsequently the mating choices of females (Zuk *et al*., 1990), though the parasitic infection was nematode infection of the gut in early life rather than the current bacterial community of an ejaculate. However, if the male reproductive micrcobiota could similarly affect male sexual

ornaments, or provide a cue to information about the host, this could have important consequences for sexual selection through influencing female mate choice.

Collectively, these results indicate that reproductive microbiotas play an important role in the health of both male and female hosts, and in their reproductive success (e.g. by impacting on sperm fertilising efficiency and the risk of stillbirth). In addition to these effects on the host's own health and reproductive success, the reproductive microbiota of a host also possesses the unique property of being able to influence the health and reproductive success of the host's sexual partners, through sexual transmission at mating.

Mating-Associated Variation, Disruption and Transmission of the Reproductive Microbiota

As a physically intimate act 'designed' for the efficient transfer of gametes, with close contact and exchange of bodily fluids, mating in internal fertilisers is a potential source of disruption to the microbiota, particularly the reproductive microbiota. Equally, mating creates scope for this reproductive microbiota to affect the outcomes of mating, including the likelihood of fertilisation and the sexual transmission of pathogens, as well as the reproductive health of the host more generally. Additionally, whilst the negative effects of pathogen transmission during mating are often considered, there is also the potential for the transmission of beneficial members of the reproductive microbiota (Lombardo *et al.*, 1999). Study of the reproductive microbiota and its interaction with mating is therefore important not only in terms of better understanding their role in healthy host function, but also within the body of work investigating sexual selection and the evolution of mating systems. Understanding the role of, and variation in, the reproductive microbiota adds an important contextual dimension when considering the fitness cost and benefit of mating outcomes or decisions, thus shedding light on the adaptive significance of mating behaviours or strategies, both in terms of the potential transmission of beneficial microbes or pathogens, as well as the disruption of a healthy and functional microbiota. There has been a suite of theoretical work investigating the potential for sexually-transmitted microbes to affect the

evolution of mating systems by modulating the cost of mating to a given individual (e.g. Thrall *et al*., 1997; Thrall *et al*., 2000; Kokko *et al*., 2002). The predictions of such work could be extended to incorporate the potential for a mutual interaction between mating and reproductive microbiotas (Lombardo *et al*., 1999). In his review of the occurrence and evolutionary significance of Sexually Transmitted Diseases (STDs) in birds, Sheldon (1993) highlights the dearth of good data available to systematically test the predictions generated by this body of theory.

Relatively little is known about the effect of potentially disruptive events such as mating on reproductive microbiota composition. In humans, bacterial concentration in the ejaculate is positively associated with seminal HIV1-RNA levels and having reported insertive anal sex in the past three months, whilst anti-retroviral therapy is associated with a lower seminal bacterial concentration (Korhonen *et al*., 2017). Additionally, patients with HIV have reduced bacterial species richness and diversity compared to uninfected men (Liu *et al*., 2014). Considering the reproductive microbiota of both males and females, studies of the cloacal microbiota of four avian species, found that the promiscuous and polyandrous species were more frequently infected with cloacal microparasites (Poiani and Wilks, 2000a), were more likely to harbour potentially pathogenic bacteria and that there were higher levels of such bacteria detected in the cloaca during the breeding season compared with the non-breeding season (Poiani & Gwozdz, 2002). Further, a literature review of seven studies of avian microparasites suggested that microparasites were distributed according to mating system with polygamous species having a higher species richness (Poiani, 2010). There has arguably been more of a focus on how the seminal microbiota may be transmitted to females and more generally how the female, rather than the male reproductive microbiota may be disrupted by mating. This bias makes sense, at least in terms of the effect of mating, as in internally fertilising species, the female is exposed to the reproductive microbiota of the male for much longer than the male is exposed to the female's, and thus the risk of contracting sexually-transmitted bacteria is generally higher for females than males. Indeed, (Kulkarni & Heeb, 2007) demonstrated sexual transmission of an experimentally

infected parasite was significantly more likely from the male to the female than from the female to the male in the zebra finch, *Taeniopygia guttata*. The observation of bacteria attached to chicken sperm in an ejaculate (Vizzier-Thaxton *et al*., 2006) also provides a means of transmission from male to female, which is obviously lacking in the opposite direction. Thus, variation in the female reproductive microbiota has been linked to mating history and behaviour. Higher diversity was found in the vaginal bacterial communities of females from the more promiscuous of two sympatric rodent species (MacManes, 2011) and in cloacal swabs taken from polyandrous compared with monandrous female common lizards (White *et al*., 2011), whilst across nine primate species female promiscuity (polyandry), mating group size and male testes mass (both proxies of polyandry) were associated with increasing individual vaginal microbiome diversity (Yildirim *et al*., 2014). In birds, Lombardo *et al*. (1996) found that analysis of cultured cloacal swabs of tree swallows grouped mating pairs together based on numbers and identity of species isolated. On the other hand, Hupton *et al*. (2003) found that there was no association between either the ejaculate or cloacal microbiota of a male red-winged blackbird and the cloacal microbiota of the females in his territory. Similarities in composition of the reproductive microbiota between mates have also been used to suggest the sexual transmission of bacterial species. Correlation of the composition of reproductive bacterial communities between breeding pairs above that between other community members has been demonstrated in house sparrows (Stewart & Rambo, 2000) European barn swallows (Kreisinger *et al*., 2015) and humans (Mändar *et al*., 2015), and in female baboons a history of shared sexual partners predicted similarity of vaginal bacterial communities (Miller *et al*., 2017). the effect of co-habitation rather than mating per se cannot be discounted. The cloacal microbiota of black-legged kittiwake breeding pairs became increasingly dissimilar after the prevention of insemination, with decreasing diversity of females and the loss of species from females that were shared with the male partner (White *et al*., 2010).

The signature of mating and insemination may also be found in more general disruption of the reproductive microbiota rather than (or in addition to) the transmission of specific

bacterial species. Changes in community composition have been observed following mating, demonstrating its potential as a disruptive event. In an investigation of human women infected with, or at risk of infection with HIV, the sole variable investigated that correlated with symptoms of vaginal bacterial dysbiosis in human females was detection of spermatozoa, indicating recent intercourse (Gallo *et al*., 2011). More directly, within twelve hours after mating there are shifts in the bacterial species isolated from human vaginal swabs, with increases in some species and decreases and with more pronounced changes in women whose partners have inflammatory prostatitis (Borovkova *et al*., 2011). A decrease in the relative abundance of *Lactobacillus* (Mändar *et al*., 2015), temporary increase in *Gardnerella. vaginalis* diversity (Vodstrcil *et al*., 2017) as well as an abundance in overall *G. vaginalis* (Mitchell *et al*., 2012), have all been observed following mating in human females. Functionally, a deviation from a more stable, protective state was associated with sexual activity in a study of the temporal dynamics of the human vaginal microbiota (Gajer *et al*., 2012), and an increase in vaginal *E. coli* and other gastroenteric species was observed following mating in humans both with and without condom use, but only suppressed by vaginal H_2O_2-producing lactobacilli when condoms were used (Eschenbach *et al*., 2001). This set of findings is interesting for two reasons. An increase in bacterial species even with condom use is evidence of a disruptive effect of mating in the absence of transmission of micro-organisms from the sexual partner (as prevented by the condom. Additionally lack of *E. coli* suppression with condom use suggests a further male factor that is introduced through insemination and suppresses the action of the vaginal microbiota. though to what extent this is due to transmission of microorganisms from the male or general physiological changes in the vagina due to mating cannot be determined although it is likely to be both. Changes as a result of mating could be due to disruption changing conditions unbalancing existing microbiota, or transmission of bacteria.

Mating is therefore a potentially disruptive event, at least for the female reproductive microbiota, with the potential to cause changes in the community composition, either through direct transmission from sexual partners, or due to changes in the environment as a result of

insemination or the female's own physiological response to mating. However, the field still lacks a comprehensive study fully characterising the reproductive microbiota of individual males and females in a population, and combining this with fine-grained data on socio-sexual interactions in order to establish the potential for microbial disruption and transmission through mating. Whilst some progress has been made to this end in humans (Mändar *et al*., 2015; Vodstrcil *et al*., 2017), studies using humans and wild animals are limited in the information they can realistically gain regarding subject behaviour, how often and well samples can be collected and to what extent subjects can be controlled and manipulated. Further, much of the scope for extrapolating results from humans to make conclusions regarding other species is limited.

The avian reproductive system in particular is an interesting candidate for studies of the reproductive microbiota and its dynamic interaction with mating systems and behaviours. The presence of the cloaca, the common opening for the digestive, urinary and reproductive tracts, allows incorporation of digestive bacteria into the reproductive microbiota, and the potential for the sexual transmission of digestive bacteria. Additionally, various bird species have been used as model systems in studies of mating systems. For example in studies of female mating preferences for male ornamentation in mallards (Omland, 1996), exploring patterns of extra pair paternity in the Maluridae (Rowe & Pruett-Jones, 2013), investigations of determinants of the outcomes of sperm competition in domestic fowl (Birkhead *et al*., 1999) and cryptic female choice via the effect of female reproductive fluids on sperm performance in house sparrows and Spanish sparrows, *Passer hispaniolensis* (Cramer *et al*., 2014).

Next-Generation Sequencing: Opportunities and Challenges

The study of microbiota has been catalysed by rapid advances in genetic sequencing technology, which now allow for the simultaneous sequencing of thousands of bacterial DNA sequences isolated from multiple samples, which are becoming increasingly accessible (Consortium *et al*., 2012). A particularly common approach employs next-generation sequencing technology on the Illumina MiSeq platform, as used to characterise the seminal microbiotas of

mice (Javurek *et al*., 2016) and humans (Hou *et al*., 2013), the cloacal microbiota of swallows (Kreisinger *et al*., 2015) and the gut microbiota of domestic chickens (Juricova *et al*., 2013). This approach typically sequences the 16S region to identify the bacterial species of origin of amplicons; this region of the gene encoding ribosomal RNA is highly variable across and conserved within species. Sequencing of the 16S gene region on the Illumina MiSeq platform therefore allows for the simultaneous sequencing of many amplicons from multiple samples. DNA amplicons (target fragments of DNA extracted and amplified from biological samples) are tagged with a 'barcode' and an adaptor then pooled with other samples before being loaded on to the Illumina Miseq. They are then anchored in place by the adaptor to the Illumina cell, and fluorescently labelled nucleotides are added to the cell, which attach to the first available complementary nucleotide of the anchored amplicon and are imaged before more nucleotides are added to attach to the next nucleotide of the amplicon and the process repeated. This then provides a sequence of images revealing the sequence of the amplicon anchored at each position, including the sequence of the barcode which indicates the sample of origin of each amplicon. This approach offers resolution of bacterial RNA to the species (or Operational Taxonomic Unit, OUT) level, through comparisons with existing sequence database the identification of many of the taxa present, thus allowing for the analysis of the composition and structure of complex bacterial communities.

However, whilst this approach overcomes the biases in the organisms identified often suffered by *in vitro* culture-based methods, and is much more time efficient than early culture-independent sequencing methods, it presents a new set of challenges. These include contamination of samples, amplification bias at several stages of the process, and variation in the number of sequences produced per sample. To some extent, sterile procedures and conditions should control for control for contamination. However, a particularly difficult source of contamination to eliminate is laboratory reagents. Lusk, (2014) reports that the issue of contamination in microbiota studies is widespread, and Salter *et al*., (2014) find that contamination in 'sterile' DNA extraction kits changes the conclusions drawn from sequencing

work. It can be difficult to remove these contaminants without compromising the integrity of the reagents (Champlot *et al*., 2010). Despite its undoubtedly ubiquitous nature, the challenge of contamination, nor its resolution, is rarely quantified or discussed in microbiota studies employing next-generation sequencing. For example, in four studies employing next-generation techniques to study male reproductive or digestive microbiotas (Hou *et al*., 2013; Javurek *et al*., 2016; Kreisinger *et al*., 2015; Videnska *et al*., 2013) contamination is not reported, nor ways of minimising or addressing it. There is therefore no established procedure to minimise or address such contamination issues, and so the initial aim of this work was to address this in our own microbiota pipeline, working to minimise contamination, assess the effectiveness of such measures and produce a method to deal with the remaining contaminants.

Another challenge produced by next-generation sequencing techniques is the potential for amplification bias by the different barcodes used to identify samples of origin of the different amplicons (Aird *et al*., 2011). For instance, if each barcode binds differentially well to different bacterial taxa this could lead to differences between communities that are an artefact of barcode amplification rather than genuine biological differences between communities. Further, despite the standardisation of DNA prior to next-generation sequencing to ensure that the same amount of bacterial DNA is sequenced for each sample the number of total sequence reads produced (the sampling depth) commonly varies widely between samples (e.g. Kreisinger *et al*., 2015; Filippo *et al*., 2010; Videnska *et al*., 2013). This is analogous to differing sampling efforts between samples and is often dealt with by rarefaction to the number of reads in the smallest sample (e.g. Kreisinger *et al*., 2015).

The Thesis

The overarching goal of my doctoral thesis is to characterise male and female reproductive microbiotas in a polyandrous population of birds, explore patterns of variation and underpinning biological drivers. To do this, I studied a captive population of red junglefowl, *Gallus*

gallus, by combining next-generation sequencing approaches with *in vitro* culturing as well as mating and insemination experiments.

The red junglefowl is the wild ancestor of the domestic chicken (Fumihito *et al*., 1994; Fumihito *et al*., 1996). Red junglefowl populations are socially structured (Collias & Collias, 1996) and both males and females mate multiply, with variation in male mating success determined by dominance status (Collet *et al*., 2012). This variation in male mating success provides natural variation to investigate the effect of mating on the reproductive microbiota, and this polyandry, combined with the ability of female fowl to store sperm for prolonged periods (Pizzari & Birkhead, 2000) means that ejaculates compete for fertilization within the female. If bacteria within the ejaculate or the female reproductive tract can affect sperm function, as suggested by *in vitro* studies (Haines *et al*., 2013; Triplett *et al*., 2016) then microbiota composition can mediate this competition and will have important consequences for sexual selection. Just as sexual selection can act on male aggressiveness (McDonald *et al*., 2017) and sperm allocation (Gillingham et al., 2009), post-copulatory sexual selection could also act on microbiota, selecting for male microbiota less damaging to host function or ejaculate characteristics better able to resist the damaging effect of both male and female microbiotas. Further selection could also be imposed by cryptic female choice; female junglefowl are able to bias sperm retention against related partners (Løvlie *et al*., 2013) so could do the same to males with more harmful microbiota, or lower sperm quality due to microbiota, in response to either a pre- or post-copulatory cue. The red junglefowl therefore provides a very relevant system for the study of reproductive microbiota, and work exploring its interaction with mating systems and behaviours would lend to context to studies such as those cited above relating to sexual selection and the evolution of mating traits through understanding how the microbiota may mediate such processes.

The red junglefowl is also a relevant system in which to study the Major Histocompatability Complex (MHC) as a potential as driver of variation in microbiota composition; there has been shown to be a MHC heterozygote advantage in pathogen resistance

(Worley *et al*., 2008). Additionally, MHC-based-mate choice has been demonstrated in both males (Gillingham *et al*., 2009) and females (Løvlie *et al*., 2013) has been demonstrated and thus any MHC-driven microbiota variation could provide a microbiota-mediated cue allowing MHC-based mate choice, or provide a reason for choosing MHC-dissimilar mates.

Further, prior work means there is much information available regarding our captive population and procedures for the study of the animals. For instance, in methods of ejaculate collection (e.g. Gillingham *et al*., 2009)and artificial inseminations (e.g. Løvlie *et al*., 2013), mating behaviour assays (e.g. McDonald *et al*., 2017) and understanding and typing of the red junglefowl MHC (e.g. Kaufman *et al*., 1999; Worley *et al*., 2008).

Thesis Aims & Structure

The thesis develops over four integrated data chapters. In the first data chapter, I develop a work pipeline Using this approach and semen samples collected from males across multiple breeding seasons, I aim to characterize the seminal microbiota, and compare different types of samples, to see how closely each sample microbiota reflects the microbial communities contained in natural ejaculates and how related the seminal microbiota is to the faecal microbiota. I develop a methodological pipeline based on next-generation sequencing technology on the Illumina MiSeq platform to characterise the microbiota of male semen (seminal microbiota) and explore patterns of variation, while addressing the often unexplored methodological issues of contamination, rarefaction issues and protocol biases. The proposed pipeline reduces the risk of contamination and deals systematically with residual (e.g. reagent-linked) contamination and methodological issues such as barcoding amplification. Using this methodology, I show that male semen hosts a rich microbial community with an overall 698 OTUs, and on average 65.5 OTUs per sample. I show that this seminal microbiota largely overlaps with the faecal microbiota of these birds, suggesting an important role of the cloaca for the integration of gut microbiota. Importantly, I show that different males produce semen samples with consistently different microbial communities and that this variation is driven by specific clusters of bacterial species.

Whilst high throughput sequencing approaches allow us to gain a picture of bacterial species present using the DNA isolated from a sample, they do not tell us the viability of the organisms identified. Therefore, in my second data chapter, I combine next-generation sequencing with in vitro culturing to test whether bacteria detected in male semen are alive, and explore the ecology of seminal microbiotas by investigating differences in communities cultured under different environmental conditions. Semen samples were cultured in aerobic and anaerobic conditions and on three different media, and the resulting bacterial cultures harvested and characterised using our next-generation approach. I demonstrate that these bacteria are alive in the ejaculate, validating the investigation of theories based on the function of these bacteria in the junglefowl ejaculate or reproductive tract, as well as their potential for transmission to, disruption and colonisation of a female. Additionally, through investigation of the effect of culture conditions I demonstrate that the structure and composition of the bacterial community produced varies with culturing condition, aiding interpretation of differing results in previous culture based studies and informing design and interpretation of future culture based investigations.

In my third data chapter, I investigate biological sources of within- and between-male variation in seminal microbiota composition. First, I consider temporal variation across a breeding season, then age, cohort and immunogenetic effects using ejaculate samples taken from males across two separate breeding seasons, with repeated samples from several males for which information on two MHC loci is available. I find that allelic richness and amino acid divergence at an MHC locus were the only factors which significantly explained variation in community composition of ejaculates. I then focus on the effect of mating on the seminal microbiota, using a subset of these males that were sampled immediately after inclusion in a mating group with females, where detailed information on the mating behaviour, mating rates and the identity of mates was recorded for each male. I find evidence suggestive of an effect of mating on the seminal microbiota as mating group significantly explained variation in community composition and structure between samples, and within group variation was linked to the group's sexual network

metrics. However, this mating effect was not realised in analysis of mating metrics at an individual level.

In my fourth data chapter, I focus on the microbiotas of the female cloaca and investigate the influence of insemination from males on the female cloacal microbiota in terms of disruption and transmission. Using artificial insemination and sampling experiments carried out in a past breeding season I track changes in the microbiota of virgin females throughout a series of artificial inseminations, and for a period of time after inseminations are stopped. I find an effect of insemination upon the reproductive microbiota of the female red junglefowl, particularly the initial insemination of a virgin female, and a potential return towards the virgin state of a female with loss of male OTUs following the cessation of inseminations.

CHAPTER 2. THE SEMINAL MICROBIOTA OF THE MALE RED JUNGLEFOWL: METHODOLOGICAL APPROACHES AND PATTERNS OF VARIATION

Rosie Eccleston[a], Stephen Preston[b], Kiyono Sekii[c], Adrian L. Smith[b], Tommaso Pizzari[a]

[a] Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK

[b] Department of Zoology, University of Oxford, Oxford, UK

[c] Department of Biology, University of Hirosaki, Hirosaki, Japan

ABSTRACT

Microbiotas are emerging as fundamental modulators of the host ecology, evolutionary biology and health. Microbiotas inhabiting the reproductive tract of the host, should be particularly relevant as they can have a direct impact on the host reproductive success and potential for transmission through socio-sexual interactions. Yet, we know almost nothing of the reproductive microbiotas in non-human organisms, a gap in knowledge that in part reflects the challenges of studying reproductive microbiotas arising from contamination and other methodological biases. Here, we study the microbiota of male semen in a captive population of red junglefowl, *Gallus gallus*, using 16S rRNA Illumina Next-generation sequencing. We first present a work pipeline to reduce multiple sources of contamination and an analytical approach to deal with residual contamination, rarefaction issues and laboratory biases. Using this approach, we then characterize the semen microbiota, comprising of 698 Operational Taxonomic Units. We show that these microbiotas largely reflect the microbial communities contained in natural ejaculates (and thus transmitted to females during mating), and that there is large overlap with the faecal microbiota.

INTRODUCTION

Microbiotas, the microbial communities living in or on a host organism, can be considered as the host's 'extended phenotypes', not strictly speaking part of the host body but playing an important, in some cases essential, role in host function. This can be physiological, with evidence of the gut microbiota allowing efficient nutrient uptake from

diet in humans (Filippo *et al*., 2010) and influencing the maturation of the immune systems in domestic chickens, *Gallus (gallus) domesticus*, (Crhanova *et al*., 2011). Microbiota can also influence host behaviour through the mediation of interactions between hosts, for instance there is evidence of scent gland microbiota mediating kin recognition in Hyenas, *Crocuta crocuta* (Theis *et al*., 2013). The host is therefore expected to generate selective pressures, e.g. through physiology or behaviours to promote and maintain favourable communities of bacteria and eradicate unfavourable one (e.g. diet drives convergence in the gut microbiota across mammalian species, (Muegge *et al*., 2011)). Thus, as well as microbiotas being crucial to the understanding of host physiology and ecology, they are also crucial in understanding their evolution.

Reproductive microbiotas, the bacterial communities living within the reproductive tract, are particularly relevant as modulators of host fitness. First, because reproductive microbiotas can influence a host's reproductive health, and reproductive success; for example, the vaginal microbiota can protect against HIV infection in humans (Borgdorff *et al*., 2014), and bacterial load in the male ejaculate is associated with inflammatory prostatitis (Ivanov *et al*., 2009). Second, reproductive microbiotas can have considerable impact on the reproductive success of a host. In females, reproductive microbiotas can influence the risk of miscarriage and preterm birth (Eckert *et al*., 2003; Nelson *et al*., 2015). In males, the microbiotas of the vas deferens and the ejaculate can impact on the chances of fertilisation, though *in vivo* evidence is contradictory e.g. in boars, *Sus scrofa*, increased *Escherichia coli* concentration of ejaculates was associated with reduced litter size (Maroto Martín *et al*., 2010) but Hou *et al*. (2013) found that the fertility status of men did not predict similarity in bacterial composition of samples. An impact on fertilisation could be imagined through the direct action of bacterially produced compounds upon gametes, or through induction of an immune response to the microbiota which then targets gametes. Reduced sperm motility was observed when ejaculates were inoculated with cultured intestinal microbiota members in both domestic

chickens (Haines *et al*., 2013) and turkeys, *Meleagris gallopavo,* (Triplett *et al*., 2016). Third, reproductive microbiotas can also modulate host fitness through influencing the fitness of a host's sexual partners through sexual transmission. The potential for sexually transmitted infections (STIs) to influence the evolution of host mating systems and behaviours has been demonstrated theoretically (e.g. Graves & Duvall, 1995; Thrall *et al*., 1997; Thrall *et al*., 2000; Kokko *et al*., 2002) and this could be expanded to include the influence of microbiotas, both through the mediation of STI transmission and contraction, and through the transmission of microbiota members themselves.

The presence of bacteria in male semen has been demonstrated in a number of species - mammals (e.g. Javurek *et al*., 2016), birds (e.g. Westneat & Birch Rambo, 2000), insects (e.g. Andere *et al*., 2011) and reptiles (e.g. Johnston *et al*., 2014). Variation in the bacterial load and composition of semen has been linked to host health in humans, including in the incidence of inflammatory prostatitis (Ivanov *et al*., 2009) and in the viral load and shedding of HIV (Liu *et al*., 2014). However, most work exploring variation in the ejaculate microbiota and its possible causes and consequences has been done in humans, and often with a clinical focus, meaning a subset of the population, rather than a cross-section, is focused upon. Recently, Javurek *et al*., (2016) have presented the seminal microbiota of the mouse, *Mus musculus*. In this study, however, the seminal fluid was extracted through dissection of the seminal vesicles, thus it is unclear whether these results represent the microbiotas of ejaculates produced during copulation.

Birds represent a particularly interesting clade for the study of seminal microbiota and the ecological and evolutionary implications of sexually transmitted infections (Sheldon, 1993; Lombardo et al., 1999; (Poiani and Wilks, 2000b). The presence of a cloaca, the common opening for the digestive, urinary and reproductive tracts in the majority of bird (and some reptile) species, increases the potential for the incorporation of gut microbiota into the reproductive and seminal microbiota. Evidence

of bacteria in the semen of a number of avian species has been demonstrated in red-winged blackbirds, *Agelaius phoeniceus,* (Hupton *et al*., 2003), tree swallows, *Tachycineta bicolor* (Lombardo & Thorpe, 2000) domestic turkeys, (Donoghue *et al*., 2004) and domestic chickens (Reiber *et al*., 1995). However, these studies use culturing methods subject to potential biases associated with the different growth requirements of different bacterial species. Here, we characterise the microbiotas present in the semen of male red junglefowl, *Gallus gallus*, a species of birds with a highly promiscuous mating system, with variation in male mating success (Collet *et al*., 2012) providing natural variation to explore the relationship between the microbiota and mating rate. Polyandry and sperm retention by females (Pizzari & Birkhead, 2000) create sperm competition between co-occurring ejaculates within the female, and the ability of the female to bias retention against certain mates (e.g. Gillingham *et al*., 2009). This creates an opportunity for the seminal microbiota to influence sexual selection for instance, if it can affect sperm performance or provide a pre- or post-copulatory cue to bias female sperm retention.

Establishing a system of characterisation of seminal microbiotas in red junglefowl can therefore provide a platform to investigate patterns of variation in the reproductive microbiota and its interaction with, and consequences for, host reproductive health and success and consequently mating systems in this and other species. In particular, since the red junglefowl is the ancestor of the domestic chicken, *Gallus (gallus) domesticus* (Fumihito *et al*., 1996), this work can have considerable applications for the management of the biosecurity, health and production of commercial poultry flocks.

Next-generation sequencing techniques are currently considered the most time and cost-effective method of characterizing microbial communities and have been employed in the study of the seminal microbiota of humans (Hou *et al*., 2013) and mice (Javurek *et al*., 2016); the cloacal microbiota of the barn swallow, *Hirundo rustica* (Kreisinger *et al*., 2015); and in the study of gut microbiota in domestic chickens

(Videnska *et al*., 2013). To characterise the red junglefowl seminal microbiota, both qualitatively and quantitatively, we used a next-generation approach to sequence the 16S ribosomal RNA (rRNA) gene. This gene is common to all eukaryotic organisms and is conserved within species, so is a useful target to detect and distinguish a wide range of bacteria. In this study, we used the next-generation sequencing platform Illumina MiSeq, which enables the simultaneous sequencing of thousands of DNA amplicons from several different samples through a 'barcoding' system, capable of identifying the sample of origin of each of the resulting sequences. The high-throughput sequencing technology involves the attachment of many bacterial amplicons to a sequencing plate and subsequent amplification creating 'DNA clusters', which in turn allows the simultaneous sequencing of a large number of amplicons through the attachment and visualisation of nucleotide-specific fluorescent markers. As valuable an asset as these techniques are in the accurate characterisation of complex microbial communities, their sensitivity, the complexity of the process and the sheer volume of the information produced creates a whole new set of challenges, which are often left unaddressed, or unreported. Two potential sources of bias are particularly important: contamination and laboratory biases.

Regarding contamination, as the protocol is designed to extract, amplify and identify very small amounts of a wide range of bacterial DNA, any non-sample bacterial DNA that is erroneously incorporated into the sample at any point between sample collection and sequencing, will also be inadvertently sequenced, and outputted by the sequencing platform as a component of the original sample. Thus, even seemingly minor contamination from the laboratory environment, reagents and other samples can obscure the characterization of the genuine microbial community of a sample, and may lead to erroneous conclusions as to composition and variation in the microbiota of interest. Despite the widespread issue of contamination (Lusk, 2014), and the demonstration of potentially contaminating DNA in supposedly sterile reagents, and their potential to

influence the results of microbiota studies (Salter *et al*., 2014), very few studies have explicitly addressed this issue. For instance, out of the four studies of microbiota cited above (Hou *et al*., 2013); Javurek *et al*., 2016; Kreisinger *et al*., 2015 and Videnska *et al*., 2013), none report laboratory contamination, or any measures taken to minimise such contamination and its effects either at the laboratory or analysis stages. Thus, there seems to be no commonly followed protocol or known list of common contaminants that can be used to inform procedure and analysis in managing this problem in studies similar to ours.

Second, the multistage process of preparing the samples for sequencing presents a number of factors between the stage of collection and that of sequencing, which may affect the bacterial composition of the samples as a result of laboratory protocols. In particular there is potential for amplification bias due to the different barcodes used to identify samples of origin of the different amplicons (Aird *et al*., 2011). The differential attachment of barcodes, which allow the identification of the sample of origin of a sequence, to individual bacterial taxa, may lead to bias in the amplification of different bacterial taxa, resulting in sample composition differences that do not represent biological variation. Despite standardization of DNA concentration for each sample before submission for sequencing, previous studies (e.g. Kreisinger *et al*., 2015; Filippo *et al*., 2010; Videnska *et al*., 2013) suggest that disparities remain in the number of sequences produced at the end of Illumina MiSeq sequencing and initial quality control. As with any study of ecological communities, the composition of the community sampled, and thus the conclusions based on the comparisons of these communities, is affected to some extent by the amount of sampling effort, in this case the number of sequences. It is therefore important that, in order to accurately compare samples, a method of assessing the extent to which they represent the actual sample community, and of standardising this across samples, needs to be developed and the potential for any variation and subsequent

method of standardisation to affect conclusions based on the resulting communities needs to be evaluated.

An additional challenge that is specific to studies of seminal microbiotas is that samples do not always represent ejaculates naturally produced by males, raising questions as to their biological relevance. In birds, two techniques for semen collection exist; manual stimulation and controlled copulation. In poultry, manual stimulation is obtained through 'abdominal massage' where gentle pressure is applied to the abdominal muscles of the male, resulting in a reflexive ejaculation (Burrows & Quinn, 1937). In controlled copulations, a male mounts and mates with a female. In some species, a subset of males will perform when exposed to a female dummy fitted with a 'false cloaca' (Pellatt & Birkhead, 1994; Westneat & Birch Rambo, 2000). In poultry, males can be exposed to a live female, which wears a harness to prevent insemination and ejaculate samples are collected on this harness (Pizzari, 2007). Both these methods have been used previously to describe ejaculate characteristics in red junglefowl (e.g. Gillingham *et al*., 2009; Løvlie *et al*., 2013) and each may provide complementary information on the seminal microbiota. For instance, abdominal massage is thought to capture the majority of the extragonadal sperm reserves stored in the vas deferens of a male (Etches, 1996), and may contain a greater amount of faecal contamination. Controlled copulation samples on the other hand, more closely represent the natural ejaculates that males pass to females, and as such will display a higher amount of intra-male variation. The preferred sample method will therefore depend on the specific aims of the study, and one should compare the two types of sample to understand if there is a systematic difference between them, and what the nature of this might be; in particular how much does an abdominal massage sample represent a natural copulation.

The aims of this chapter were to:

1) Develop a methodological pipeline for the sampling, processing and analysing of the seminal microbiota of the red junglefowl, which seeks to establish and reduce sources of methodological bias;

2) Characterise the seminal microbiota and describe patterns of variation within individual males and across the population;

3) Compare microbiotas between semen samples obtained through manual collection and natural ejaculates produced by males while mating with a female, as well as between semen samples and faecal samples to explore the relationship between gut and reproductive microbiotas in male birds.

MATERIAL AND METHODS

Study population

We studied a captive population of red junglefowl, housed in outdoor aviaries at the University of Oxford John Krebs Field Station in Wytham, Oxfordshire. This population (~200-250 adults) was established in 2006 and has been the subject of research on different aspects of reproductive behaviour and sexual selection. Samples used here are from 65 sexually mature males sampled in the 2013 and 2015 breeding seasons (from May to September each year), between the ages of one and nine years old. (2013: 23 one years old (yo), 6 two yo, 11 three yo, one 5 yo, one 7 yo; 2015 16 two yo; 12 three yo, 4 4 yo, 5 5 yo, one 7 yo, two 9 yo).

Adult males are housed in pens in male-only groups of six to thirty individuals of mixed age and relatedness. Throughout the periods of sample collection, birds were routinely moved across pens and groups as a result of husbandry and logistic considerations imposed by other experimental work on this population. In addition, some males were temporarily deployed in mating trials, where they were housed in mixed-sex groups for periods of ten days, before being returned to the male-only groups.

Samples were collected opportunistically when males became available (e.g. caught for blood sampling or to be moved across pens). Sampling effort in 2015 focussed largely on four male-only pens where more systematic sampling sought to obtain repeated samples from individual males. Sample collection protocol followed was consistent – as was the bird husbandry – throughout and across the two seasons.

Sample collection and processing

Males were sexually-rested for at least 48 hours to restore maximal extragonadal sperm reserves (Etches, 1996). Seminal samples were obtained from male red junglefowl using one of two established techniques: abdominal massage and controlled copulation. Abdominal massage takes advantage of the ejaculation reflex in the male following pressure on the abdominal muscles (Burrows & Quinn, 1937). Abdominal massage was performed in a procedure room separate to the pens housing the birds, and samples were directly into a sterilized microtube. All handling of the birds and samples was done with clean, gloved hands and wearing overalls. Gloves were replaced and cleaned between each bird. Efficient handling of the bird and sample minimized the amount of time the sample was exposed to the environment; the tube was opened only when the abdominal massage had begun and closed within one minute of opening and immediately following sample collection. If a sample had not been produced within one minute of the tube being opened, the tube was disposed of and a fresh sterile tube utilised. Controlled copulation sampling took place in pens housing males. The experimental female was fitted with a specially designed harness, which covers her cloaca to prevent insemination, placed on a plastic pad on the ground, and gently held in a sexually soliciting position in front of a male (Pizzari, 2007). A single male was allowed to mount and copulate with the female, and when this was completed, the female was quickly removed from the pad, and the pad and harness removed from further contact with the birds. The ejaculate was immediately collected from the harness and pad using a pipette and sterile filter tip, and placed in a

sterile microtube before being taken to the laboratory for processing within five minutes of ejaculation. The amount of time the sample was exposed to the environment was minimised; the tube was only opened once ejaculation occurred and the sample transferred into the tube and sealed within one minute. The harness was thoroughly cleaned with ethanol before each copulation, and covered with smooth, non-absorbent tape to facilitate this. The pad was covered with a disposable plastic sheet, which was replaced and cleaned with ethanol before each copulation. All the equipment used repeatedly in collecting samples (e.g. pipettes, filter tips, pads, harnesses, glove boxes, sample boxes and marker pens) was cleaned with ethanol before and after every use and stored in sealed containers. As in abdominal massage, all handling of the birds, samples and equipment was done with clean, gloved hands and wearing disposable overalls. A schematic representation of the workflow of the pipeline is presented in Fig. 1.

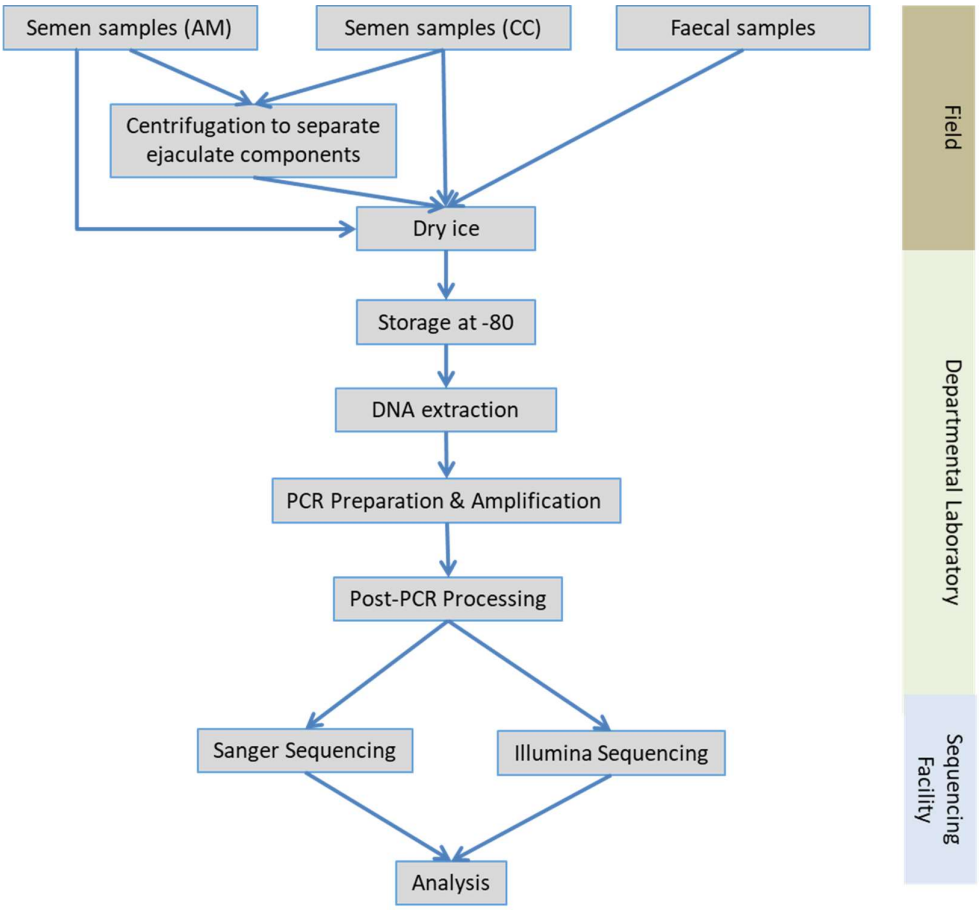


Figure 1 *Schematic representation of the workflow of sample collection and processing. AM indicates collection by the abdominal massage method and CC collection by the controlled copulation method.*

Semen samples were processed in the laboratory of the field station within five minutes of collection by either method. The laboratory surfaces were cleansed thoroughly with DNAZap PCR DNA Degradation Solutions (ThermoFisher Scientific), a potent nucleic acid degrading reagent, before and after each use and clean gloves were worn and replaced between handling of each sample. Pipettes were used only for this work and were used with filter tips. If the semen sample was sufficiently large, it was divided into three sample types; whole semen, seminal fluid and sperm pellet. Up to five 10uL whole semen aliquots were taken from each sample before centrifugation, and a 1uL aliquot placed in 200uL Phosphate-Buffered Saline for motility analysis. The remaining sample was centrifuged at 400g for five minutes to obtain a sperm pellet and seminal fluid supernatant. The supernatant was then pipetted off as the seminal fluid aliquot, leaving the sperm pellet behind. When the sample volume was not sufficiently large, only the whole semen aliquot was used. Once processed, all samples were immediately placed on dry ice before storage at -80°C.

Additionally, faecal samples were obtained from individuals at the time of semen sampling. Birds were temporarily housed in cardboard poultry transport boxes whilst awaiting sampling, and afterwards before being returned to the pen. The boxes were lined with clean plastic sheets, which had been sterilized with ethanol. Faeces produced whilst the bird was in the box were collected into sterilized microtubes using sterile plastic spoons as soon as the bird was removed from the box. The sample was therefore a maximum of fifteen minutes old when it was collected and was placed on dry ice within five minutes of collection before storage at -80°C. Frozen semen and faecal samples were taken to the Department of Zoology within the same day of collection for further processing.

DNA Extraction

DNA was extracted and purified using the DNeasy Blood and Tissue Kit (QIAGEN) and according to the manufacturers guidelines, with temperatures, timings and volumes optimised for bacterial DNA extraction for either semen, cloaca or faecal samples. Briefly, for semen samples, 200uL lysis buffer mix and 190uL water was added to a 10uL sample then incubated at 37°C for 30 minutes. 50ul Proteinase K were then added, followed by 400uL Buffer AL, then incubated at 56°C for 70 minutes. 8uL RNase were then added and the mix incubated at 37°C for 10 minutes. Extraction for faecal samples was performed in the same way, with the same reagents and protocol but InhibitEx Tablets (QIAGEN) were added to the sample, and the samples centrifuged prior to lysis to allow pelleting and removal of any PCR inhibitors in the faecal matter.

DNA was purified by precipitation with 750uL of 100% ethanol and passed through a DNEasy spin column, then washed with 500uL Buffer AW1 then again with 500uL of Buffer AW2 and eluted with 60uL molecular biology reagent water. Purified DNA was frozen and stored at -20°C.

PCR Preparation and Amplification

PCR mix for a 25uL PCR reaction was prepared with 0.25uL Phusion High-Fidelity DNA Polymerase (New England Biolabs), 5uL 5X Phusion HF Buffer (New England Biolabs), 0.5uL 10mM dNTP mix (Sigma-Aldrich), 17.25uL molecular biology reagent water (W4502, Sigma-Aldrich) and 0.5uL each of 10uM forward and reverse bacterial primers 347F and 803R to amplify an approximately 456bp fragment containing the V3 and V4 regions of bacterial 16S rDNA.

This primer pair has been used in previous studies investigating the faecal microbiota of the domestic chicken (Videnska *et al*., 2013) and the forward primer, 347F, also has a 6bp long barcode sequence attached, designed in-house as described by Bystrykh (2012). As sample DNA was pooled following PCR Amplification and before the

sequence run, a different barcoded 347F (taken from a set of 100, all with a different 6bp in-house designed barcode attached) was used for each sample so that the 6bp sequence – unique to each barcode - can then be used to establish the individual PCR reaction of origin of Illumina MiSeq sequence reads. These barcodes are different to the sequencing barcodes attached to PCR amplicons later in the pipeline. The sequencing barcodes are attached prior to Illumina MiSeq sequencing and identify the Illumina library of origin of sequence reads. The two barcoding levels in combination allow the in-house barcodes to be used more than once in a sequencing run, and together indicate the sample aliquot of origin of the sequence reads.

'The PCR was carried out according to the Phusion manufacturer's protocol and with reference to the primer melting temperatures. One uL of the DNA template was added to each reaction mix and for negative controls, 1uL molecular biology reagent water was used in place of the DNA, and 0.5uL lab cultured *Escherichia coli* DNA for positive controls. There was one positive and up to three negative controls for each PCR plate used. DNA from these positive controls was used to assess PCR amplification success through gel electrophoresis visualisation but was not sequenced.'

In order to avoid contamination from the DNA product of previous amplifications with the same target, DNA extraction and PCR preparation was performed in a separate laboratory to that where amplification and post-amplification processing was performed. Separate sets of equipment and reagents are also used (including pipettes, filter tips, plasticware, glassware, gloves and laboratory coat) and kept in sealed boxes. All plasticware and reagents were specified for use in molecular biology work to keep them contaminant-free as far as possible. Where reagents required dilution or aliquoting into smaller amounts this was done under the same sterile conditions as the DNA extractions and PCR preparation.

All surfaces were cleaned thoroughly with DNAZap PCR DNA Degradation Solutions before each use and were exposed to UV radiation for 1 hour in order to avoid contamination from previous experiments of my own and other laboratory users. Pipettes were also cleaned with DNAZap Solutions before each use and exposed to UV radiation for one hour along with microtubes, PCR plates and pipette filter tips to prevent cross-contamination from previous use and from any environmental contamination that has occurred during storage or handling. All work was done under a laminar flow hood to reduce environmental and experimenter contamination, but with no air flow to reduce the risk that air movement might possibly result in contamination of samples with microbial DNA on surfaces, within the air system or from other samples. Gloves were changed regularly and whenever anything from outside the hood is handled. Samples did not enter the hood apart from for the few seconds it took to load DNA template into the reaction mix.

The PCR regime consisted of an initial heat activation of one minute at 98°C, followed by 30 cycles of denaturation at 98°C for 15s, annealing at 51°C for 15s and elongation at 72°C for 30s, then a final extension at 72°C for two minutes. Amplification was always performed in the same Veriti 96-well Thermal Cycler (Applied Biosystems).

Post-PCR Processing

Amplification of the target fragment was confirmed, and amount of DNA in the PCR product quantified, by adding 1uL glycerol loading dye to each sample, which was then run for one hour at 70V on a 1% agarose gel stained using ethyl bromide. The gel was then viewed under UV light, photographed and the sample bands compared with those of a 50bp Hyperladder (Bioline) to establish that there was signal of DNA in the amplification product, and that DNA fragments matched the target fragment size.

Sanger Sequencing

We conducted a preliminary investigation to assess the success of DNA extraction and PCR amplification, as well as the effect of contamination reduction techniques, on a set of semen samples and PCR water negative controls, which were Sanger-sequenced prior to the commencement of Illumina MiSeq sequencing. PCR amplification products from five semen samples and 13 water negative controls from before, during and after the implementation of the strict contamination reduction protocol outlined above, were cloned in competent *E.Coli* (NEB 10 beta Competent *E. coli*) using the pGEM-T Easy Vector System I (Promega), and according to the manufacturers' protocols as follows.

In separate reactions, amplicons produced from each sample/control were inserted into the vector (through ligation when incubated for an hour at room temperature with a DNA Ligase) which was used to transform chemically-competent *E.coli* (NEB) cells. The resultant amplicon-containing bacteria were spread onto a solid agar medium in a petri dish, one dish per sample/control, and incubated at 37°C for 48h to produce colonies, each colony made up of clones all containing the same amplicon. Twelve colonies per sample were picked at random, vector plasmid DNA isolated from the cells using the Wizard Plus SV Miniprep DNA Purification System, and the original target DNA fragment released from the plasmid using a single enzyme digestion and PCR amplified. The resultant amplicons were Sanger sequenced using Big Dye terminators, using the Zoology department's sequencing facility. The resulting sequencing output was uploaded into the Geneious version 7.1 (Kearse *et al*., 2012), and taxonomic identity of the sequence (Operational Taxonomic Unit, OTU) assigned through the Basic Local Alignment Search Tool (BLAST) (Altschul *et al*., 1990) and NCBI nucleotide database (National Center for Biotechnology Information, 1988). The taxonomic level of the OTUs was not consistent as BLAST was able to identify some sequences to a finer taxonomic level than others. All taxonomic information was retained, rather than reduced, such that

all OTUs were defined to the finest taxonomic level possible rather than to the taxonomic level available for all sequences. This approach did not affect our quantitative or qualitative outcomes, as it does not change the number or classification of OTUs isolated in each sample, only the nomenclature.

Illumina MiSeq Sequencing

Sample PCR products were combined in DNA libraries at approximately equal concentrations. Between 30 and 80 samples were multiplexed per library, with approximately 150 samples per run (up to five libraries in a single run), and the libraries combined and DNA quantified such that approximately the same amount of DNA per sample were used in each run. PCR product from positive controls was not sequenced, but up to three negative water controls from each PCR amplification plate were sequenced, 36 PCR negatives in total. Electrophoresis gels were used to give a relative estimate of DNA concentration of the PCR product obtained from each sample, and the band was classified to low, medium and high intensity signal. 1.25uL, 2.5uL and 5uL were added to the library for high-, medium- and low-band intensities, respectively. The samples were then purified using QIAGEN MiniElute PCR Purification Kit according to manufacturer's instruction and quantified using Qubit dsDNA BR Assay Kit, again to manufacturer's instructions. These purified libraries were prepared for sequencing on the Illumina MiSeq platform, with attachment of theIllumina barcoded index primer (to allow subsequent identification of the library of origin of sequences, as all libraries were pooled onto a single MiSeq run), universal Illumina primer and Illumina adaptor using the NEBNext Ultra DNA Library Prep Kit, and then once again quantified before submission to the sequencing facility.

The samples were paired-end sequenced at the Oxford Zoology Department sequencing facility, using an Illumina MiSeq (Illumina, San Diego, CA, USA) with v2 600

cycle kit and a 10% Phi-X spike, producing an approximate 300bp sequence for each amplicon end.

The MiSeq output is already demultiplexed according to the Illumina library barcodes and the resulting files of sequence reads for each library were demultiplexed according to the in-house barcodes using custom Python scripts, with those sequence reads lacking either of the two barcodes discarded, as well as those with barcodes of insufficient quality (bases with Phred score < 20 i.e. < 99% confidence the base call is correct according to the accompanying Mi-Seq generated qual string, were removed and barcodes not 7bp long discarded). The resulting fastq files were further processed using the QIIME pipeline (Caporaso *et al*., 2010). Paired end reads are assembled using the script 'join_paired_ends.py' using the fastq-join method (Aronesty, 2011) and then are assigned to their samples of origin according to a barcode mapping file and quality filtered using the QIIME script 'multiple_split_libraries.py'. Base calls with a Phred score < 20 are removed, with sequence reads truncated after six consecutive low-quality base calls, or homopolymers of greater than six nucleotides in length. Sequence reads with an average Phred score < 25 are discarded, as are sequence reads less than 200 base pairs in length. No chimera detection step is included, as subsequently closed reference OTU picking is done against a chimera-checked reference database (DeSantis *et al*., 2016).

Resulting sequence reads were clustered into OTUs, aligned and taxonomy assigned, and an OTU table produced, using the 'pick_open_reference_otus.py' QIIME script with the UCLUST algorithm (Edgar, 2010). Closed reference OTU picking is performed with reference to the GreenGenes database (http://qiime.org/home_static/dataFiles.htmlftp://greengenes.microbio.me/greengenes_release/gg_13_5/gg_13_8_otus.tar.gz), creating OTUs consisting of clusters of sequence reads that match sequences in the database by 97% similarity. Residual sequences that do not match the database and so are not assigned to OTUs in this previous step are then

clustered Open reference OTU picking, and a cluster 'centroid' used as the reference sequence for each OTU, and OTUs with fewer than two sequence reads (i.e. singletons) discarded. The two groups of resulting OTUs are then combined and aligned and a phylogenetic tree built, with taxonomy assigned to OTUs to as fine a taxonomic level as possible using the GreenGenes database and from this the OTU table is produced.

A single pseudo-OTU 'Unassigned' (on average 3% of reads per sample) was subsequently discarded as this did not represent a single OTU resulting from clustering. There was no evidence of 'Unassigned' OTUs being retained for further analysis in studies using a similar approach in semen (Javurek *et al*., 2016), cloacal (Kreisinger *et al*., 2015) or avian gut samples (Videnska *et al*., 2013). All subsequent analyses were carried out using R software version 3.3.2 (R Core Team, 2016).

Identifying and Minimising Contamination

A stepwise approach was taken to identify the sources of contamination and the identity of contaminants at the PCR Amplification stage, and to produce a protocol for sample processing and analysis that minimises contamination, allowing us to deal with residual contaminants. Table 1. outlines the sequential steps of this methodology and the potential sources of contamination associated with them.

Table 1 *The five main stages in the methodology for male junglefowl seminal microbiota sample collection and processing, and the contamination risks associated with each stage.*

Sample Collection	Contamination with junglefowl (i.e. non-seminal) bacteria from the focal bird or birds in the near vicinity. Cross contamination between samples via equipment as well as environmental bacteria on the equipment and surfaces used. Contamination from the skin, hair and clothes of the experimenters.
Sample Processing	Cross contamination between samples via equipment as well as environmental bacteria on the equipment and surfaces used. Contamination with bacteria from the skin, hair and clothes of the experimenters.
DNA Extraction	Cross contamination between samples via equipment as well as environmental bacteria on the equipment and surfaces used, including that from other laboratory users' experiments. Contaminants contained in the reagents used. Contamination with bacteria from the skin, hair and clothes of the experimenters. Contamination with the DNA product from previous extractions. Contamination with the post-PCR amplification product from previous experiments.
PCR Preparation and Amplification	Cross contamination between samples via equipment as well as environmental bacteria on the equipment and surfaces used, including that from other laboratory users' experiments. Contaminants contained in the reagents used. Contamination with bacteria from the skin, hair and clothes of the experimenters. Contamination with the DNA product from previous extractions. Contamination with the post-PCR amplification product from previous experiments.
Post-PCR Processing	Minimal risk as post-PCR the amount of DNA is very large compared to any contamination, and no further step will amplify the contamination. However, there is still a risk of cross-contamination between samples.

Contamination was first identified in PCR negative controls, following optimization of non-barcoded primers 347F and 803R with semen samples using the HotStarTaq Master Mix Kit (QIAGEN). A rigorous stepwise process was consequently implemented with the aim of minimizing, identifying and managing contamination at the

PCR stage. Reagents were switched to higher-grade molecular biology reagents, with fresh batches of primers and dNTPs being ordered and aliquoted. Microtubes and PCR Plates (Sarstedt) of a quality certified for use in molecular biology were used and a new cleaning protocol was adopted, with all surfaces and equipment cleaned with DNAZap PCR DNA Degradation Solutions and exposed to UV radiation for one hour before experimental use. DNA Extraction, PCR preparation and post-PCR processing were all spatially separated, different sets of equipment and laboratory coats were used for each stage, and for DNA and non-DNA use. To eliminate the possibility of cross contamination from samples and *E. coli* DNA positives within the hood and to check the consistency of contamination, ten negative water controls were prepared and completely sealed before DNA positives were added.

The Sanger Sequencing output was used to assess the success of the measures put in place to reduce contamination and thus deduce the likely source and extent of the remaining contamination. A comparison of semen samples with contaminated negatives from the same PCR run and a comparison of contaminated negatives from the first identification of contamination with contaminated negatives from the end of the stepwise process described above was therefore carried out.

Criteria were set using Illumina MiSeq sequencing output, which defined OTUs as contaminants. To do this, the prevalence and average relative abundance of OTUs in 36 PCR negative water controls were calculated, and those OTUs that appeared in 90% of control samples and had an average relative abundance equal to or greater than 0.1% in control samples in which they appeared were classed as contaminants. We distinguished between 'whole dataset' or 'per miseq' contaminants. 'Whole dataset' contaminants were identified from all PCR control samples collected and processed in the course of all sample analysis for this work, whilst 'per miseq' were OTUs which met the contaminant criteria within all the control samples for a MiSeq sequencing run (between .. and … negative

water controls per run). This distinction was made to separate those OTUs that were potentially contaminating across all sequencing runs and needed to be dealt with in all samples, and those that were only present as potential contaminant within a single run and only needed to be dealt with in the samples of the MiSeq run which they contaminated.

'Whole dataset' contaminants are likely to have originated from laboratory or reagent sources, which were constant across all runs, and so were removed from the dataset prior to further analysis. Those 'per miseq' OTUs, which met these criteria just for one MiSeq run are likely to represent potential cross-contamination across different semen samples within that run. These contaminants are more difficult to deal with as these OTUs are likely to be genuine components of biological samples and thus it would not be appropriate to simply remove them from all samples of that run. A parsimonious approach is removing a baseline relative abundance as defined by the average relative abundance in the negative controls. A relative abundance that is equal to- or lower than this threshold was deemed likely to represent contamination, whilst a level above the threshold is likely to reflect a genuine part of the sample. This is not a perfect solution as the baseline relative abundance of an OTU in controls may be much higher than the relative abundance due to contamination in a biological sample, due to there being fewer bacteria present in the control compared with the sample. This would mean that removing to a baseline as defined by relative abundance in controls could result in a loss of naturally occurring abundance of those OTUs from samples. However, this is preferred to complete loss of biologically relevant OTUs, or treating contamination as a genuine component of the reproductive microbiota.

Dealing with Variation in Sequencing Depth

Samples differed in the number of reads generated for all the sequences (sequencing depth). Because sequencing depth represents sampling effort, we

investigated whether there was an effect of variation in sequencing depth on the structure and composition of the community produced, and whether samples of differing sequencing depths were comparable. To do so, from all 442 sequenced semen samples we selected the ten samples with the highest number of reads ('High'), the ten samples centred around the average per sample number of reads ('Medium'), the ten with the lowest number of reads ('Low'), and the ten samples with the lowest number of reads greater than 1000 ('Low1000', as the 'Low' samples had such low read numbers as to be intractable for analysis and comparison). The relationship between average within sample OTU relative abundance and frequency (number of samples in which an OTU was detected) was plotted for each sequencing depth category to check if there were consistent differences in community 'shape' between the samples in each category. Species accumulation curves were plotted for the forty samples (i.e. 'High', 'Medium', 'Low', and 'Low1000'), using the 'rarefaction' function in the R package 'vegan', to investigate whether samples with different sequencing depths follow the same patterns of species accumulation. To ensure that samples had comparable levels of coverage at which the actual community is well represented, coverage was calculated for all semen samples using the R package 'entropart' according to Chao, Good's and Zuang-Huang coverage indexes, and samples not achieving 95% coverage in all three measures were discarded from further analysis. All samples were then rarefied to the same number of reads as that of the lowest sample retained as in other studies of avian microbiota using the same data type (e.g. Kreisinger *et al*., 2015; Videnska *et al*., 2013).

Assessing Potential Laboratory Bias: Barcoding steps and repeatability

We investigated the effect of the two barcoded primer steps (i.e. Illumina Library Barcode and 347F PCR Amplification Barcode) upon bacterial composition. DNA extracted from three semen samples were each amplified using five different in-house designed 347F Primer Amplification barcodes, and the PCR product of these fifteen amplification reactions was split into three separate libraries, each of which was then

prepared for sequencing using a different Illumina barcode**s** to give fifteen repeats of each sample (Fig. 2). This approach enabled us to assess differences between samples due to methodological effects including not only primer bias, but also, in the case of the 347F PCR Amplification Barcode, any factors that might lead to differential PCR amplification between reactions.

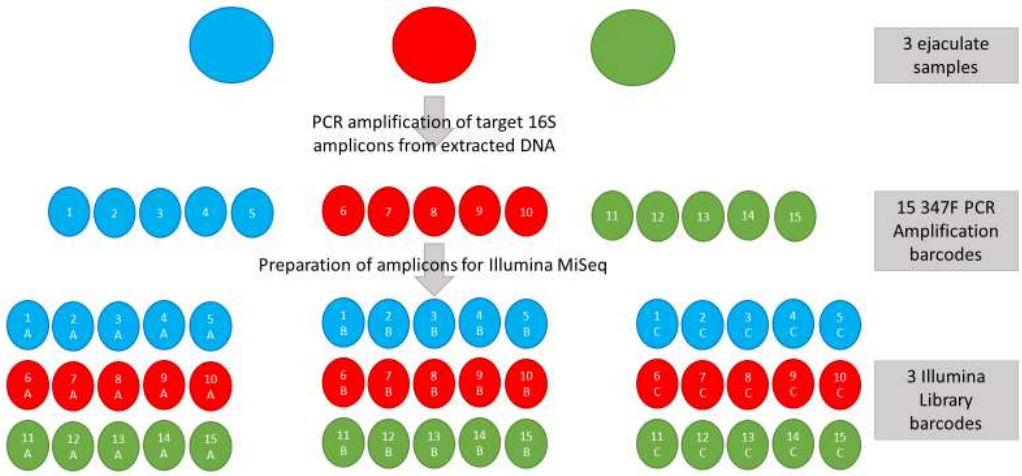


Figure 2 *Diagrammatic representation of the experimental design used to investigate the effect 347F PCR Amplification and Illumina barcodes. Colour represents ejaculate sample, numbers the 347F Primer Amplification barcode and letters the Illumina barcode.*

Three community diversity metrics commonly used to characterise microbiota (Species Richness (e.g. Bolnick *et al*., 2014), Shannon Diversity and Simpson Diversity (e.g. White *et al*., 2011) were calculated from the Illumina output for each repeat, and the effect of the two barcoding levels on these metrics was visualized. The nature of any differences was also visualized through plots of OTU relative abundance and OTU frequency for each sample. To assess the amount of variation due to the two barcode levels, and the extent to which this may have obscured any differentiation of samples, linear mixed models (LMMs) with and without the two barcode levels as random effects were compared by ANOVA using the R package 'lme', with metrics transformed as appropriate to ensure they met the assumptions of the linear mixed models. Species Richness and Shannon

Diversity were ln(natural logarithm)-transformed, and Simpson Diversity was transformed using the ArcSine Square Root transformation.

In addition, we assessed within-sample repeatability. To explore how much variation between samples is due to overall post-collection methodological factors rather than biologically meaningful factors, we measured sample repeatability across aliquots of the same semen sample. These aliquots were created at the time of sample collection, where the sample was split into two or three 10uL whole semen aliquots before storage. Any differences between these aliquots was likely be due to artefacts of the post-collection laboratory process rather than any property of the sample itself. The Illumina sequencing output of nine semen samples from nine different birds, each with two or three aliquot repeats was used. Repeatability measures were calculated for the three community diversity metrics described previously according to two methods, one LMM-based and the other ANOVA-based (Nakagawa & Schielzeth, 2010). The LMM-based used the following model: Community Metric ~ 1 + (1|Sample), and calculated repeatability as:

$$R = \frac{VA}{VA+VE} \qquad \textit{Equation 1.}$$

where VA is the between-group variance, and VE the within-group variance. The ANOVA-based approach used the following model: Community Metric ~ Sample, and measured repeatability as:

$$R = \frac{MSA-MSW}{MSA+MSW(n_0-1)} \qquad \text{Equation 2.}$$

where $n_0 = \frac{1}{k-1}(N - \frac{\sum {n_i}^2}{N})$ (see Table 2). *Equation 3.*

Table 2. *Symbols for the variables in Equations 2 and 3*

Symbol	Variable
K	Number of individuals
N	Total sample size
n_i	Sample size of individual *i*
MSA	Mean between-individual sum of squares
MSW	Mean within-individual sum of squares (i.e. residual)
R	Repeatability

n_0	Sample size of an individual, if all were equal

We also assessed repeatability by ranking OTUs by relative abundance in each sample, and comparing the Spearman's Correlations of these ranks between aliquots from the same sample to those from different samples. Where a semen sample had three aliquot repeats, two of the three were randomly selected for the analysis. The aliquots were then randomly paired (8/9 of which were pairs from different samples) and the Pearson's correlation coefficients of original pairs were compared to those of random pairs with an un-paired student's T-test. To ensure that the dataset used was not zero-inflated, and differences between samples exacerbated by low-abundance OTUs, we performed the analysis using two subsets of OTUs within each sample, one subset containing only OTUs occurring at an average relative abundance of 1%, and the other only OTUs occurring at 0.1%. Additionally, using the same random pairs of aliquots and the same OTU subsets, Bray-Curtis dissimilarities were calculated for original and random pairs and then a difference between the two groups investigated using an un-paired student's T-test.

Characterisation of the Seminal Microbial Community and Basic Patterns of Variation

Overall, 171 whole semen samples (145 through AM, 26 through CC) were collected and successfully sequenced from 65 males throughout the 2015 and 2013 breeding seasons, and were used to characterise the red junglefowl seminal microbiota. The phyla that contribute most to the seminal microbiota were identified, as well as those with the highest average relative abundance per sample. To look for patterns of variation in microbiota composition across individuals and identify OTUs potentially driving this variation, a Principal Component Analysis (PCA) was performed on the average relative abundance of OTUs in each bird. To eliminate the effect of influential outliers in the form

of infrequently observed OTUs, the same analysis was also performed using a subset of 39 OTUs, which were detected in at least 10% of birds and at an average relative abundance equal to, or greater than 1%. These thresholds were deemed appropriate as they gave a distribution free of influential outliers when the first PCA axis was plotted against the second.

Controlled Copulations and Abdominal Massage

All first collected whole semen samples from the 2015 season were used to investigate the differences in community diversity and the relative abundances of OTUs in seminal microbiota resulting from the two collection methods, comprising 26 controlled copulation samples and 73 abdominal massage samples from 38 birds. The number of OTUs shared between the two types was calculated as well as the number unique to each.

The three community diversity metrics (species richness, Shannon diversity and Simpson diversity) were calculated for each sample and transformed as appropriate to meet the assumptions of statistical tests. To establish whether the collection method (i.e. abdominal massage *versus* controlled copulation) was a source of variation an ANOVA test was performed in R as described above. To allow for the uneven numbers of samples for the two collection methods, a subset of the 26 controlled copulation samples and 26 randomly selected abdominal massage samples was used. Again, for this subset, the number of OTUs shared between the two types was calculated as well as the number unique to each, and variation in diversity metrics was investigated using ANOVA. Further, to investigate differences between samples collected by the two methods from the same male a third subset was used comprising six pairs of samples where samples were collected by both methods from the same male within three days of each other. The number of OTUs shared between the two types was calculated, as well as the number unique to each, and variation in diversity metrics was investigated using ANOVA.

Additionally, for this third subset, the Pearson's Correlation of community diversity metrics between the two sample types from each male was also calculated, as well as the Pearson's Correlation of the rank of OTUs by relative abundance in each sample. As previously, two subsets of OTUs within each sample were used to calculate rank, one subset containing only OTUs selected to occurring at an average relative abundance of 1%, and the other only OTUs occurring at 0.1%.

Whole semen, seminal fluid and sperm samples

Thirty-four semen samples collected in the 2015 season had aliquots of whole semen (EJA), seminal fluid (SF) and sperm pellet (SP) as described in methods. These included both controlled copulation and abdominal massage samples, from 22 birds and includes second, third and fourth as well as first semen samples of a bird. The number of OTUs shared between all three types was calculated as well as the number shared between the three combinations (i.e. EJA *vs.* SF, EJA *vs.* SP, and SF *vs.* SP) and those unique to each type. The three community diversity metrics (species richness, Shannon diversity and Simpson diversity) were calculated for each sample and transformed as appropriate to meet the assumptions of statistical tests. To establish whether sample type was a source of variation in these metrics an ANOVA test was performed as previously. Differences in these measures between the three groups was further investigated using Welch Two Sample T-tests to compare each type with both the others (three tests in total). Samples, for which we were unable to collect all three sample types (EJA, SF and SP) were removed from the dataset, and Pearson's Correlations between the types for each community metric were performed using the remaining 31 semen samples.

OTUs were ranked by relative abundance in each sample and the Pearson's Correlation of these ranks for each combination of for each semen sample was calculated i.e. between EJA and SF, EJA and SP and SF and SP for each semen sample. As previously, two subsets of OTUs within each sample, one subset containing only OTUs selected to

occurring at an average relative abundance of 1%, and the other only OTUs occurring at 0.1%.

Comparing the seminal microbiota to the faecal microbiota

To investigate how much the seminal microbiota reflects the faecal microbiota, in terms of both community diversity and OTU relative abundance, matching (i.e. from the same bird within fifteen minutes of each other) faecal samples were collected and analysed for 26 semen samples, which represented the first sample from a particular bird in that day, each from a different bird. Only first collected semen samples were used as the extent and pattern of variation in microbiota structure and composition between subsequent samples on the same day is unknown. The number of OTUs shared between the two microbiotas within a male was calculated, as well as the number unique to each microbiota. The three community diversity metrics (species richness, Shannon diversity and Simpson diversity) were calculated for each sample and transformed as appropriate to meet the assumptions of statistical tests. To establish whether sample type (faecal or semen) was a source of variation in these community diversity metrics an ANOVA test was performed on a linear model with each metric as the response variable and sample type as a fixed factor. Differences in these measures between the two sample types was further investigated using a Paired T-test and the Pearson's correlations of community diversity metrics between the two sample types from each male. OTUs were ranked by relative abundance in each sample and the Pearson's Correlation of these ranks between each faecal-ejaculate pair was calculated. As previously, two subsets of OTUs within each sample, one subset containing only OTUs selected to occurring at an average relative abundance of 1%, and the other only OTUs occurring at 0.1%.

RESULTS

1. Establishing and reducing methodological bias

Contamination

Contamination was first identified following optimization of non-barcoded primers (Fig. 3). After the implementation of a more rigorous protocol and the switching to higher molecular grade reagents, there was still a signal in all the negative controls (Fig. 4). The consistency of contamination in all controls and across PCRs performed on different days would suggest a reagent source rather than environmental contamination. Additionally, the presence of contamination in negatives following PCR preparation, which did not use semen samples, and when positive controls were added after negative control reactions were sealed is inconsistent with the possibility of cross contamination from samples and positives handled during the same PCR preparation.

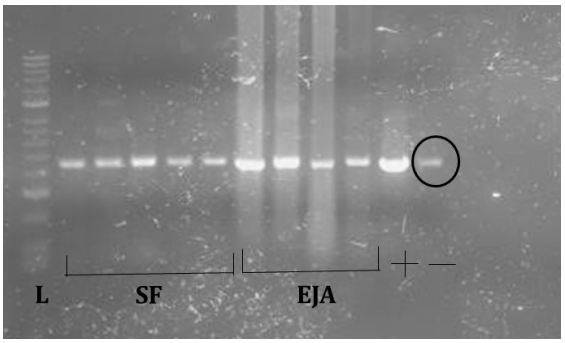


Figure 3 *UV photograph of gel following electrophoresis of PCR amplification products from five seminal fluid samples (SF), five whole semen samples (EJA) a Neisseria positive control (+) and a water negative control (- and circled) following PCR amplification with the HotStarTaq Master Mix Kit. Reference of the position of the sample and negative bands in relation to the 50kb ladder (L) bands indicate that there are fragments of DNA in these samples the same size as the target fragment. The intensity of the band is a rough indicator of DNA concentration.*

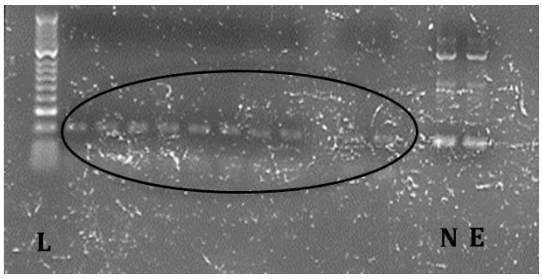


Figure 4 *UV photograph of gel following electrophoresis of the PCR amplification products of ten negative water controls (- and circled), a Neisseria positive control (N+) and an E.Coli positive control (E+). Reference to the 50kb ladder (L) indicates that DNA in the sample and negative bands matches the target fragment size. The intensity of the band is a rough indicator of DNA concentration.*

Identifying Contaminant Species

Considering results of Sanger sequencing (Fig. 5) of negative controls taken before (Negatives 1-4), during (Negatives 5-9) and after (Negatives 10-13) the implementation of the strict contamination minimisation protocol, there was a shift in species found in negative controls, and where there are negative and biological samples from the same PCRs (Negatives 5-9), there were species found in samples additional to those in negative controls, suggestive of species originating from the ejaculate sample and not from contamination. Across the later negative controls (10-13) two bacterial groups were consistently identified, *Halomonas sp.* and *Shewanella sp*. These two residual contaminant groups are species not commonly reported in the human or chicken microbiota as far as we are aware, and are not commonly used laboratory strains. *Shewanella* species are found in, often extreme, marine environments (Dikow, 2011) and *Halomonas* in highly saline environments (e.g. Vreeland *et al*., 1980). We therefore concluded that they are unlikely to be part of the reproductive microbiota and most likely to have originated from the reagents supplied. This means that these species can be removed from the analysis of sequences, as contamination is consistent across all samples.

Of the species found in the semen samples, additional to those in the contaminated negatives (i.e. *Halomonas sp*. and *Shewanella sp*.), many are likely to be a physiological part of the junglefowl microbiota. *Lactobacillus* (Samples 1, 2 and 5, Fig. 5) and *Bifidobacterium* (Sample 1 and 2, Fig. 5) are found in the domestic chicken faecal microbiota (Videnska et al., 2013), *Corynebacterium* (Sample 5) and *Actinomyces* (Sample 5, Fig. 5) have all been reported in human ejaculates (Liu *et al*., 2014) along with *Streptococcus* (Sample 2 and 5, Fig. 5) and *Staphylococcus* (Sample 2, Fig. 5) although these latter two were also found in Negatives 1 and Negative 9, respectively. We therefore concluded that although some level of bacterial DNA persists in the negative water

controls, only two of a collective 48 colonies from the latest four negative controls were organisms that are likely to be members of the junglefowl microbiota, and therefore the remaining bacteria represents contamination via laboratory reagents or equipment that can be eliminated from the analysis as described below. These results suggest that as long as negative water controls are systematically collected throughout the process to allow differentiation between contamination and the target microbiota, sequences can be isolated from biological samples. Signal in sample of DNA following gel electrophoresis therefore is highly likely to represent relevant sample bacterial DNA rather than contamination.

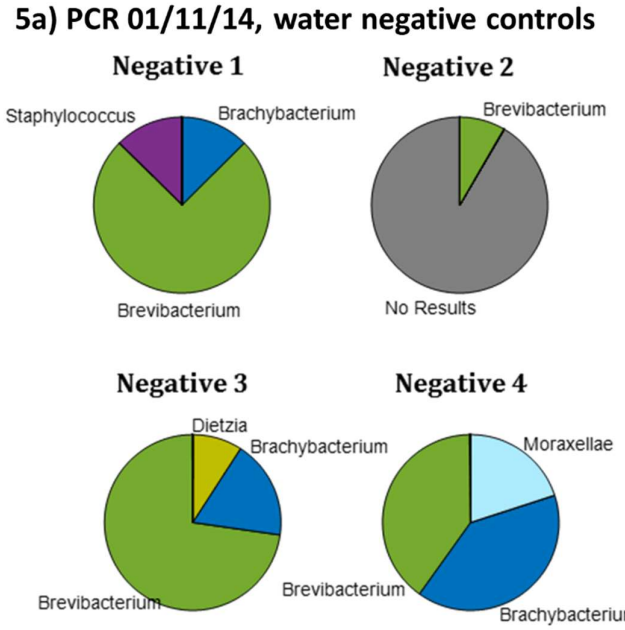


Figure 5a *Pie charts displaying the results of sequencing of water control negatives. Segment size represents the proportions of the 12 colony sequences per sample that were identified as belonging to the phylotypes shown. 'No results' = the sequences produced from that colony did not have sufficient signal identify the organism. 'Uncultured bacterium' in Sample 4 encompassed 'Uncultured rumen bacterium' and 'Uncultured faecal bacterium', and as such is likely to reflect a component of junglefowl microbiota than a PCR contamination source.*

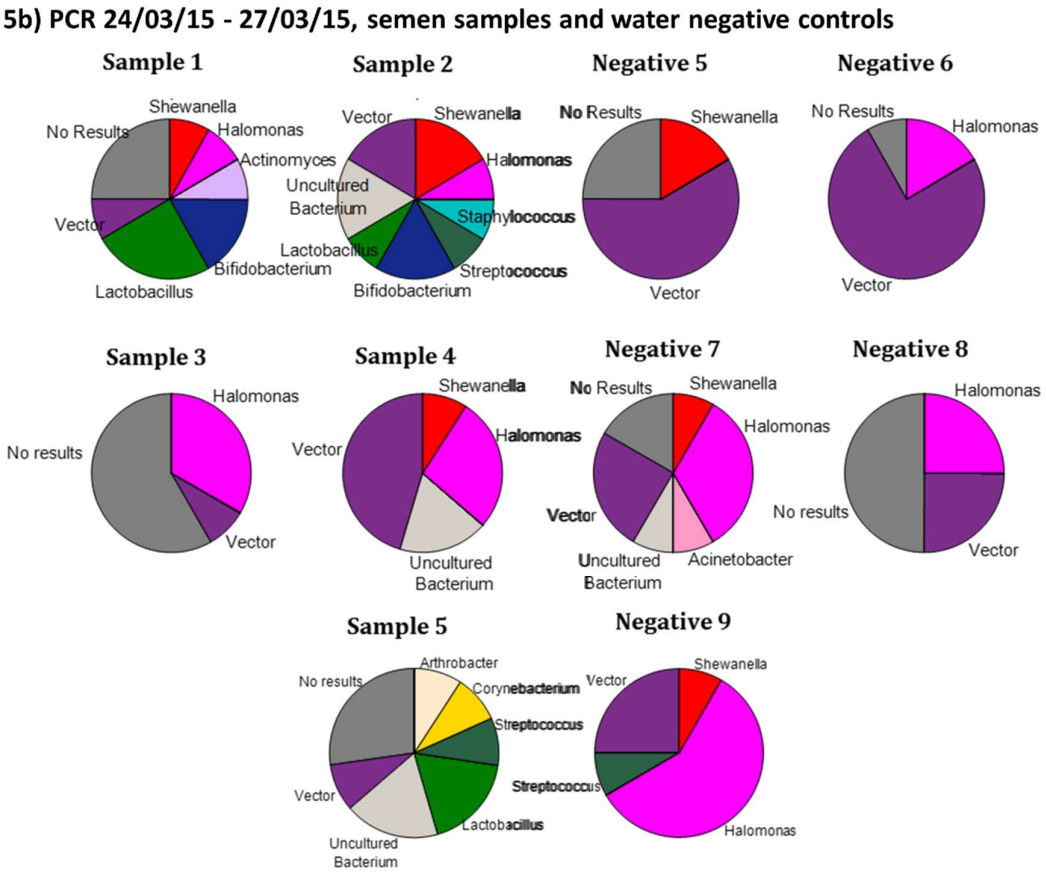


Figure 5b. *Pie charts displaying the results of sequencing of semen samples (Samples 1-5) and water control negatives (Negatives 5-8). Segment size represents the proportions of the 12 colony sequences per sample that were identified as belonging to the phylotypes shown. 'No results' = the sequences produced from that colony did not have sufficient signal identify the organism. 'Vector' = the sequence produced was of the vector used to insert amplified DNA into the E. coli for cloning. Biological samples 1-5 contain several OTUs which are not found in the matching water negative controls (Negatives 5-8) and which are therefore likely to have originated from the red junglefowl semen. Samples 3 and 4 look quite similar to negative controls amplified in the same run (Negatives 7 and 8).*

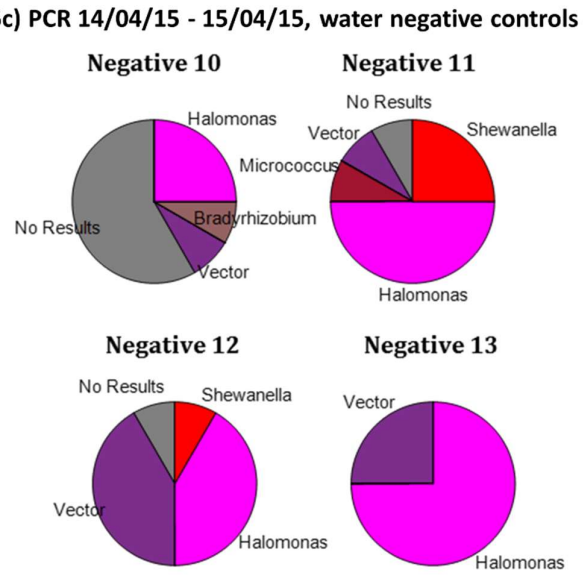


Figure 5c *Pie charts displaying the results of sequencing of water control negatives. Segment size represents the proportions of the 12 colony sequences per sample that were identified as belonging to the phylotypes shown. 'No results' = the sequences produced from that colony did not have sufficient signal identify the organism. 'Vector' = the sequence produced was of the vector used to insert amplified DNA into the E. coli for cloning.*

Dealing with Contaminants

Having identified the nature and the source of contamination, and having implemented a contamination reduction protocol and verified that this likely represents contamination from reagents and does not prevent detection of genuine microbiota components, this protocol was used to process samples across eight Illumina MiSeq runs, i.e. each run was a collection of samples into a single library and sequenced on the same lane of the Illumina MiSeq. Using negative water controls taken and sequenced during these runs, five OTUs (three *Halomonas,* one *Shewanella*, and one *Enterobacteria*) met the criteria of whole dataset contaminants i.e. having a prevalence across all controls greater than 95% and a mean relative abundance greater than 1% (Table 3). In addition, three OTUs were classified as 'per MiSeq' contamination in MiSeq1, three in MiSeq2 and one in Miseq4 (Table 4).

Table 3 *Whole Dataset Contaminants. Frequency indicates the proportion of negative control samples that each OTU was detected in. (1) and (2) simply differentiate the two species of the Halomonas genus that cannot be more finely assigned taxonomically.*

OTUID	Frequency	Average Relative Abundance
Shewanella algae	1	0.30
Enterobacteriaceae sp.	0.98	0.03
Halomonadaceae sp.	0.98	0.04
Halomonas sp (1)	1	0.25
Halomonas sp. (2)	0.98	0.25

Table 4 *Per-MiSeq Contaminants. MiSeq indicates the Illumina sequencing run in which each OTU was defined as a contaminant. Frequency indicates the proportion of negative control samples in that run in which each OTU was detected.*

OTUID	MiSeq	Frequency	Average Relative Abundance
Actinomycetaceae sp.	2	1	0.02
Actinomyces sp.	2	1	0.01
Corynebacterium sp.	2	1	0.02
Planococcaceae sp.	1	1	0.01
Clostridiaceae sp.	4	1	0.02
Neisseria sp.	1	1	0.01
Neisseria cinerea	1	1	0.01

Variation in sequence number

Sequencing on the Ilumina MiSeq platform produced 162,081,446 sequence reads in total from seven MiSeq runs. This includes 925 samples of ejaculate, faecal, cloacal, cultured and negative control samples. Following demultiplexing, according to in-house barcode, approximately 60% of sequence reads were lost, and a further 4% of the remainder were excluded during quality filtering and clustering in QIIME. Further sequence reads (2% of those from the QIIME output) were removed as they belonged to the 'Unassigned' OTU, and approximately 40% of the reads remaining following the removal of the 'Unassigned' OTU were removed as either whole-dataset or per-run contaminants.

From 442 pre-rarefaction ejaculate samples there were 11,213,850 sequence reads in total, with an average of 25,371 (SD = +/- 30,701) per sample.

Considering plots of OTU relative abundance and OTU frequency there was not an immediately obvious underlying difference in the community structure of samples due to sequence number, across the ten 'High', 'Medium', 'Low1000' and 'Low' samples (Fig. 6). The 'Low' category is not shown here as sequence numbers were intractably low and it was difficult to properly interpret the patterns.

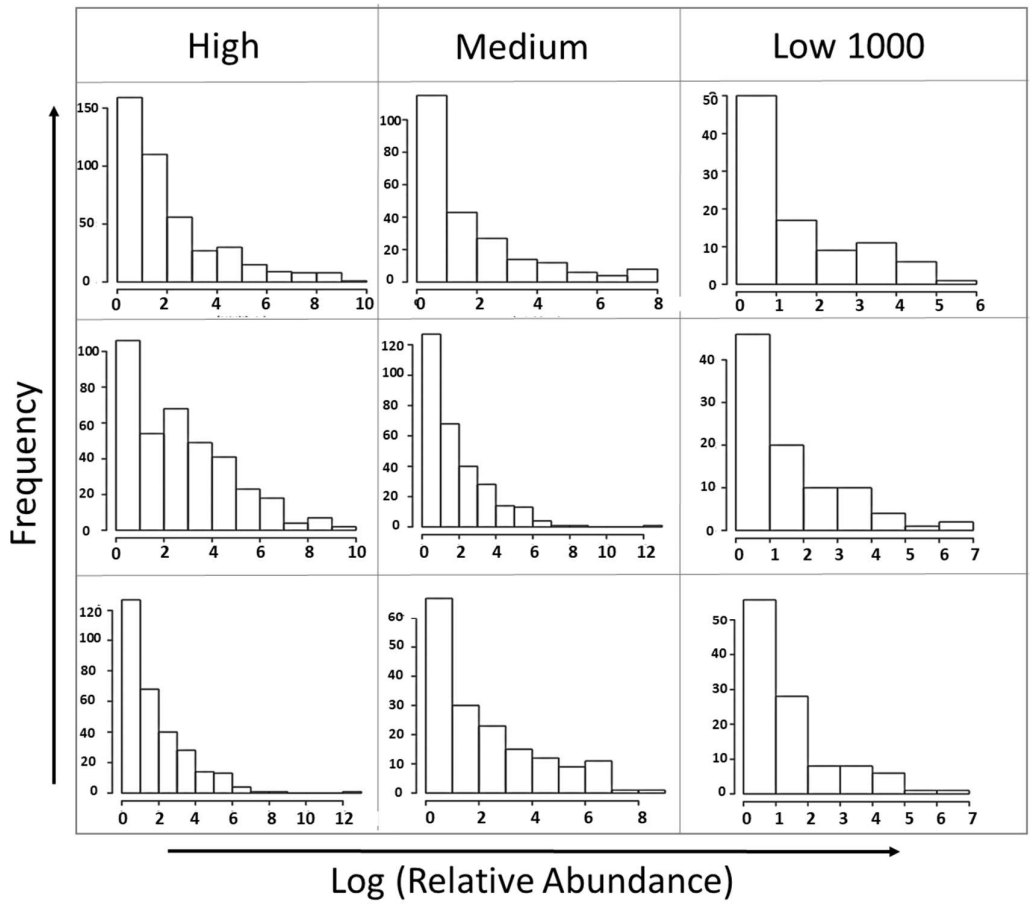


Figure 6 *Plots of OTU Frequency against the natural log of OTU relative abundance for a random selection of three from the ten samples in each depth category, showing that community 'shape' does not differ across depth categories. The Low category is not shown as sequence numbers are intractably low.*

Species OTU accumulation curves, showed a decelerating asymptotic pattern, except in the case of very low ('Low') samples, with most slowing in their OTU accumulation by a sequencing depth of 1000 sequences, although none reached complete saturation (Fig. 7). However, there was a difference across the read number categories in the eventual number of OTUs following rarefaction at this depth (Fig. 8).

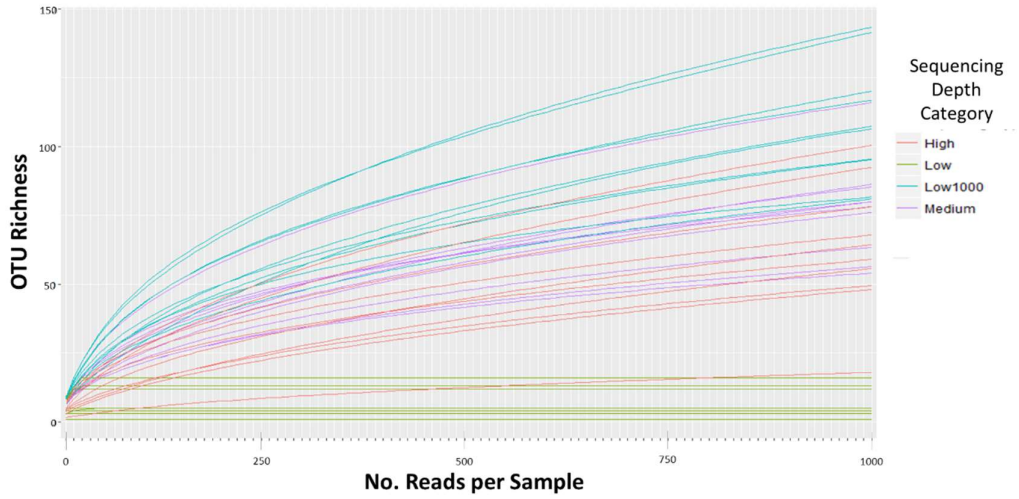


Figure 7 *Sample accumulation curve for all 40 selected samples up to 1000 sequence reads, showing that at 1000 reads per sample, species accumulation has slowed but continues for High, Medium and Low 1000 samples. NB those with low sequence numbers and a horizontal curve have not reached saturation, but are 'flat-lining' as rarefaction has reached their actual number of sample reads; this function does not predict numbers of OTUs at higher read numbers.*

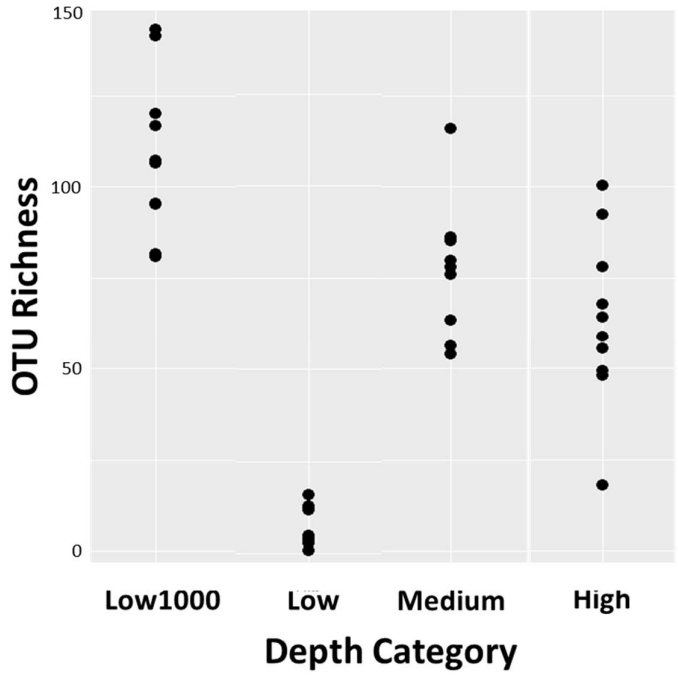


Figure 8 *OTU richness at rarefaction to 1000 reads per sample across different sequencing depth categories.*

An analysis of variance revealed that read depth category did indeed explain OTU richness at rarefaction to 1000 sequences ($F_{(3,36)} = 51.5$, $p < 0.001$) and the effect remained even when the 'Low' samples were removed ($F_{(2,27)} = 12.1$, $p < 0.001$). This was further explored in relation to potential laboratory bias, to establish its impact on conclusions

from analyses investigating variation across reproductive microbiota. Despite variation in sequencing depth, and associated variation in OTU richness at equivalent rarefaction, the majority of samples (396 of 442) achieved 95% coverage in all three coverage indices (Fig. 9). Those samples that achieved less than 95% coverage in any of the three coverage indices (Good's, Chao or Zuang-Huang) were discarded from subsequent analyses. All samples retained were rarefied to 523 sequences, the lowest read number for a retained sample. Rarefaction, rather than relative abundance, was used to achieve equal sample sequence numbers. 'This approach was deemed more robust because relative abundance will still retain 'rare' OTUs, i.e. those defined as occurring at a sufficiently low per sample absolute abundance to be lost when samples are rarefied. Retention of these rare OTUs, even at miniscule values of relative abundance, may affect diversity measures and would increase species richness compared to if the sample were rarefied, potentially confounding differences across samples of different read number (and similarities between samples of similar read numbers).

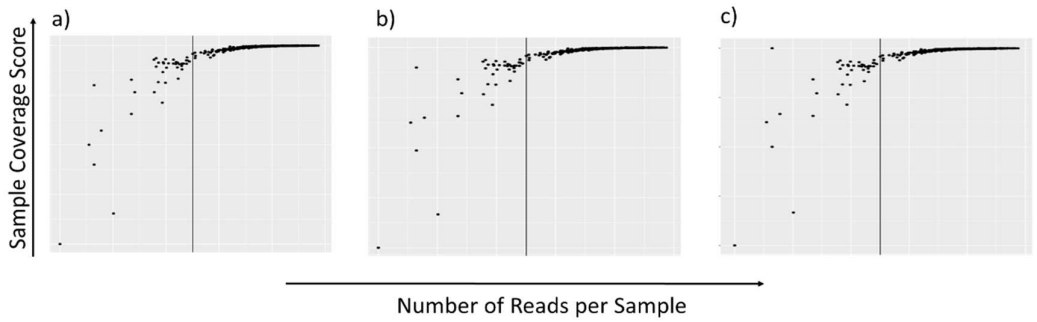


Figure 9 *Plots of three coverage indexes a) Good's coverage b) Chao coverage c) Zuang Huang coverage against the natural log of total read number for all samples in the dataset, demonstrating that the majority of samples achieve a high level of coverage. The vertical line identifies the smallest number of reads (523) in samples achieving 95% coverage in all three measures.*

The Effect of Illumina Library Barcode and In-house PCR Amplification Barcode

We investigated the effect of two different levels of barcoding on community diversity metric, to clarify whether variation between samples is an artefact of barcode bias rather than due to differences originating in the sample composition itself. Plots of the three community metrics (species richness, Shannon diversity, and Simpson diversity) indicate that variation between repeats of the same sample was generally smaller than variation across samples (Fig. 10). 347F barcode did not appear to have any consistent effect on any of the three community metrics. Illumina barcode on the other hand, whilst not preventing discrimination of samples, did appear to have consistent effects on all metrics across barcode and sample.

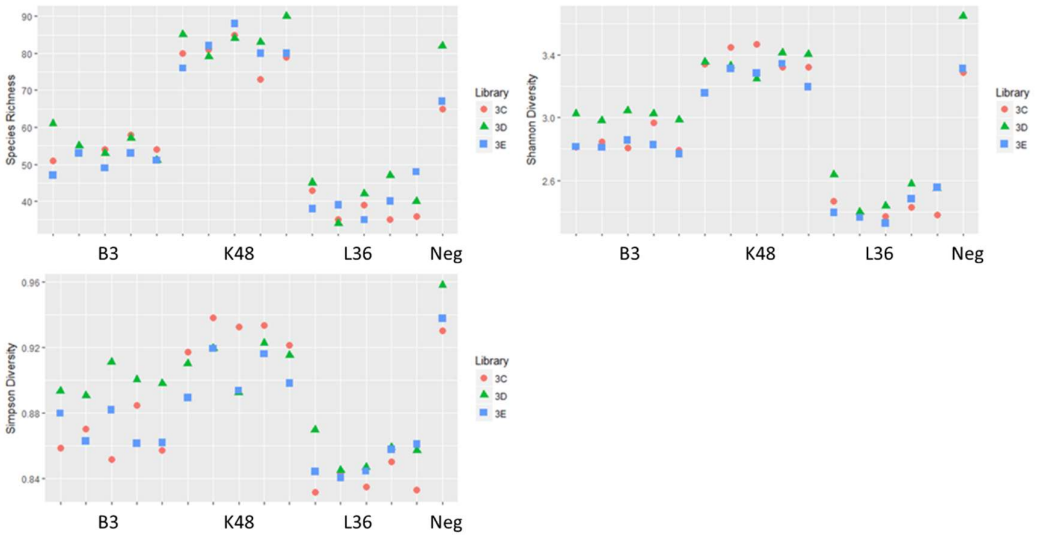


Figure 10 *Plots of three community diversity metrics for all 347F barcode and Illumina library barcode replicates of each of four samples, showing that there is variation in community metrics with the two barcodes, but that samples are still distinguishable from each other. These are grouped vertically by 347F barcode within original sample indicated on the x-axis (five per sample except one for PCR negative) and colour and shape coded by Illumina library barcode.*

There were many OTUs that did not appear in all fifteen repeats of a sample. Looking at the relationship between within-sample OTU frequency and mean relative abundance however, it is evident that those OTUs not present in every repeat occur at low relative abundances when they did occur (Fig. 11). This suggests that between-repeat differences in OTU presence are the result of OTU rarity or a ubiquitous inability to be detected, rather than bias of amplification by 347F PCR Amplification or sequencing bias by Illumina barcodes.

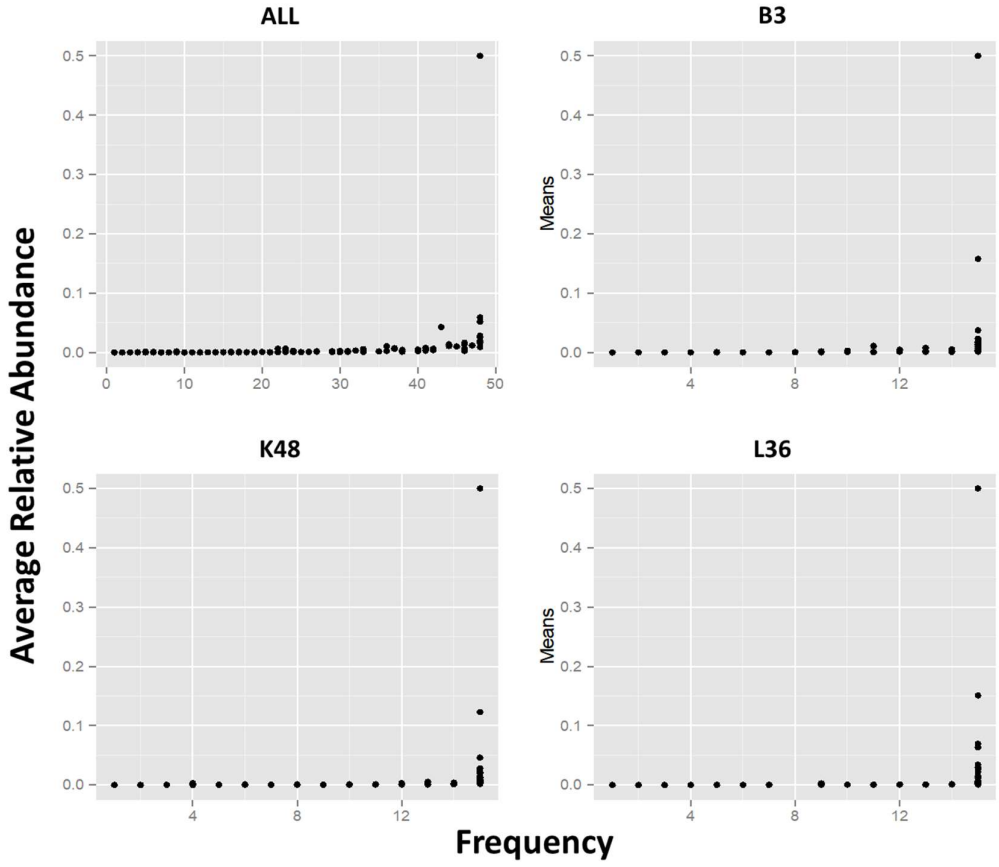


Figure 11 *Plots of OTU frequency against mean relative abundance in the 15 replicates for each of the three biological samples (B3, K48, L36) and for all 45 grouped (ALL) demonstrating that OTUs not found in every sample are usually those that occur at very low relative abundance.*

LMMs confirmed these visual trends by discriminating between samples from different birds without the inclusion of 347F PCR Amplification or Illumina barcode (species richness ($F_{(3,44)}$ = 232 , $p < 0.001$) Shannon diversity($F_{(3,44)}$ = 218 , $p < 0.001$) Simpson diversity ($F_{(3,44)}$ = 64.1, $p < 0.001$), and this was still the case when the two were included as random effects (SimpDiv (ChiSq = 66.9, df = 3, $p < 0.001$), ShanDiv (ChiSq = 68.4, df = 3, $p < 0.001$), SimpDiv (ChiSq = 49.1, df = 3, $p < 0.001$). Thus, we can conclude that whilst barcodes appeared to explain some variation in community metrics between sample repeats, they did not prevent the detection of biologically relevant between sample differences originating previous to the use of the barcodes, and are confident that results based on variation in these metrics between samples are not simply due to barcode amplification bias. Given the previous results displaying an effect of sequencing depth category on OTU richness in rarefied samples, it was also explored whether this variation with Illumina barcode was due to differences between these libraries as to total number of sequence reads. Illumina library barcode did explain variation in total read number ($F_{(2,42)}$ = 3.29, $p = 0.047$; Fig. 12) and we therefore explored the effect of including number of reads per sample on variation between birds on the community metrics. This did not affect the ability of the model to differentiate between bird samples, and read number did not explain any further significant variation above that explained by BirdID for any community metric (species richness $F_{(2,41)}$ = 390, $p < 0.001$; Shannon diversity $F_{(2,41)}$ = 342, $p < 0.001$; Simpson diversity $F_{(2,41)}$ = 71.1, $p < 0.001$). Nevertheless, we included Illumina library barcode as a fixed factor in later analyses in this chapter to investigate its possible effect on patterns of variation in seminal microbiotas.

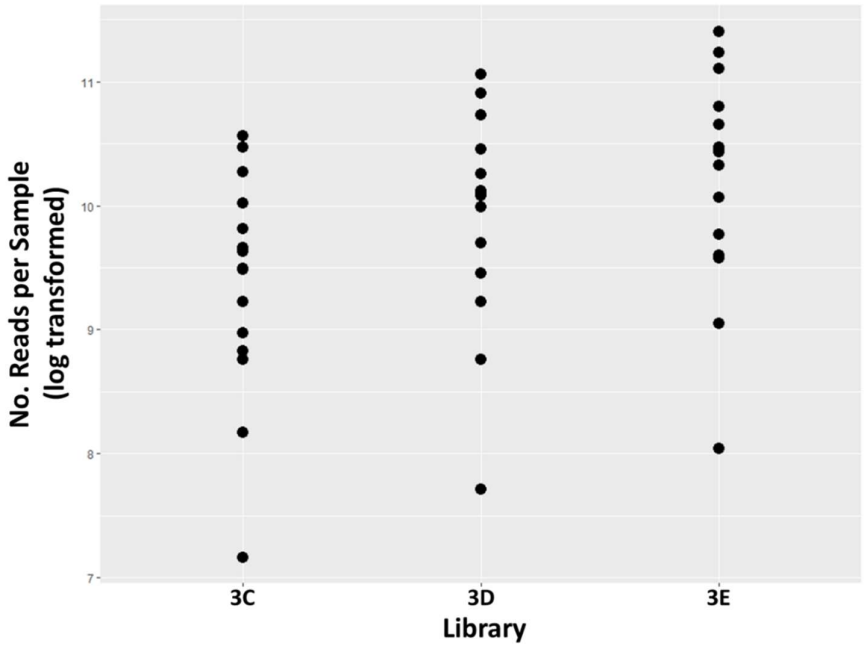


Figure 12 *Log transformed numbers of reads in samples from each Illumina barcode library, showing that there is a difference between libraries.*

Sample Repeatability

Repeatability scores based on the two methods (LMM and ANOVA) were congruent and indicated that the aliquots of a sample were considerably repeatable, although only the Simpson diversity significantly differentiated samples (Table 5). it should be pointed out however, that lack of significant differentiation between samples in species richness and Shannon diversity could be due to the fact that the samples were not different in these community metrics upon collection, rather than differences between aliquots accrued during sample processing and sequencing obscuring between sample differences.

Table 5 *Repeatability scores calculated based in Linear Mixed Models (LMM) and ANOVA, and significance of sample ID in predicting the three sample community metrics*

Repeatability scores			
	Species Richness	Shannon Diversity	Simpson Diversity
LMM	R = 0.319	R= 0.390	R= 0.443
ANOVA	R= 0.316	R = 0.389	R = 0.438
Significance of sample ID			
ANOVA	$F_{(8,14)} = 2.17$ p = 0.098	$F_{(8,14)} = 2.62$ p = 0.055	$F_{(8,14)} = 2.98$ p = 0.036

When comparing the ranks of OTUs based on relative abundances, actual and random pairs of aliquots from the same semen sample were significantly more positively correlated than those from randomly aliquoted pairs in both OTU subsets both, when including average relative abundance >1% ($t = 4.89$, $df = 9.98$, $p = 0.001$), and when including average relative abundance >0.1% ($t = 4.19$, $df = 14.8$, $p = 0.001$). Likewise, in both OTU subsets random pairs had higher Bray-Curtis dissimilarity scores than actual pairs ($t = -5.20$, $df = 15.1$, $p < 0.001$ when including average relative abundance >0.1%; $t = -5.08$, $df = 14.3$, $p < 0.001$ when including average relative abundance >1%). Dissimilarity scores were higher when the relative abundance for inclusion was lower. When comparing actual sample pairs, mean Bray-Curtis dissimilarity score was 0.390 when OTU inclusion criterion was 0.1% average relative abundance compared with 0.328 for 1% inclusion criterion, and 0.648 at 0.1% inclusion vs 0.609 at 1% inclusion when comparing random pairs.

2. Characterisation of the seminal microbiota and basic patterns of variation

From 171 whole semen samples taken from 65 males by abdominal massage (145) and controlled copulations (26) throughout the 2015 and 2013 breeding seasons we obtained 698 OTUs following contamination removal and rarefaction, with an average of 65.6 OTUs per sample. Histograms of OTU frequency and average relative abundance show that most of these OTUs occurred at very low abundance and only in few samples (Fig. 13).

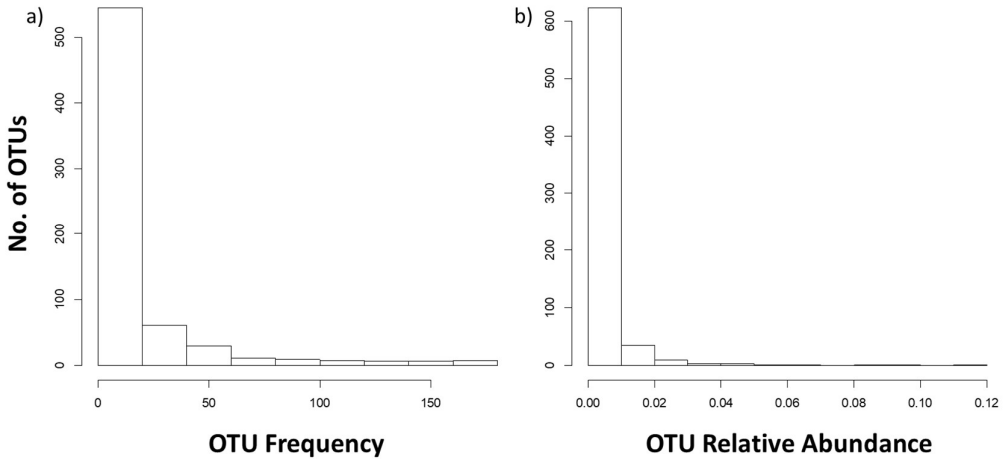


Figure 13 *Histograms illustrating a) the number of OTUs at different frequencies, with majority of OTUs occurring in a low proportion of samples and b) the number the OTUs occurring at different average relative abundances across samples, with the majority of OTUs occurring at a low relative abundance on average.*

Identified OTUs belonged to 22 phyla. The seminal microbiota was dominated by four phyla, which made up >99% of the sequences; *Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Bacteriodetes*. The order of phyla by overall abundance differed slightly to order by average within sample relative abundance (Table 6).

Table 6 *Phyla detected in red junglefowl ejaculate samples. Average relative abundance is the average per sample relative abundance of each phylum, and Proportion of Overall Abundance is the Proportion of all reads in all samples in this dataset belonging to each phylum. The phyla are ranked by Proportion of Overall Abundance. Figures are not rounded to 3 s.f. to illustrate differences between phyla.*

Phylum	Average Relative Abundance	Proportion of Overall Abundance
Firmicutes	0.006101	0.475602
Actinobacteria	0.005875	0.423783
Proteobacteria	0.003815	0.074645
Bacteroidetes	0.003554	0.012596
Fusobacteria	0.016171	0.006860
TM7	0.002640	0.001912
Cyanobacteria	0.009858	0.001804
Chloroflexi	0.006455	0.001015
WPS-2	0.005354	0.000302
Synergistetes	0.003505	0.000238
Tenericutes	0.002362	0.000227
Acidobacteria	0.002413	0.000216
Other	0.002206	0.000162
Gemmatimonadetes	0.003027	0.000151
Deferribacteres	0.002103	0.000119
SR1	0.017208	9.72E-05
Nitrospirae	0.003107	7.56E-05
GN02	0.002868	6.48E-05
Elusimicrobia	0.001912	4.32E-05
Fibrobacteres	0.001912	2.16E-05
Lentisphaerae	0.001912	2.16E-05
Spirochaetes	0.003824	2.16E-05
Chlorobi	0.001912	1.08E-05
TM6	0.001912	1.08E-05

Principal Components Analysis (PCA) was used to explore the variation across individual males in composition of the seminal microbiota, and to identify variables that capture inter-individual variation, and which can be used to identify sources of variation in the microbiota composition of the population. In this PCA axes explained little variation, with the first axis explaining just 9.78% of total variation, and the second 7.20%. This was reduced to 8.69% and 6.01% respectively, when the dataset was transformed into Heillinger distances to account for the high number of zeros. This is to be expected; the nature of the community as demonstrated by Fig. 13 was such that the

majority of the 646 OTUs occur at very low relative abundance and in very few of the birds. This means that a lot of variation between birds was explained by many OTUs occurring in just a few birds at very low abundance. To address this, we performed another PCA using a subset of 39 OTUs that occurred at minimum average relative abundance of 1% and in more than 10% of the birds.

In this PCA, the first axis described 12.7% of the variation and the second 10.1%. We can see that the birds are now separated along PC1 and PC2 (Fig. 14a), although no clustering of birds was immediately evident, and a number of OTUs seem to be correlated with the two axes (Fig. 14b).

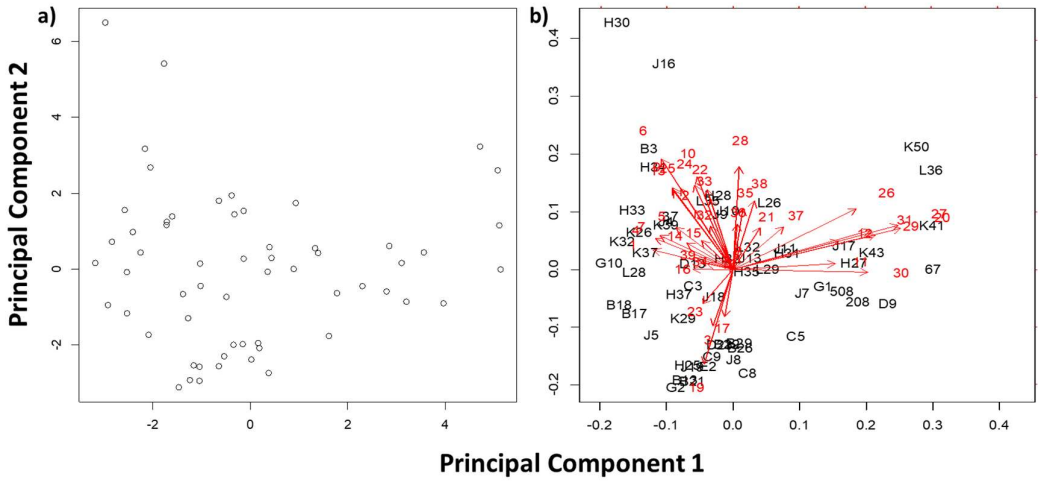


Figure 14 *Graphical representation of each bird's values of the first two PCA axes. Figure b) includes the identity of the bird for each point and red arrows indicate the alignment of relative abundances of the OTUs with axis values. Numbers indicate OTU. Birds are separated along both axes, and a number of OTUs appear to be correlated with one of the components.*

To further explore these patterns, the Spearman correlation of occurrence profiles of the subset of 39 OTUs was calculated and this is displayed in Fig. 15. with the dendrogram representing hierarchical clustering of the OTUs. The first branching of the dendrogram was explained by OTUs which co-occur and were positively correlated with PC1 (predominately *Firmicutes* species) and the second with those negatively correlated with PC2 (predominantly *Actinomyces* species). The heatmap indicates that the OTUs

within these two clusters appear to be the most tightly associated with each other in terms of co-occurrence.

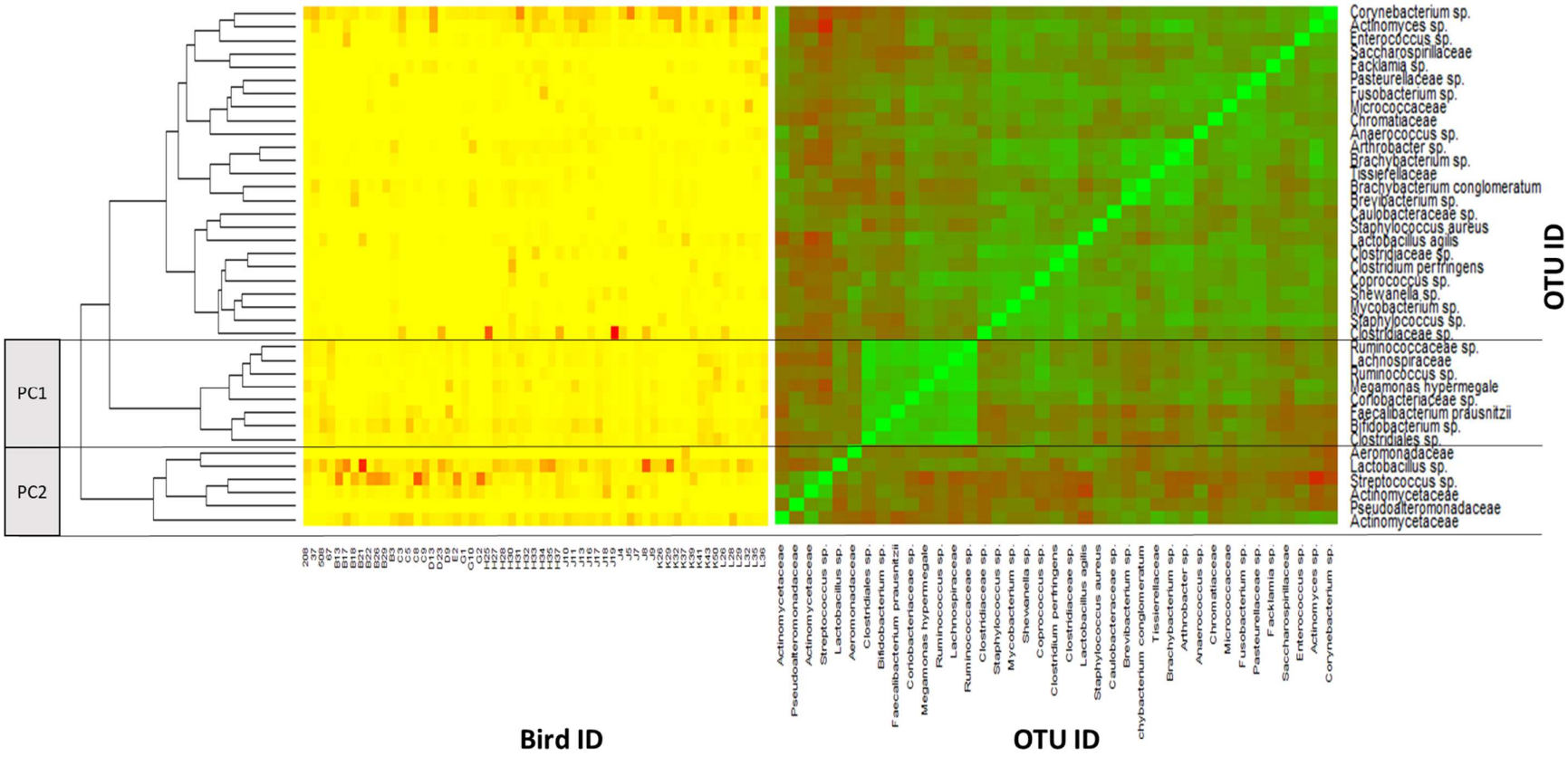


Figure 15 *The average relative abundance of each OTU per bird, with yellow indicating low average relative abundance and red high, BirdID on the horizontal axis and OTU on the vertical (left heatmap). The Spearman correlation of occurrence profiles of OTUs with red a negative correlation and green a positive correlation, and the stronger each colour intensity (i.e. bright red or bright green) the more positive or negative the correlation (right heat map). The dendrogram to the left illustrates hierarchical clustering of OTUs according to Spearman Correlations of their occurrence profiles PC1 and PC2 indicate the OTU clusters which correlate positively with PCA Axis 1 and PCA Axis 2 respectively.*

3. Comparing microbiotas

Abdominal massage versus controlled copulations

The differences in microbiota of the two sample collection methods, abdominal massage and controlled copulation, was explored using three datasets: i) 99 first semen samples (i.e. produced by a male at the beginning of a sampling trial on a given day), comprising 73 samples obtained from abdominal massage (AM) and 26 obtained from controlled copulation (CC), ii) a subset of this dataset, comprising 26 randomly selected AM samples and the 26 CC samples, and iii) an additional subset of 12 matched samples, where six pairs of AM and CC samples collected from the same bird in the same collection period (i.e. within two days of each other) are used.

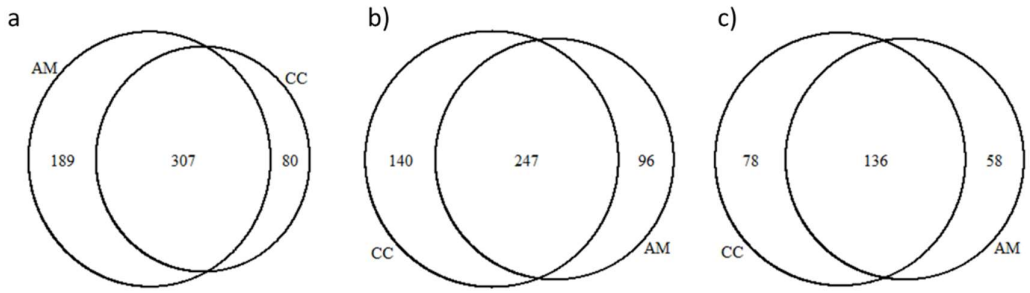


Figure 16 *Venn diagrams showing the OTUs shared by and unique to the two ejaculate collection methods using a) all 99 samples b) 52 samples with 26 of each collection method c) 12 samples of six pairs collected from the same bird within the same collection period. CC indicates controlled copulations and AM abdominal massage.*

From all first semen samples – 26 controlled copulations (CC) and 73 abdominal massage (AM) – 576 OTUs were found overall, with 496 in AM samples and 387 in CC samples. 307 (53%) of these were found in both, 80 (14%) were unique to CC, and 189 (19%) unique to AM (Fig. 16a). When a subset of samples was used for AM so that both CC and AM had 26 samples each, 483 OTUs were detected overall, with 247 (51%) found in both, 96 unique to AM (20%), and 140 (29%) unique to CC (Fig. 16b). When only those 12 samples making up the six matching pairs were used, 272 OTUs were detected overall, 58 (21%) were unique to AM, and 78 (29%) unique to CC, whilst 136 (50%) were shared (Fig. 16c). When all samples were analysed in a linear model, ANOVA indicated collection method was a significant factor explaining variation in species richness ($F_{(1,97)} = 12.0$, $p < 0.001$), Shannon diversity ($F_{(1,97)} = 10.3$, $p = 0.002$), and Simpson

diversity ($F_{(1,97)}$ = 6.79, p = 0.011), with controlled copulations having higher richness and diversity (Fig. 17).

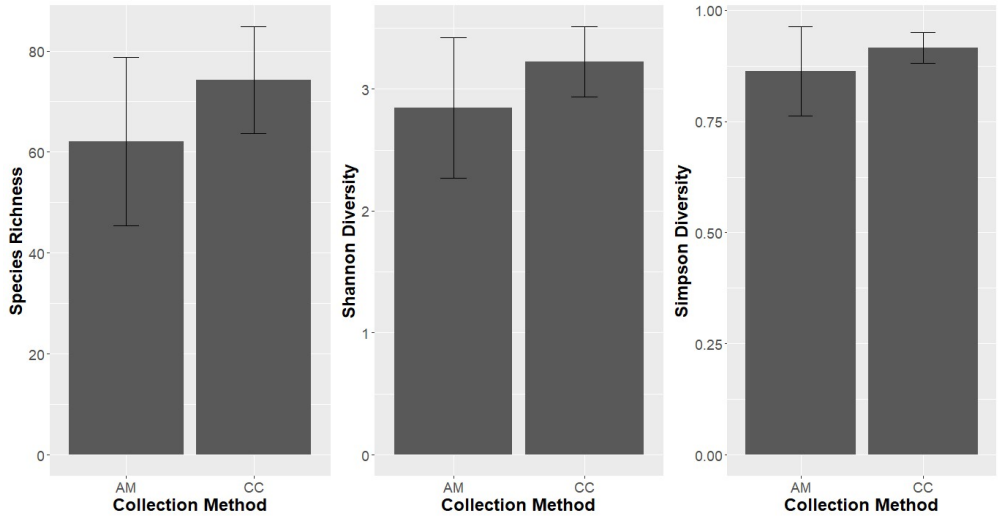


Figure 17 *Community metrics of semen samples collected by the two different methods. CC indicates controlled copulations and AM indicates abdominal massage. All ninety-nine samples were used, vertical bars indicate standard deviations.*

When the same analysis was performed using the reduced subset of 52 samples species richness and Shannon diversity were significantly higher in CC samples ($F_{(1,50)}$ = 5.46, p = 0.023, and $F_{(1,50)}$ = 4.24, p = 0.045, respectively; Fig. 18), though Simpson diversity was now no longer significantly different between the two sampling methods.

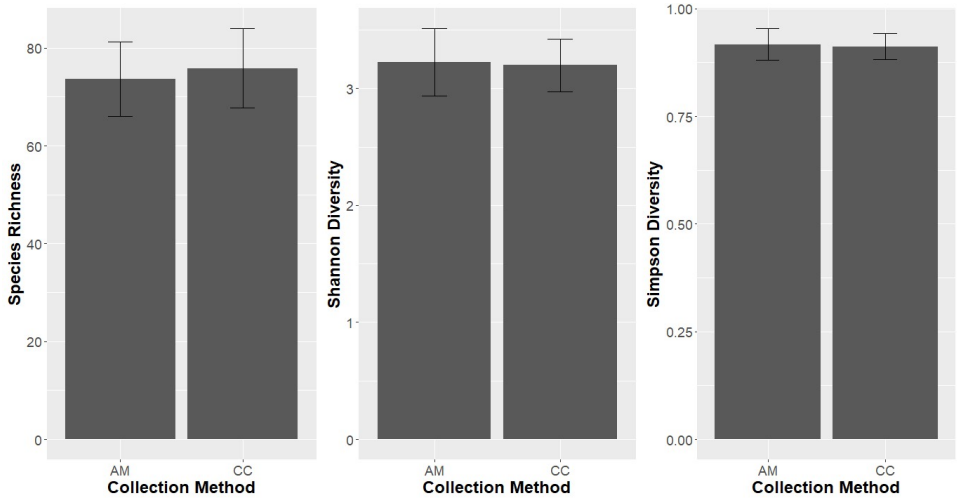


Figure 18 *Community metrics of semen samples collected by the two different methods. CC indicates controlled copulations and AM indicates abdominal massage. A subset of 52 samples (26 of each collection method) was used, and vertical bars indicate standard deviations.*

None of the three community metrics were significantly explained by collection method when only the 12 samples of the matching pairs (species richness: $F_{(1,10)} = 0.226$, $p = 0.801$; Shannon diversity: $F_{(1,10)} = 0.033$, $p = 0.859$; Simpson diversity: $F_{(1,10)} = 0.226$, $p = 0.645$). These results did not change qualitatively for any subset when Illumina library barcode was included as a fixed effect (Table 7).

Table 7 *The significance of collection method in predicting variation in community metrics, according to ANOVA of linear models containing both collection method and Illumina library barcode as fixed effects. Bold italics indicate significant effects.*

	Species Richness	Shannon Diversity	Simpson Diversity
All 99 samples	$F_{(1,89)}$ 12.5, $p < 0.001$	$F_{(1,89)}$ 10.7, p=0.002	$F_{(1,89)}$ 7.06, p=0.009
52 sample subset	$F_{(1,42)}$ 5.35, p= 0.026	$F_{(1,42)}$ 4.70, p=0.035	$F_{(1,42)}$ 3.02, p=0.089
12 sample subset	$F_{(1,9)}$ 0.210, p=0.657	$F_{(1,9)}$ 0.052, p=0.825	$F_{(1,9)}$ 0.105, p= 0.754

When we considered ranks of OTUs by relative abundances, two of the six pairs were significantly positively correlated ($t = 0.930$ $p < 0.001$; $t = 0.596$, $p = 0.003$) in the rankings of OTUs present at an average relative abundance greater than 1%. When this analysis was extended to include OTUs present at an average relative abundance greater than 0.1%, all pairs were significantly positively correlated ($t = 0.239$, $p = 0.028$; $t = 0.399$, $p < 0.001$; $t = 0.534$, $p < 0.001$; $t = 0.756$, $p< 0.001$; $t = 0.427$, $p < 0.001$; $t = 0.426$, $p < 0.001$).

Whole Semen, Seminal Fluid and Sperm Pellets

The three samples resulting from different post-collection sample processing methods are compared using 33 semen samples for which all three types were collected. In 99 samples, 599 OTUs were detected overall. 237 of these were shared across all three types (39.6%), 41 unique to whole semen samples (EJA) which contained 379 overall, 141 OTUs were unique to seminal fluids (SF) which contained 447 in total and 55 unique to sperm pellets (SP), which contained 372 OTUs in all (Fig. 19).

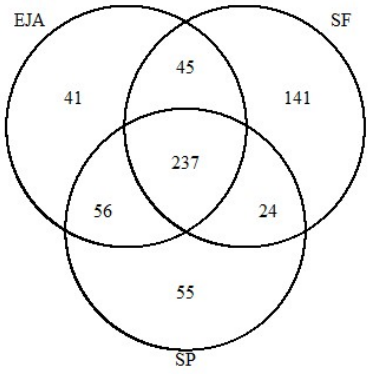


Figure 19 *Venn diagram showing the OTUs shared by and unique to the three different semen sample types. EJA indicates whole semen samples, SF indicates seminal fluid samples and SP indicates sperm pellet samples.*

ANOVA indicated that sample type was a significant source of variation in measures of species richness ($F_{(2,96)} = 4.15$, $p = 0.019$) and Shannon diversity ($F_{(2,96)} = 4.10$, $p = 0.020$), but not Simpson Diversity, with seminal fluid samples having higher Shannon diversity and species richness (Fig. 20). The inclusion of library as a fixed factor did not affect the conclusion of this analysis (species richness: $F_{(2,90)} = 4.73$, $p = 0.011$; Shannon diversity: $F_{(2,90)} = 4.28$, $p = 0.017$; Simpson diversity: $F_{(2,90)} = 2.58$, $p = 0.081$).

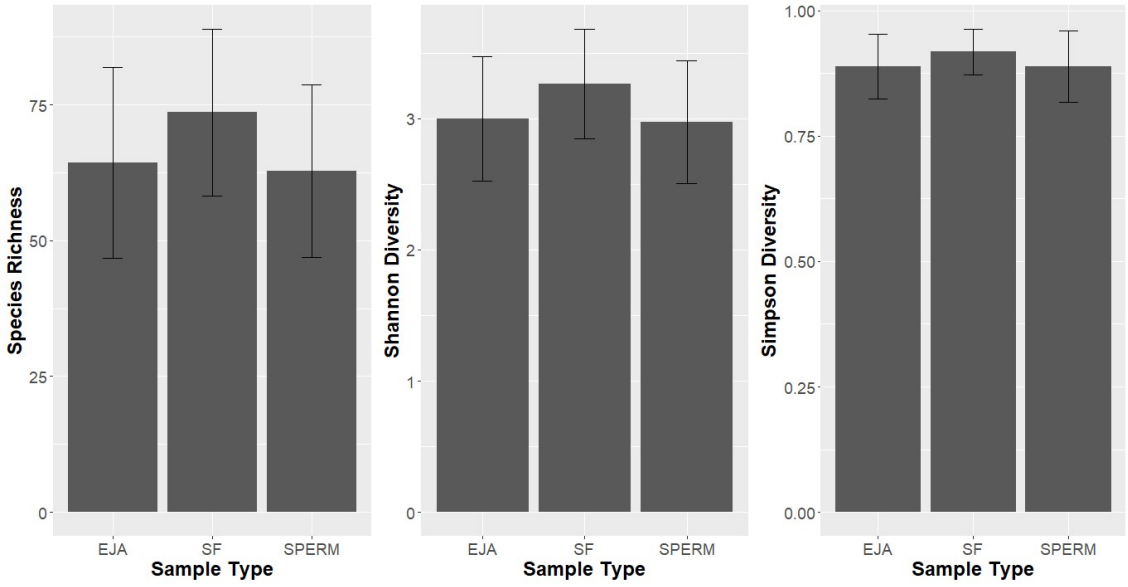


Figure 20 *Community metrics in different semen sample types. EJA indicates whole ejaculate samples, SF indicates seminal fluid samples and SP indicates sperm pellet samples. Vertical bars represent standard deviation.*

SF samples had significantly higher measures of species richness, Shannon diversity and Simpson diversity than both EJA and SP samples (Table 8).

Table 8 *Output from T-tests comparing community metrics between different semen sample types. EJA indicates whole semen samples, SF indicates seminal fluid samples and SP indicates sperm pellet samples. Bold italics indicate significant test statistics.*

	Species Richness	Shannon Diversity	Simpson Diversity
EJA V SF	***t = -2.27, df = 63.7, p = 0.026***	***t = -2.41, df = 63.8, p = 0.019***	***t = -2.14, df = 59.6, p = 0.037***
EJA V SP	t = 0.390, df = 64.7, p = 0.698	t = 0.236, df = 65.0, p = 0.814	t = -0.016, df = 64.0, p = 0.987
SF V SP	***t = 2.90, df = 65.0, p = 0.005***	***t = 2.78 df = 64.7, p = 0.007***	***t = 2.03, df = 57.1, p = 0.047***

Similarly, EJA and SP samples from the same semen sample were significantly positively correlated in Shannon diversity, Simpson diversity and species richness. This significant positive correlation was also present between SF and SP samples although not as strong, whereas EJA and SF samples were only significantly correlated in species richness (Table 9).

Table 9 *Correlation co-efficients and p-values of Pearson correlation analyses comparing community metrics between the three sample types. EJA indicates whole ejaculate samples, SF indicates seminal fluid samples and SP indicates sperm pellet samples*

	Species Richness Correlation	Shannon Diversity Correlation	Simpson Diversity Correlation
EJA v SF	0.287, p = 0.019	0.287, p = 0.117	0.252, p = 0.171
EJA v SP	0.454, p = 0.010	0.502, p = 0.004	0.447, p = 0.012
SF v SP	0.383, p = 0.033	0.391, p = 0.030	0.434, p = 0.015

Considering the rank of OTUs by relative abundance, all three sample types were significantly positively correlated with each other using both OTU relative abundance inclusion criteria, with all correlations more positive with the higher inclusion criteria. EJA and SP samples were more positively correlated to each other than either were to SF samples (Table 10).

Table 10 *Correlation co-efficients and p-values of Pearson correlation analyses comparing OTU rank relative abundances between the three sample types. EJA indicates whole ejaculate samples, SF indicates seminal fluid samples and SP indicates sperm pellet samples. 1% indicates output when only OTUs present at a relative abundance equal or greater than 1% are included. 0.1% indicates output when OTUs present at a relative abundance equal to or greater than 0.1%*

	1%	0.1%
EJA v SF	0.704, $p < 0.001$	0.632, $p < 0.001$
EJA v SP	0.900, $p < 0.001$	0.847, $p < 0.001$
SF v SP	0.789, $p < 0.001$	0.698, $p < 0.001$

Comparison to Faecal Microbial Communities

The overlap of the faecal and seminal microbiota was investigated, using 26 faecal samples, 26 controlled copulation samples and 149 abdominal massage semen samples, and then a subset of these including matching faecal and semen samples from the same birds which comprises 20 faecal samples and 20 semen samples, 4 of which are controlled copulation samples and 16 of which are abdominal massage samples.

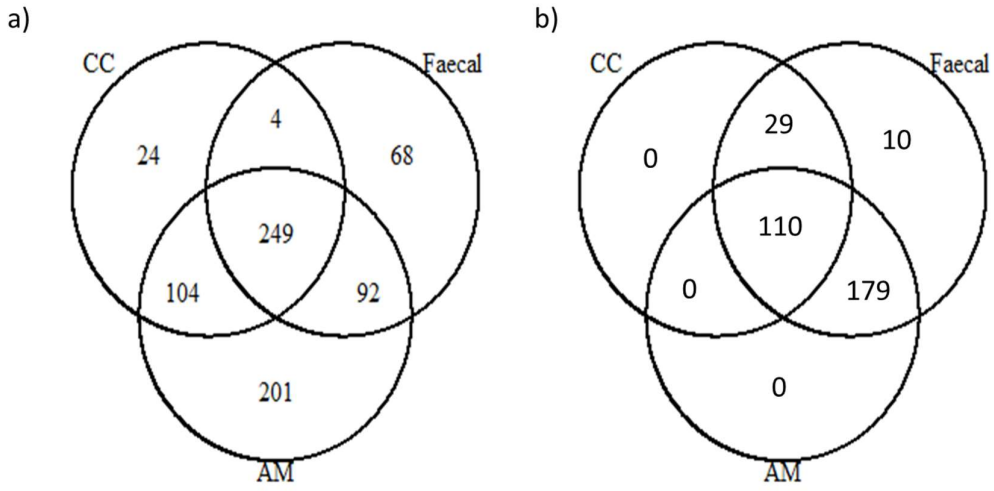


Figure 21 *Venn diagram showing the numbers of OTUs shared between different combinations of the three sample types, and the number of OTUs unique to each, using a) all 175 semen and 26 faecal samples and b) 20 faecal samples with 20 ejaculate samples from the same bird. CC indicates controlled copulation samples and AM indicates abdominal massage samples.*

Of 381 OTUs found in 26 CC samples, 66.4% were found in faecal samples and 25.8% of 646 OTUs from 149 AM samples were found in faecal samples. 83.5% of 413 OTUs found in 26 faecal samples were found in semen samples, 61.3% in CC samples and 82.6% in AM samples.

When sample sizes were matched such that there were 20 samples of all three types, 100% of OTUs in semen samples collected by either method were found in faecal samples, and 97.0% of faecal OTUs were found in ejaculate samples (Fig. 21).

Faecal and semen samples were significantly different in species richness, with faecal having a higher number of species, and this was the case for both controlled copulation and abdominal massage samples. They were also significantly different in Shannon and Simpson diversity, with semen samples being more diverse than faecal samples, but this difference was restricted to control copulations (Table 11; Fig. 22).

Table 11 *Table of T-test statistics and p-values from Welch Two Sample T-tests comparing all semen samples with faecal samples, just those collected by controlled copulation (CC) with faecal and just those collected by abdominal massage (AM) with faecal. Bold italics indicate significant t-statistics.*

	All Ejaculates	CC Ejaculates	AM Ejaculates
Species richness	***t = -4.60, df = 29.0, p < 0.001***	***t = -2.53, df = 39.5, p = 0.015***	***t = -4.90, df = 29.8, p < 0.001***
Shannon diversity	t = 1.43, df = 28.9, p = 0.162	***t = 2.76, df = 38.9, p = 0.009***	t = 1.12, df = 29.8, p = 0.269
Simpson diversity	***t = 2.07, df = 27.3, p = 0.048***	***t = 2.95, df = 30.0, p = 0.006***	t = 1.88, df = 28.0, p = 0.071

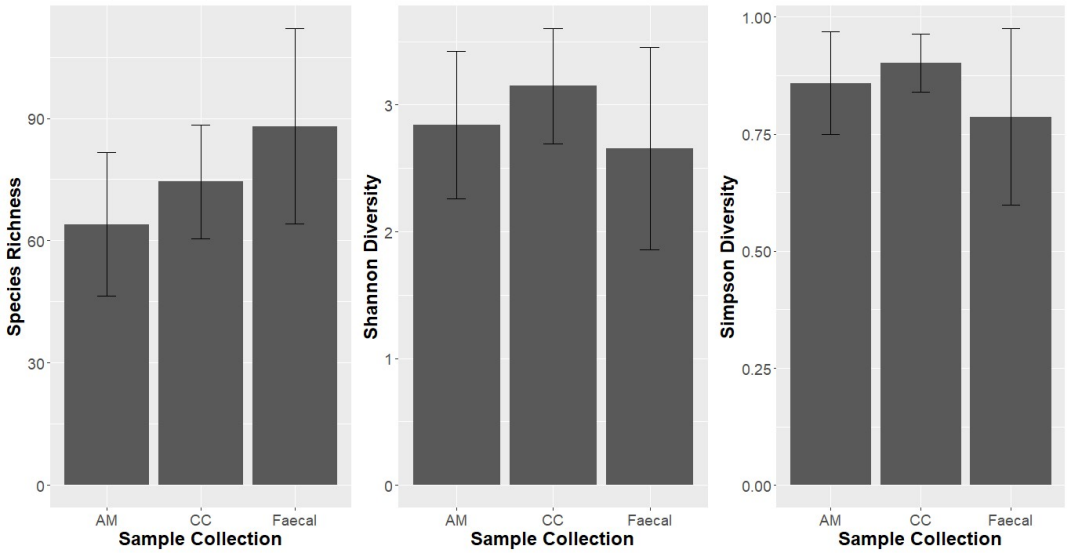


Figure 22 *Graph illustrating mean species richness, Shannon diversity and Simpson diversity for abdominal massage (AM), controlled copulation (CC) and Faecal samples. Vertical bars indicate standard deviation for each sample type.*

Exploring how closely the composition of faecal samples matched semen samples taken from the same bird, we used 20 matching pairs of faecal and semen samples and compared community diversity measures and ranks of OTU abundances between pairs. Of the 20 semen samples, four were CC and sixteen AM. In this comparison, semen and faecal samples only differed in Simpson diversity, with semen samples having greater diversity than faecal samples (Table 12; Fig. 23).

Table 12 *Table of T-test statistics and p-values from paired T-tests comparing 20 semen samples with 20 matching faecal samples, just those collected by controlled copulation (CC) with faecal and just those collected by abdominal massage (AM) with faecal. Bold italics indicate significant differences.*

	All Ejaculates	CC Ejaculates	AM Ejaculates
Species richness	t = 0.803, df = 22, p = 0.431	t = 1.22, df = 3, p = 0.311	t = 0.347, df = 18, p = 0.733
Shannon diversity	t = 2.01, df = 22, p = 0.057	t = 1.33, df = 3, p = 0.275	t = 1.52, df = 18, p-value = 0.146
Simpson diversity	***t = 2.65,*** ***df = 22,*** ***p = 0.015***	t = 1.32, df = 3, p = 0.279	***t = 2.23,*** ***df = 18,*** ***p = 0.038***

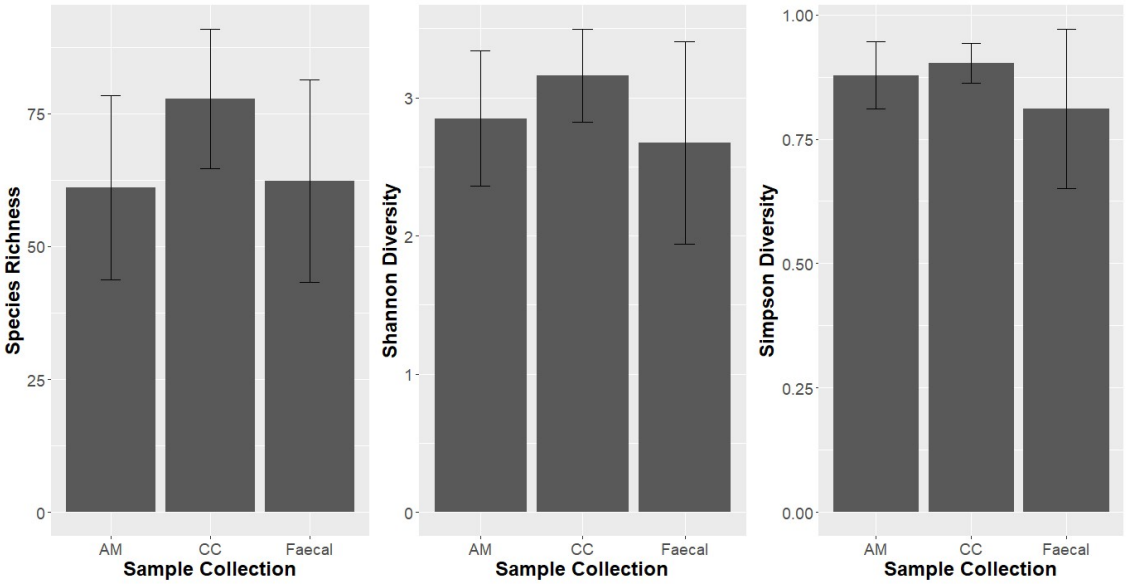


Figure 23 *Graph illustrating mean species richness, Shannon diversity and Simpson diversity for abdominal massage (AM), controlled copulation (CC) and faecal samples, using the subset of matching faecal and ejaculate samples. Vertical bars indicate standard deviation for each sample type.*

When OTUs present in the samples were ranked by abundance, the correlation between samples from the same bird ranged from 0.244 to 0.473, with no clustering of controlled copulation or abdominal massage samples at either end of the range. Significance could not be accurately computed due to the high number of ties where an OTU was absent from both samples. When a subset of inclusion of 0.1% average relative abundance was used, there is now a wider range of correlations -0.021 to 0.508. When a subset of inclusion of 1% average relative abundance was used, there was now a wider range of correlations -0.210 to 0.601. Again, there was no pattern of correlation distribution between controlled copulation and abdominal massage.

DISCUSSION

Evidence is emerging that male semen plays host to diverse communities of bacterial species, which play potentially fundamental roles in the evolutionary ecology of host populations through their effects on host reproductive success and fitness. However, very little remains known about the seminal microbiota of non-human organisms, and the field suffers from a lack of standard methodological approaches. We employed a next-generation sequencing approach to characterise the seminal microbiota of male red junglefowl. A methodological pipeline of our sampling and sequencing approach was devised and evaluated, and the effect of differences in semen collection and initial processing methods was investigated, as well as the overlap of the seminal microbiota with the faecal microbiota.

Part of this work focused on addressing the challenges of high-throughput sequencing and targeting of the 16S rRNA gene. One of the most problematic issues that may arise, given the broad range and sensitivity of this approach, is contamination of the samples; indeed, it has been shown by Salter *et al*. (2014) that the results of microbiota studies could be affected by not taking account of contamination. We implemented a strict protocol to minimise contamination and a method to deal with contaminant OTUs that remain. Samples processed in this way generate communities that are likely to be genuine avian ejaculate communities and allow exploration of

variation between samples. Whilst we did not directly compare sequenced samples with and without the contamination, our present work is consistent with Salter *et al*. (2014) and indicated that contamination did occur. The methodology proposed in this study reduced the risk of contamination and was able to identify and analytically remove persistent contaminants, likely originating from reagents.

A second challenge relates to variation in total number of reads per sample following sequencing and initial processing. We found that the majority of our samples achieved a high level of community coverage despite this variation, and that whilst standardisation through rarefaction to the minimum number of reads per sample means a loss of information, and variation in read number per sample leads to differences in the number of OTUs in rarefied samples, this did not obscure differentiation in samples. Further, we found that this variation in sequence number was associated with Illumina library barcode, but inclusion of library barcode did not affect conclusions of later analyses. Therefore, we can conclude that variation between samples in total read number due to differential sequencing linked to Illumina barcode does not obscure or confound patterns of variation across samples.

A similar pattern was confirmed by our analysis of samples repeated with PCR and sequencing library barcodes; despite variation, samples can still be distinguished from one another. We also explored the potential of variation between samples due to post collection processing; aliquots of the same ejaculate taken immediately post sample collection was used to investigate the collective effect of all the factors and steps between collection and sequencing that may affect the bacterial composition of the samples as a result of laboratory process. ID of the semen sample, as opposed to aliquot, predicted significant variation in Simpson diversity and in none of the other community metrics, with Simpson diversity also giving the highest semen sample repeatability scores. This suggested that post collection factors added to variation between samples, on top of biologically relevant variation originating in the actual composition of the original semen samples. However, aliquots from the same semen sample had more closely

correlated OTU abundances and lower dissimilarity measures to one another than to aliquots from other ejaculates. This suggests that variation between aliquots of the same ejaculate lies in the high number of rare OTUs, which is further supported by the fact that Simpson diversity (which puts more weight on more common species, Morris *et al*., 2014) is the most repeatable and differentiating of the three sample types used. This emphasises the benefit of using multiple measures of community structure.

In light of the fact that we found some repeatability in the bacterial communities of different samples, that Simpson diversity differentiates samples, and that differences in all community measures detect differences in samples due to collection method and type in this study, we can conclude that variation due to laboratory process, if properly quantified and taken into account, does not necessarily confound variation due to actual biological differences between semen samples. Our study however, does illustrate that it is necessary to ensure that samples are processed in as standard and consistent a manner as possible, and that caution does need to be exercised when interpreting results of analyses based on just single measures of community composition and structure. Where effect sizes being investigated are expected to be very slight, multiple repeats of every semen sample used would increase confidence in conclusions that there is no difference between samples, but this is not always possible for instance given the time and cost involved in such studies and given that the sample is not always large enough to be split into multiple aliquots.

The red junglefowl seminal microbiota that we describe comprises of an average of approximately 65 OTUs per sample, and 698 OTUs overall, from 171 semen samples from 65 males. 23 identified phyla were represented in all, and the seminal microbiota is dominated by the same four phyla that dominate the microbiota of the domestic chicken gut (Videnska *et al*., 2013) and the cloaca of the European swallow (Kreisinger *et al*., 2015): *Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Bacteriodetes*. These studies also found a similar pattern of sequence read distribution across phyla as in this study, with the top phylum making up

approximately half of the sequence reads and the other three of these four the vast majority of the remainder. Other phyla never contribute more than 1% of sequence reads. Further, in the swallow only four OTUs were found in every sample and approximately 5% of OTUs in more than half of the samples, which is reflective of the junglefowl seminal microbiota described here, where the majority of OTUs are found at low frequencies. However, in the swallow, the *Proteobacteria* phylum made up the highest proportion of sequence reads across all samples, whereas in the junglefowl *Firmicutes* made up the highest proportion. In the domestic chicken, faecal samples the highest phylum in terms of proportion of sequence reads was also *Firmicutes*, but *Proteobacteria* and *Actinobacteria* contributed a lower proportion and *Bacteroidetes* a higher proportion, than in the junglefowl seminal microbiota. These similarities in identity and patterns of the dominating phyla gives us further confidence that the communities we have described largely capture genuine seminal microbiota. We detected some variation amongst individual male birds, exemplified by the Principal Components Analysis, where males are spread out along the two axes of variation when the most frequent and on average most abundant OTUs are included. It is worth mentioning that whilst these axes separate birds, they do not explain all the variation amongst birds, but this does not mean that these axes cannot be useful in looking for a source of the variation that is captured by the axes of this analysis. A heatmap and hierarchical clustering reveals two clusters of co-occurring OTUs apparently driving these axes of variation; organisms in one cluster predominantly belonging to the *Firmicutes* column and their average relative abundance in samples also all positively correlated with score on PCA Axis 1, and organisms in another cluster predominantly belonging to the *Actinomyces* phylum and their average relative abundance in samples also all positively correlated with score on PCA Axis 2. Therefore, it would seem that at least part of the variation between birds is driven by abundance of these two phyla. This could be leading to differences in function of the microbiota between birds or it could be that these two microbiota clusters represent two types of stable microbiota state. For instance, in humans the vaginal microbiota exists in one of five states, four of which are dominated by

Lactobacillus species, and the fifth, which is characterised by not being dominated by one single taxonomic group, and is linked to prevalence of symptoms of vaginal dysbiosis (Ravel *et al*., 2011) and even protection against STIs (Borgdorff *et al*., 2014). Similarly, the healthy gut microbiota in humans exists in one of three stable microbiota states, dominated by *Bactereoides sp., Prevotella sp*., or *Ruminococcus sp*., which do not differ in 'functional richness' i.e. the range of functions provided by the bacterial community. It is worth mentioning that as the seminal microbiota is a transient community not sourced all from the same physiological location, then the analogy in terms of a stable, functional microbiota with gut and vaginal microbiota might be less relevant here.

Another challenge in the study of seminal microbiotas pertains the collection of semen samples. These are often obtained invasively (i.e. through dissection, e.g. Javurek *et al*., 2016) or manually, raising questions as to the degree to which they capture the microbiotas contained in natural ejaculates. Our study was able to compare and contrast the microbiotas of semen samples obtained manually, through abdominal massage and from natural copulations (through controlled copulations). We showed that whilst samples collected by the two methods are by no means identical, more than half of OTUs detected are present in both types of samples. Samples from controlled copulations (i.e. natural ejaculates) however, have a greater diversity and species richness than samples from abdominal massage. The fact that this difference was not significant in just the subset of matched samples could either be due to insufficient sample size or may indicate that differences between non-matched abdominal massage and controlled copulation samples reflect confounding factors, such as individual male effects. OTU rank abundances are significantly positively correlated between samples from the same birds collected within the same collection period, suggesting that whilst the two sample types do vary in overall content of OTUs, they still reflect each other, which again could explain why differences in community metrics are not apparent when matched samples are used. Thus, whilst both types of samples can be used to describe and explore patterns of variation in the seminal microbiota, it is important to realise that whichever method is used will influence the community generated and so one in

isolation may not be appropriate if the aim of a study is to comprehensively describe the entire reproductive microbiota.

We also found differences between semen sample types - whole semen, seminal fluid and sperm pellet. The potential advantage of using seminal fluid samples is that spermatozoa cells are removed, the presence of which poses problems downstream in the laboratory process; the presence of large amounts of fowl DNA released from the sperm cells during DNA extraction can obscure the detection of bacterial DNA in gel electrophoresis, used to quickly assess the success of extraction and amplification, as well as quantification of bacterial DNA using Qubit. However, whilst centrifugation to pellet and remove sperm cells could significantly reduce this problem, there is the chance that some bacteria are pelleted along with the sperm, and this in itself may reduce the chance of obtaining sufficient bacterial DNA to characterize samples. Further if some bacteria are more likely to be pelleted than others, either due to size differences or closer association with sperm cells (Vizzier-Thaxton *et al*., (2006) observed *Campylobacter* and *Salmonella* bacteria attached to sperm in domestic chickens using electron microscopy, with more *Salmonella* than *Campylobacter* attached to sperm cells on average), the resulting bacterial DNA may not accurately reflect the actual composition of the whole ejaculate microbiota. Investigation of the differences in the three sample types is therefore useful in informing the design and interpretation of future work. We found that while all types of samples shared some OTUs, a number of OTUs were unique to each type, with whole semen and sperm pellet samples sharing more with each other than seminal fluid samples, and seminal fluid samples containing a higher proportion of OTUs unique to that type. Seminal fluid samples contained more OTUs overall than the other two types, and per sample species richness and Shannon diversity is also greater in the seminal fluid sample than in the sperm pellet or the whole semen samples, which do not differ from each other. We conclude that some bacteria might be pulled down with the sperm pellet during centrifugation, but not all, and that the removal of spermatozoa from a sample may actually increase the rate of OTU detection. However, seminal fluid samples are not significantly

positively correlated with semen samples in terms of OTU rank abundances and so it would seem that there is some differentiation in OTUs that remain in the seminal fluid and thus such a sample might not entirely represent the whole ejaculate microbiota. It is thus more appropriate to use the whole semen sample as the presence of spermatozoa does not prevent the detection of bacterial DNA. Reasons as to the differential presence of OTUs in the different sample types may represent a mechanical effect where larger bacterial cells are more likely to be pelleted, or a closer physical association between certain bacteria and the sperm cells as observed by Vizzier-Thaxton *et al*. (2006), and so further work using such samples may be useful in exploring associations of bacteria with spermatozoa.

One potential source of seminal microbiota is the gut microbiota, especially in birds where the cloaca presents increased opportunity for the incorporation of gut microbiota into the ejaculate microbiota, and indeed vice-versa. Interestingly, we saw a complete overlap between the two sample types when equal numbers of faecal and semen samples were compared. All seminal OTUs were found in faecal samples, and all but 10 (of 328) faecal OTUs were found in semen samples. Samples differed in Simpson diversity, with semen samples having a higher diversity than faecal samples. This suggests that the majority of OTUs in either or both samples are also present in the cloaca, confirming that the common opening of digestive and reproductive tracts leads to increasingly similar microbiotas. This is in contrast to the findings in mice (Javurek *et al*., 2016) which showed that there was a unique seminal microbiota, with OTUs found in the seminal microbiota and not in the faecal microbiota. As there is evidence of bacteria in the chicken gut, and it seems to be dominated by similar phyla as the seminal samples characterised here the suggestion is that the ejaculate microbiota in birds is dominated by the faecal microbiota. However, given that only ten OTUs are unique to faecal samples, and that Javurek *et al*., (2016) find bacteria when observing ejaculates dissected directly from the seminal vesicles, it is unlikely that the ejaculate is completely sterile prior to passing through the cloaca. Rather, there could be a bacterial community in the cloaca made up of both gut and reproductive microbiota members, and that both faecal and ejaculate samples pick up OTUs from this microbiota upon ejaculation

and defecation. Faecal samples contained more species than semen samples, but were less diverse, so they seem to be more dominated by a small number of OTUs than semen samples. This could be interpreted as a result of the ejaculate most likely being a combination of different communities originating from different parts of the reproductive tract and sites with differing physiological conditions (Lake, 1957), including the faecal microbiota itself. However, there is no compelling evidence of correlation in diversity metrics or OTU rank abundance between matched samples suggesting that any variation in ejaculate microbiota is not being driven by the gut microbiota. Further dissection of birds would elucidate the source of the microbiota present in the ejaculate produced. Whilst this might not be relevant to consideration of the effects of the microbiota during mating itself, where the presence of the microbiota rather than its origin is more important, effects of exposure of sperm to bacteria prior to ejaculation could be explored.

To conclude, we have developed a methodological approach to characterise and explore variation in the seminal microbiota of male red junglefowl. Whilst no single approach is perfect, and the best method depends on the aims and scope of different studies, this work provides an understanding of the potential consequences of the different approaches which can be used to inform experimental design and interpretation of results in future work. We show that this microbiota is taxonomically similar to other avian and reproductive microbiota previously studied, with very larger overlap in OTU content to the red junglefowl faecal microbiota, and that it varies repeatedly across individual male birds. Whatever the sources of the ejaculate microbiota or consequences of such variation, the ejaculates transferred to female junglefowl contains diverse and distinct communities of bacteria. If this variation has consequences for female health, or the fertility of the male, then it certainly would lead to differential consequences of mating. The reproductive microbiota as a whole bacterial community should therefore be considered in terms of its potential to affect mating, and perhaps to be affected by mating, both in theoretical studies of the effect of sexually transmitted micro-organisms on the evolution of

mating systems and in empirical studies considering the mating decisions, behaviour and fertility outcomes in the red junglefowl.

CHAPTER 3. CHARACTERISING SEMINAL MICROBIOTAS OF RED JUNGLEFOWL LIVING UNDER DIFFERENT ECOLOGICAL CONDITIONS BY COMBINING *IN VITRO* CULTURING WITH GENOMIC SEQUENCING

Rosie Eccleston[a]; Harrison Ostridge[a]; Wook Kim[b]; Stephen Preston[c]; Tommaso Pizzari[a]

[a] Edward Grey Institute, Department of Zoology, University of Oxford, UK

[b] Department of Biological Sciences, Duquesne University, US

[c]Department of Zoology, University of Oxford, Oxford, UK

ABSTRACT

Next-generation sequencing techniques have hugely expanded the capacity to characterise and explore microbiotas through the simultaneous sequencing of DNA from hundreds of bacterial species across multiple samples. However, such techniques do not distinguish between live and dead bacteria. This distinction is crucial when studying the consequences of these communities for the host, particularly in the context of seminal microbiota where live bacteria could affect sperm performance, and affect females through sexual transmission. Here, we use *in vitro* culturing to demonstrate the presence of live bacteria in the seminal microbiota of male red junglefowl, *Gallus gallus*, and combine with Illumina next-generation sequencing to characterise and explore cultured communities, and compare to communities generated by direct sequencing of the same semen samples. Overall, 60% of the Operational Taxonomic Units (OTUs) directly-sequenced from semen samples were identified through culturing. Cultured communities were dominated by the same four phyla as those directly sequenced. However, 49% of the 217 cultured OTUs were not found through direct sequencing of the semen samples. This was to some degree due to the fact that different culturing conditions favoured the growth of OTUs that were very rare or occurred at very low levels in semen samples. Further, by exposing a sample to a set of different culturing conditions (aerobic, anaerobic and three different growth media) we show that structure and composition of cultured communities varied across culturing conditions i.e. the media used and the presence of oxygen. Collectively these results indicate that a large proportion of the bacteria in the red junglefowl seminal microbiota is live, and that these communities change dynamically with ecological conditions emphasising the need of combining *in vitro* culturing with next-generation molecular techniques when exploring similarly transient, complex and phenotypically diverse communities.

INTRODUCTION

The advent of powerful next generation sequencing genomic tools, which allow for the simultaneous and accurate identification of thousands of bacterial DNA sequences isolated directly from samples, has catalysed the study of host microbiotas in recent years (e.g. Consortium *et al*., 2012). Genomic tools, however, amplify bacterial DNA without distinguishing between live or dead bacteria. This issue is particularly relevant for certain environments such as faeces, parts of the gut and reproductive tract, and semen, which can in principle contain DNA from dead bacteria originating from other regions of the host. Characterising the community of bacteria living in a certain region of the host is critical to understand local ecological dynamics, interactions with the host and their fitness consequences. One way to address this challenge is to use *in vitro* culturing.

Culturing-based techniques have been traditionally used to study live bacteria. While useful in identifying live bacteria, culture-based approaches are however likely to offer a biased view of the microbiota as a whole, favouring species better adapted to the conditions of the culturing environment. The fact that many species will not be culturable or will only be cultured under certain conditions limits the power of culture-based technique to survey microbiota and can lead to contradictory outcomes across different studies. Therefore, rather than an exhaustive representation of the entire microbiota living within a host region, cultured communities represent a non-random subset of the microbiota, reflecting the component of the community best suited to grow under the particular ecological conditions determined by culturing conditions. While advances in the culturing techniques have shown that a larger proportion of the gut microbiota is culturable than previously believed (Eckburg *et al*., 2005), it is likely that even a wide range of culturing conditions may not capture the entire community.

A promising way forward is to combine the complementary strengths of next-generation sequencing with *in vitro* culturing. This is exemplified by Browne *et al*.'s (2016) approach, which demonstrated the value of culture-based techniques in combination with genomic sequencing

and computational modelling in the phenotypic analysis of the human gut microbiota. This two-pronged approach presents a number of advantages. First, genomic sequencing results in extensive coverage of the entire microbial community. Second, *in vitro* culturing tests whether bacterial DNA belongs to live organisms and sheds some light onto basic ecological conditions favoured by different bacterial species. The goal of this chapter is to combine next generation sequencing with *in vitro* culturing to study the communities of bacteria living in the semen of male red junglefowl, *Gallus gallus*.

Studies of live semen microbiota have typically used in vitro culturing approaches. The majority of such studies have focused on human samples with a specific clinical focus such as infertility (e.g. Vilvanathan *et al*., 2016) or chronic prostratis (e.g. Ivanov *et al*., 2009). Culture-based studies of the human seminal microbiota have typically revealed a wide range of live bacterial species, whilst also indicating considerable variation in patterns of bacterial isolation from samples, and in their functional composition. For instance, Vilvanathan *et al*. (2016) found bacteria in only 30% of ejaculate samples and Rodin *et al*. (2003) in just over half, whereas Willeén *et al*. (1996) found bacterial growth in more than 80% of samples, and Rehewy *et al*. (1979) reported culturing bacterial isolates from 68% of samples. Similarly, Kermes *et al*. (2003) found that anaerobes outnumbered aerobes in every sample, whereas Ivanov *et al*. (2009) and Shalika *et al*. (1996) did not find anaerobe isolates in any samples. Such differences most probably reflect the differences in culturing conditions used by different studies, which will select for the growth of different species and thus result in different communities. Consistent with this, several studies described the seminal microbiota of humans as being dominated by facultative anaerobic bacteria, usually members of *Enterococcus*, *Staphyloccus* and *Streptococcus* genera (Eggert-Kruse *et al*., 1998; Rodin *et al.*, 2003; Toth & Lesser, 1981). However, when culturing conditions were targeted towards the growth of anaerobic organisms, a much higher proportion of obligate anaerobes was isolated from the same samples that were dominated by aerobes and facultative anaerobes under different conditions (Eggert-Kruse *et al*., 1998), and obligate anaerobes, such as *Peptococus*, *Prevotella* and *Bacteriodes* species, began to be prevalent (Willeén *et al*., 1996).

Collectively, these results demonstrate both the potential for male semen to host complex microbiotas, and the limitations of studies based exclusively on *in vitro* culturing.

Investigations of the communities of bacteria living within male semen in non-human organisms often focus on the potential for the transmission of live bacteria to females through artificial inseminations when breeding farm animals e.g. bacteria was isolated from 63% of semen samples collected from pigs, *Sus scrofa*, with 43 different isolates found in total according to morphological and biochemical characteristics (Bresciani *et al*., 2014). *Campylobacter coli*, *Salmonella enteritidis*, *Pseudomona aeruginosa*, *Escherichia coli*, *Escherichia hermannii*, *Staphylococcus lentus*, *Staphylococcus sciuri*, and *Staphylococcus capitis* were identified in isolates grown from turkey, *Meleagris gallapavo*, semen (Donoghue *et al*., 2004) and Smith (1949) cultured bacteria from domestic chicken rooster, *Gallus (gallus) domesticus*, semen samples, commonly finding coliform bacilli, staphylococci, and diphtheroid bacilli. Reiber *et al*. (1995) also cultured bacteria from rooster semen samples, and found the genera isolated depended on the culturing medium used.

The focus on presence of bacteria in semen in the context of artificial semination is unsurprising given the potential for such communities to impact directly on male reproductive success; one could imagine an effect on fertility in principle via their effects on sperm and the fertilising efficiency of an ejaculate, either through the direct effects of bacterial products, or the eliciting of an immune response in either the male or the inseminated female. Indeed, there is evidence of an effect of bacteria on sperm quality, with reduced sperm motility when semen samples are inoculated with known intestinal bacteria in domestic chickens (Haines *et al*., 2013), and reduced quality in terms of motility, velocity and viability of sperm in stored ram, *Ovis aries*, semen following inoculation with pure cultures of the bacteria most frequently isolated from semen samples (Yániz *et al*., 2010). Effects on the mated female's health, through sexual transmission of bacteria and the disruption of local bacterial communities, could also affect

reproductive success of the male through an effect on female reproductive microbiota, female fitness and female reproductive responses.

Studies have used culturing approaches to explore live seminal bacteria in the semen of wild bird populations; Westneat & Rambo (2000) collected natural ejaculates from red-winged blackbirds, *Agelaius phoeniceus*, and cultured them on three different media before identifying colonies through biochemical assay profiles. They found bacterial growth from 11 of 18 cultured ejaculate samples, giving nine different types of bacteria isolated. Hupton *et al*. (2003) also collected natural ejaculates from red-winged blackbirds, and found 42% of 203 samples contained at least one bacterial species, with a maximum of ten species per sample and 37 species overall. In a study of tree swallow semen, *Tachycineta bicolor*, 62% of 30 samples cultured contained one or more bacterial species (Lombardo & Thorpe, 2000) and the presence of live bacteria was also demonstrated in the testes and the cloaca of house sparrows, *Passer domesticus*, which may contribute to the seminal microbiota of this species (Stewart & Rambo, 2000).

In fowl, the seminal microbiota is likely to be a highly transient community, formed of several components introduced leading up to and during ejaculation (Lake, 1957; Nishiyama, 1955) and potentially incorporated into the reproductive microbiota of the female once inseminated. The spermatozoa are produced in the seminiferous tubules within the testes, the tissues of which also produce various phospholipids before this mix passes into the vasa efferentia, the epididymis and then the vas deferens, the main sperm storage organ (Lake, 1957). From there it passes through the ducts deferens to the erectile ejaculatory duct and into the cloaca where, upon ejaculation, the largest amount of fluid is added from the lymphatic folds (Niyishima, 1955). Additionally, the erectile ejaculatory duct is in close proximity to the uretary ducts and rectal opening within the cloaca and so there is also opportunity for incidental incorporation from previous or simultaneous secretions from these ducts as the ejaculate passes through the cloaca. There is thus potential for the incorporation of bacterial species from communities formed in physiologically distinct habitats under different environmental conditions (Lake, 1957) and with

a variety of nutrient sources (e.g. phospholipids (Lake, 1957) complex proteins (Mohan *et al*., 1995) and glucose (Nishiyima & Fudishima, 1961)) leading to a phenotypically diverse ejaculate microbiota.

In the previous chapter of this thesis, we demonstrated the presence of bacterial DNA in red junglefowl semen, and characterised the seminal microbiota, revealing a complex community varying across individual ejaculates and male hosts. The method employed amplifies, sequences and identifies bacterial DNA present; it does not however tell us which of these organisms identified are alive in the ejaculate and are therefore viable members of the community. Here, we combine the next generation sequencing approach developed in the previous chapter with culture-based techniques to establish and explore the presence of live bacteria within the red junglefowl seminal microbiota under different environmental conditions. Specifically, each ejaculate was cultured on three different media both in aerobic and anaerobic conditions. A subset of ejaculates was also sequenced directly (i.e. before culturing), to allow for a comparison between the communities cultured under different ecological conditions and the community directly sequenced from the same semen sample. This enabled us to quantify the proportion of species directly sequenced that was captured under different *in vitro* culturing conditions, as well as identify the subsets of species favoured by different culturing conditions. We aim to: (i) test if the seminal microbiota of the red junglefowl contains live species that are culturable, (ii) establish the relationship between the cultured and directly sequenced microbiotas with reference to ecological and methodological implications, and (iii) explore the effect of different culturing conditions on the composition and structure of the cultured microbiota community.

MATERIALS & METHODS

Study Population

We studied a captive population of red junglefowl, housed in outdoor aviaries at the University of Oxford John Krebs Field Station in Wytham, Oxfordshire. This population (~200-250 adults) was established in 2006 and has been the subject of research on different aspects of

reproductive behaviour and sexual selection. For this culturing experiment, eighteen semen samples were obtained from 17 males during the 2015 and 2016 breeding seasons (three two year olds (yo), 4 three yo, 4 4 yo, 5 5 yo and one 7 yo).

Adult males are housed in pens in male-only groups of six to thirty individuals of mixed age and relatedness. Throughout the periods of sample collection, birds were routinely moved across pens and groups as a result of husbandry and logistic considerations imposed by other experimental work on this population. In addition, some males were temporarily deployed in mating trials, where they were housed in mixed-sex groups for periods of ten days, before being returned to the male-only groups.

Samples were collected opportunistically throughout the 2015 and 2016 breeding seasons whilst the birds were being sampled for other experiments, as described in Chapter 2, and thus differ in method of collection (abdominal massage or controlled copulation), time since mating, breeding history and age, though all males were sexually mature, sexually rested, and in male only groups when samples were collected. Sample collection protocol followed was consistent – as was the bird husbandry – throughout and across the two seasons.

Sample Collection and Processing

Semen samples were obtained using either of the two established techniques described in chapter two, i.e. abdominal massage and controlled copulations, and with the same collection and laboratory processing protocol described in chapter 2 to minimize environmental contamination. Briefly, abdominal massage yields semen samples that are thought to represent a large proportion of the extragonadal sperm reserves of a male, while samples obtained through controlled copulations are representative of natural ejaculates produced by males while mating with females.

Semen samples were collected throughout the 2015 and 2016 breeding season for a number of other experiments, and when samples were sufficiently large an aliquot was

taken for use in this culturing work. When samples were at least 50uL in volume, a 50uL aliquot was taken to be cultured and if the ejaculate was large enough a further 10uL was used to produce a matching directly-sequenced sample for comparison with the cultured sample. Once aliquoted as above, all samples were immediately placed on dry ice, before transfer to the Department of Zoology and storage at -80°C within the same day of collection for further processing.

Culturing

50uL of 18 semen samples, nine collected by abdominal massage and nine by controlled copulation, were diluted with 650uL sterile Phosphate Buffered Saline (PBS) to give a total volume of 700uL. The diluted sample was then mixed thoroughly by vortex and inversion to ensure homogeneity of the solution. All the semen samples were cultured in six conditions; three different solid agar media (RCM Reinforced Clostridium Medium (Thermoscientific), R2A Agar (Thermoscientific) and BD Bacto Brain Heart Infusion (Fisher Scientific) in both anaerobic and aerobic conditions. These three media represent a range of nutritional provision from low (R2A) to high (BBH) nutrient richness and are designed for selection of different bacterial species according to nutrient and growth requirements, thus were selected for this work, in combination with the two aerobic conditions, to achieve growth of different bacterial species and communities, capturing as wide a range of the seminal microbiota of each sample as possible, and allowing the investigation of community composition under different conditions.

Each sample solution was split into six, and 100uL spread onto six plates, two for each media. One of each media for every sample was then incubated aerobically and one anaerobically, both at 41°C for seven days. Anaerobic conditions were achieved using GazPak chambers and pouches. In total, eighteen 50uL semen samples were split into further aliquots and cultured across six different conditions, created by combinations of two oxygen treatments, aerobic and anaerobic, and three solid agar medium, BBH, R2A and RCM (one plate per sample per condition). A negative control of 70OuL of PBS was also split, spread on to six different media and incubated

aerobically and anaerobically in the same way as the sample aliquots. After seven days, all plates were photographed in order to record colony numbers and morphology and then the bacteria harvested. 2mL sterile PBS was applied to the medium and a sterilised metal paperclip spun across the surface using a magnetic stand to lift the colonies from the agar. The suspended bacteria were then pipetted into a 15mL sterile tube, mixed thoroughly by vortex and a 10uL aliquot added to 200uL pre-prepared lysis buffer mix (4uL Tric-HCl 1M, 0.8uL EDTA 0.5M, 2.4ul Triton X-100, 192.8uL dH20 and 4mg of lysozyme) and stored at -20°C until further processing.

DNA Extraction, PCR Amplification and Illumina Miseq Sequencing from Culturing Samples and Matching Ejaculate Aliquots

Information about community composition, in terms of bacterial OTUs detected, was generated from cultured samples and ejaculate aliquots via the Illumina MiSeq Platform using the pipeline developed and described in chapter 2. Briefly DNA was extracted from the aliquots using the DNeasy Blood and Tissue Kit (QIAGEN) according to manufacturer's guidelines and adapted to optimize extraction of bacterial DNA. The V3 and V4 regions of the bacterial 16S encoding genes were targeted and amplified using Phusion High-Fidelity DNA Polymerase (New England Biolabs). Reverse bacterial primers 347F and 803R, and the forward primer, 347F, had a barcode sequence attached carried out according to the Phusion polymerase manufacturer's protocol and with reference to the primer melting temperatures. Again, careful protocols were followed to minimise contamination both from the DNA product of previous amplifications with the same target, environmental contamination from lab space, equipment and reagents as well as cross-contamination between aliquots (see chapter 2).

Once amplification of the target was confirmed by gel electrophoresis and visualization under UV light, the PCR products were pooled into DNA libraries, purified using QIAGEN MiniElute PCR Purification Kit and prepared for sequencing on the Illlumina MiSeq platform, with attachment of the Illumina barcoded index primer Illumina adaptor using the NEBNext Ultra DNA Library Prep Kit before submission to the sequencing facility. Sequence reads resulting from

Illumina MiSeq were analysed following the protocol outlined in chapter 2, using custom made scripts in R and the QIIME pipeline, and with the same measures to address whole dataset and per-MiSeq contaminants, remove samples of insufficient coverage and to achieve equal read number. All subsequent analysis was performed in the R software version 3.3.2 (R Core Team, 2016).

Duplicated Samples and Potential 'Outlier' Samples

Of the cultured aliquots successfully sequenced, two were duplicates of one another, i.e. sequences produced by two different PCR amplification reactions using DNA extracted from the same aliquot, and three aliquots were potential outliers as they had up to 55 more Operational Taxonomic Units (OTUs) than the next most species rich samples (i.e. 78, 71 and 57 compared with 33). All analyses investigating culturing conditions were performed with each of the two duplicates, and including and excluding the three outliers. The results of these analyses on the effect of culturing conditions were qualitatively unaffected (Appendix 1 and 2). Therefore, results presented for all the analyses in this chapter include the three 'outlier' aliquots and the same one of the two duplicates chosen at random. Details of these 'outlier' aliquots and the 'extra' OTUs that they contain (i.e. those OTUs unique to one or more of the outliers) can be found in Appendix 2.

Live Bacteria and Characterisation of Cultured Communities

The presence of live, culturable bacteria in a sample was confirmed through the growth of colonies on culture plates, which were photographed at the end of the incubation period prior to harvesting (note that the lack of colonies does not necessarily mean that a sample lacks live bacteria, but rather that the bacteria living in the sample might not grow under certain culturing conditions). Cultured communities were characterised through the collation and analysis of the cleaned and rarefied sequencing output of each aliquot as outlined in chapter two. The phyla detected in all aliquots were described in terms of the number of OTUs detected, the average relative abundance per aliquot of each phylum and the proportion of total sequence reads of each phylum in all aliquots.

Comparison to Directly-Sequenced Seminal and Faecal Samples

We investigated the extent to which the microbiota of cultured aliquots represented the microbiota of directly-sequenced (i.e. not cultured) aliquots in three different ways. First, these results were compared to the 171 directly-sequenced semen samples and 26 directly sequenced faecal samples analysed in chapter two. The community structure 'shape' of culture aliquots was compared to that of semen aliquots, by plotting the number of culture aliquots an OTU occurs in against the average relative abundance at which it occurs. The phyla found in cultured aliquots of samples was compared to those found in faecal and semen samples, by their presence and absence in each of the three sample types, the order of the phyla according to their average per aliquot relative abundance in each sample type, and the order of the phyla according to the proportion of overall sequence reads in all aliquots of a type that they contribute. The OTUs found in cultured aliquots were also compared to faecal samples by calculating the proportion of OTUs in a cultured sample aliquot that were also found in faecal samples.

Eight of the semen samples collected in this study were sufficiently large to allow a further aliquot to be taken and directly sequenced in parallel to the cultured aliquot. Thus, for each of these eight samples, cultured communities could also be compared to the directly-sequenced community obtained from the same sample before culturing.'

The proportion of OTUs in a cultured sample aliquot that were never found in any of the directly-sequenced sample aliquots (i.e. were culture unique OTUs) was calculated, as well as the proportion of sequence reads that were assigned to these OTUs. Third, the proportion of OTUs in a cultured sample aliquot that were not found in the matching directly-sequenced aliquot (i.e. from the same semen sample) was also calculated, and the proportion of sequence reads that were assigned to those OTUs. The proportion of OTUs in a directly-sequenced aliquot that also occurred in the matching cultured aliquots was calculated.

The Effect of Culturing Condition

Community Metrics

We measured species richness (i.e. number of OTUs), Shannon diversity and Simpson diversity for each cultured aliquot using the R package 'vegan'. In addition, a Principal Component Analysis (PCA) was performed to give axes of variation across all samples using the standard 'stats' R package function 'prcomp', the first two PCA axes explaining sufficient variation to provide composition metrics for each sample. To eliminate the effect of infrequently observed OTUs that could be contributing to variation in PCA scores between individuals but with insufficient information on their abundance, the PCA analysis was performed using a subset of 25 OTUs, which appear in at least 10% of aliquots and at an average relative abundance equal to or greater than 1%. This produced five community metrics (species richness, Shannon diversity, Simpson diversity, PC1, and PC2). To test whether variation in each of these five community metrics was explained by medium, oxygen treatment, or an interaction between them, linear mixed models were constructed including medium and oxygen treatment as fixed factors with an interaction, and the semen sample of origin as a random factor. Simpson diversity was arc-sine transformed, it being a proportion. The box cox function in R package 'MASS' was then used to determine the transformation lambda value required to ensure data for the five community composition measures met the assumptions of the linear mixed model. In order to test the assumptions of normality of error distribution and of homogeneity of variance, I used diagnostic plots: frequency distribution histograms, q-q plots and plots of residual vs. fitted values, which were generated using the basic R packages 'stats' and 'graphics''

To establish the effect of each of the two fixed factors and their interaction, the full model was simplified, running less parametrised models (Table 1) using the lme function in the 'lme4' package and comparing models using the ANOVA function in the standard 'stats' R package. A p-value generated by the function using log-likelihood ratios from two models with and without the focal model component was utilised to estimate the significance of the extra variation explained by each model component. The effects of the levels of each significant treatment were then

determined through post-hoc comparisons using the lsmeans function in the 'lmerTest' package and inputting the linear mixed model including all significant terms.

Table 1 *Structure of the five models compared to establish the contribution of the different culturing treatments to variation in response variables across cultured aliquots*

Model 0 (NULL)	Response Variable ~ 1 (Random = Bird ID)
Model 1	Response Variable ~ Medium * Oxygen Treatment (Random = Bird ID)
Model 2	Response Variable ~ Medium + Oxygen Treatment (Random = Bird ID)
Model 3	Response Variable ~ Oxygen Treatment (Random = Bird ID)
Model 4	Response Variable ~ Medium (Random = Bird ID)

To establish whether there are groups of OTUs which occur in the same aliquots at comparable relative abundances, and whether these groups of OTUs explain differentiation of aliquots along the Principal Component axes, we performed hierarchical clustering using the complete linkage method on Euclidean distances between OTU occurrence profiles (i.e. the relative abundance of OTUs in all the aliquots) and visualised these clusters using a dendrogram. We also calculated Spearman's correlations between these profiles, and visualised them using a heatmap, with colour intensity representing the strength of a correlation. Clusters of co-occurring OTUs that may be of relevance were then identified by branching of the dendrograms, and clusters of more intense colour on the heatmap. Using all aliquots, we then calculated the correlation of the relative abundance of OTUs within each cluster with sample PCA score.

Phylum Relative Abundance

The effect of culturing condition on the relative abundance of OTUs from the nine different phyla detected was investigated using the same method and models as for community metrics, with the response variable now the relative abundance of each phylum. P-values were adjusted using the 'fdr' method of the p.adjust function within the 'stats' R package applied to each set of p-values (i.e. the set of p-values for each model component, one from each phyla) to correct for multiple comparisons.

A relationship between phyla affected by culturing condition and scores on the PCA axes, i.e. the variation in community composition across all aliquots, was explored graphically by i)

plotting the relative abundance of a phylum on to the 2D PCA space created by plotting PCA Axis 1 against PCA Axis 2 and ii) plotting the relative abundance of each phylum in an aliquot against the score of that aliquot along PCA Axis 1 and PCA Axis 2 individually.

Presence of Individual OTUs

To investigate the effect of culturing condition on the presence or absence of individual OTUs, the same approach was used, using general linear mixed models based on a binomial error distribution, and only those 25 OTUs used for the PCA. The model was repeated for each of these OTUs, with presence of the OTU being the response variable. P-values were adjusted using the 'fdr' method of the p.adjust function within the 'stats' R package applied to each set of p-values (i.e. the set of p-values for each model component, one from each OTU) to correct for multiple comparisons.

Proportion of 'Faecal' OTUs

Because semen is expelled from the papillae at the end of the vas deferens into the urodeum and proctodeum (Etches, 1996), faecal OTUs are likely to represent a physiological component of natural ejaculates in birds. The results of chapter 2 confirm that the seminal microbiota includes OTUs also found in faecal samples. It is plausible that faecal microbiota may have different ecological requirements than other seminal OTUs (Javurek *et al*., 2016). To test whether patterns of microbiota structure under different culturing conditions were affected by faecal OTUs, faecal OTUs were removed from all aliquots, and community metrics recalculated and visualised graphically. More formal statistical analyses were not performed as the removal of faecal OTUs left a maximum of six OTUs per aliquot, which we deemed insufficient for a meaningful analysis. The effect of culturing condition on proportion of OTUs in a sample that were also found in faecal samples was investigated using the same linear mixed model method as above. Both proportion of OTUs, and proportion of sequences assigned to faecal OTUs were used as response variables, and box cox transformed as previously described.

Proportion of 'Culture Only' OTUs

The effect of culturing condition on the five previously described measures comparing cultured and directly sequenced aliquots (proportion of overall culture only OTUs and sequence reads, proportion of matching culture only OTUs and sequence reads and proportion of directly sequenced OTUs shared with matching culture aliquots) was analysed using the same method as the previous analyses i.e. inclusion as the response variable in a linear mixed model, and box cox transformed as previously described.

RESULTS

1. Viable Bacteria and Characterisation of Cultured Communities

Bacterial growth was observed in all samples and across all treatments, with no growth on the negative control plates on which sterile PBS had been inoculated. Variation between samples and treatments was visible in the extent of growth and colony morphology, as demonstrated in Fig. 1.

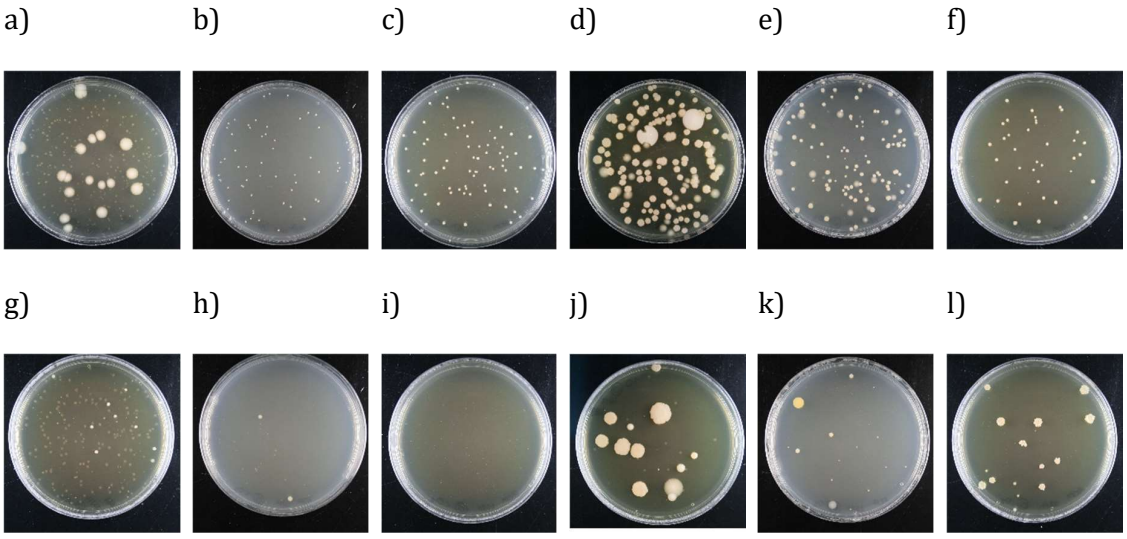


Figure 1 *Photographs of bacterial colonies on agar plates inoculated with semen samples, after 7 days of incubation. a) – f) are incoculated with semen A, g) – l) with semen C. a), b),c), g), h), i) were incubated anaerobically, d), e), f), j), k), l) were incubated aerobically. a),g),d)j) grown on BBH medium, b),h),e),k) grown on R2A medium and c),f),i),l) grown on RCM medium.*

From 98 successfully sequenced samples, a total of 217 OTUs were isolated, with between six and 78 OTUs per aliquot. Overall, aliquots were collectively dominated by species from the

phylum *Firmicutes*, accounting for over 80% of all sequence reads. The phylum *Firmicutes* was also the most abundant in each individual aliquot with an average relative abundance of 0.875% +- 0.022% (Table 2). Therefore there was an order of magnitude difference between per aliquot proportion of total reads and the average relative abundance of this phylum, which likely indicates variation in per aliquot relative abundance, either between the relative abundance of the same OTU between aliquots or between different OTUs within each phylum.

Table 2 *Average per aliquot relative abundance and proportion of overall total reads attributed to OTUs belonging to each of the nine phyla detected in cultured aliquots, as well as the number of OTUs detected for each phylum. The phyla are listed from greatest to lowest average relative abundance. Figures are not reduced to 3 s.f. in order to distinguish between abundance and read proportion of the phyla.*

Phylum	Average Relative OTU Abundance	Proportion of Total Reads	No. OTUs
Firmicutes	0.00875	0.82780	94
Actinobacteria	0.00324	0.16623	53
Proteobacteria	0.00009	0.00491	56
Bacteroidetes	0.00006	0.00047	8
TM7	0.00035	0.00035	1
Other	0.00010	0.00010	1
Fusobacteria	0.00008	0.00008	1
Cyanobacteria	0.00002	0.00004	2
Synergistetes	0.00002	0.00002	1

2. Comparison to Directly Sequenced Semen Aliquots

The four phyla with the highest per aliquot relative abundance found in the 98 cultured aliquots were the same as those found in the 177 directly-sequenced samples analysed in chapter 2. The same is also true for the proportion of total overall sequence reads, though *Firmicutes* makes up an overall lower proportion in directly-sequenced samples (47.6% of sequences compared with over 80%). The 'shape' of the cultured microbiotas is similar to that found in the microbiotas from directly-sequenced samples in chapter 2, in that both show a positive relationship between frequency of an OTU (the number of samples/aliquots in which it occurs) and its average relative abundance (Fig. 2a and b), and that the majority of OTUs are rare, occuring at very low frequency (Fig. 2c and 2d) and low relative abundance (Fig. 2e and 2f).

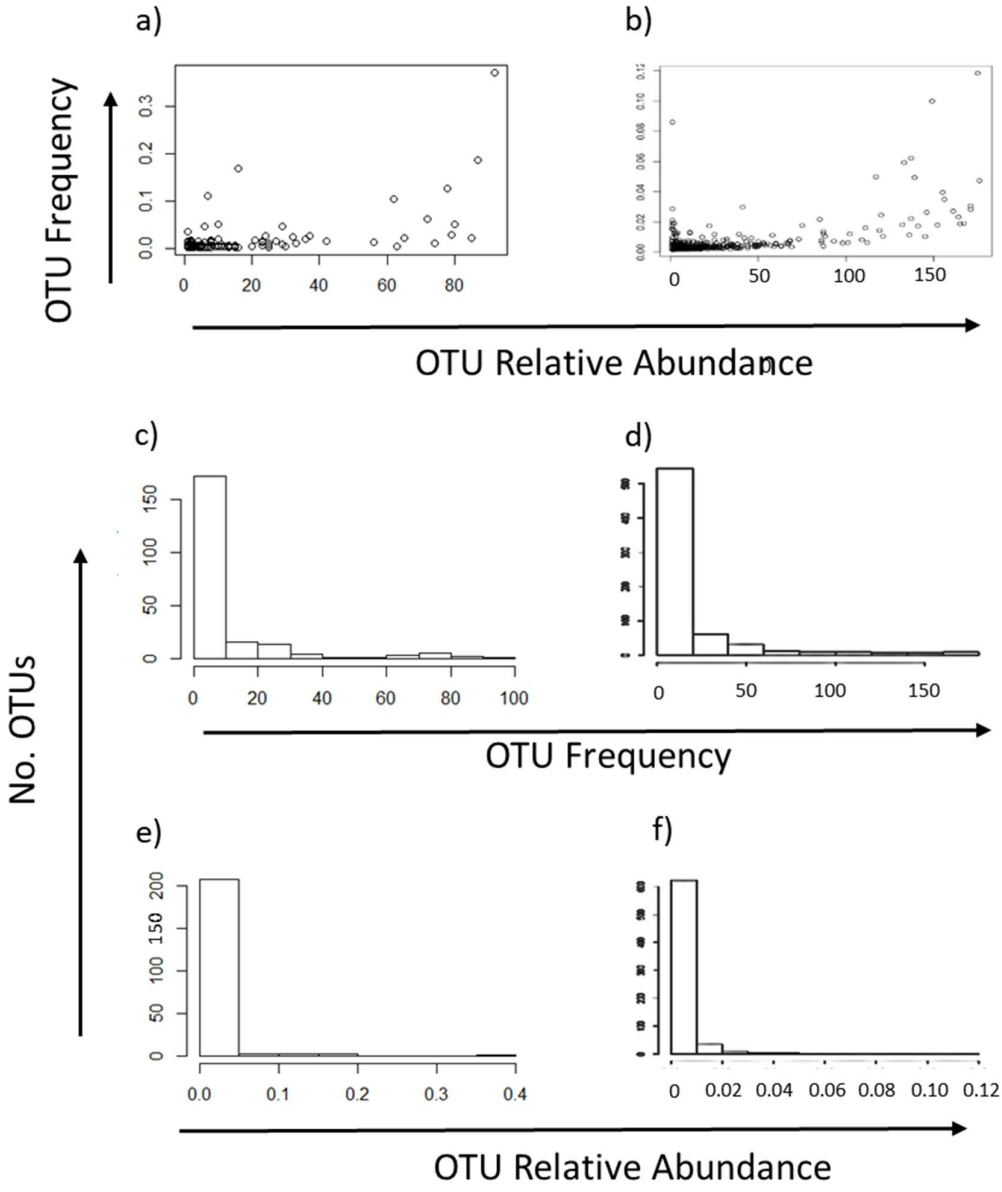


Figure 2 *Plots comparing the communities of i) cultured sample aliquots and ii) chapter 2's directly sequenced samples. a) Average relative abundance plotted against frequency (i.e. number of aliquots/samples in which an OTU occurs) for each OTU b) Histograms of OTU frequency c) Histograms of OTU average relative abundance.*

We then compared the cultured sample aliquots with the directly sequenced aliquots of eight semen samples, which were directly-sequenced as well as being cultured *in vitro*. Out of 292 total OTUs found overall, 111 were found in both directly-sequenced and cultured aliquots, 106 were only found in cultured aliquots and 75 only in directly-sequenced aliquots. Overall, 60% of the 186 directly-sequenced OTUs were found in cultured aliquots. However, 49% of the 217

cultured OTUs were not found in directly-sequenced aliquots; on average about a third of each aliquot. These make up on average 13% of sequence reads per cultured aliquot. We refer to these OTUS as 'overall culture-only OTUs'.

When comparing each of the the eight directly-sequenced sample aliquots to the matching cultured sample aliquots from the same semen sample (between four and six cultured aliquots for each semen as not all cultured aliquots were successfully processed in the laboratory and post-sequencing), on average 19% of OTUs sequenced directly from a semen sample were found each of the matching cultured aliquots. However, on average 67.8% (min = 30.8%, max = 92.9%) of OTUs per cultured aliquot were not found in the matching directly-sequenced aliquot, making up an average of 61.3% (min = 4.40%, max = 99.6%) of sequence reads per cultured aliquot that were not found in the matching directly-sequenced aliquot. We refer to these OTUS as 'matching culture-only OTUs'.

One possible reason to explain why many OTUs detected in the cultured aliquots were not detected in the matching directly-sequenced aliquot of the same semen sample is that these OTUs were in fact detected in directly-sequenced aliquots but were removed during the rarefaction process due to their low abundance. To test this idea we generated a list of OTUs lost during rarefaction from directly-sequenced aliquots. Of 323 OTUs lost during rarefaction from the eight directly-sequenced aliquots, 66 were retained in culture aliquots. These 66 OTUs represent 62.3% of the 106 'overall culture-only OTUs', indicating that loss through rarefaction can explain over half of the OTUs detected in some cultured aliquots but not in any directly-sequenced samples. Additionally, 'overall culture-only OTUs' found in each treatment are different; not all 106 are found in every condition but there are between 31 (in Anaerobic R2A samples) and 49 (Anaerobic BBH samples) OTUs found in the aliquots under each condition.

Comparison to Directly-Sequenced Faecal Samples

Compared to culturing aliquots, the 26 directly-sequenced faecal samples in Chapter 2, had 12 more bacterial phyla (21 phyla overall). The same four phyla remain the highest by overall

proportional abundance but differ by per sample relative abundance, and there is a wider spread of average per aliquot relative abundance across the different phyla in cultured aliquots. However more than half of the sequence reads in culturing samples are still from the *Firmicutes* phylum, though TM7 has the highest per aliquot average relative abundance and *Proteobacteria* contributes the highest number of OTUs. The majority (77%) of OTUs found in cultured aliquots were also found in the 26 faecal samples analysed in chapter 2, similar to the findings of chapter 2, whereas 51% of OTUs from 175 semen samples were also found in faecal samples.

3. The Effect of Culturing Condition

Alpha Diversity

The structure of the cultured microbial communities differed across culturing conditions, according to all three measures.

Variation in species richness across cultured aliquots was explained by an interaction between oxygen treatment and medium (Model 1 v Model 2, ChiSq = 7.20, df = 2, $p = 0.027$). R2A aliquots were significantly more species-rich when incubated under anaerobic conditions compared to aerobic conditions (t ratio = -3.51, $p < 0.001$). Thus, R2A aliquots consequently had a significantly lower species richness than RCM aliquots in the aerobic condition (t ratio = -2.47, $p=0.041$), and a significantly higher species richness than BBH aliquots in the anaerobic condition (t ratio = -3.35, $p = 0.004$). Conversely, BBH aliquots were significantly less species-rich than RCM aliquots in aerobic conditions (t ratio = -3.19, $p = 0.006$), but not in anaerobic conditions (Fig. 3a).

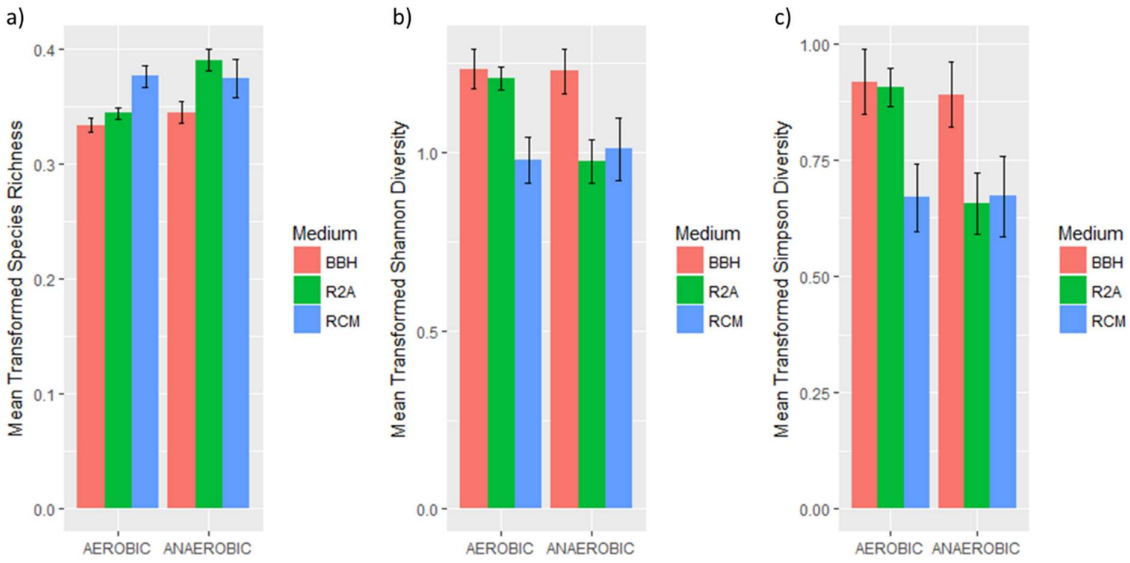


Figure 3 *Figure illustrating the relationships between a) species richness b) Shannon diversity and c) Simpson diversity and the two culturing treatments, oxygen treatment and medium. Values shown are means across the aliquots in that condition, and bars indicate standard errors. Means and standard errors are calculated from data transformed according to the criteria described in methods.*

A similar pattern was observed for Shannon diversity, where a significant portion of the variance was explained by a) an interaction between oxygen treatment and medium (Model 1 v Model 2, ChiSq = 7.05, df = 2, p = 0.029). R2A aliquots had a significantly lower Shannon diversity when incubated in anaerobic compared to aerobic conditions (t ratio = 3.01, p = 0.004). R2A aliquots had a significantly higher Shannon diversity than RCM aliquots in aerobic conditions (t ratio = 2.98, p = 0.011), and a significantly lower Shannon diversity than BBH aliquots in anaerobic conditions. (t ratio = 3.17, p = 0.006). BBH aliquots had a higher Shannon diversity than RCM aliquots in both the aerobic (t ratio = 3.05, p = 0.009) and anaerobic conditions (t ratio = 2.69, p = 0.023) (Fig. 3b).

Variation in Simpson diversity was explained by medium only (model 4 v model 0, ChiSq = 11.5, df = 2, p = 0.003). This effect was largely due BBH samples having a significantly greater Simpson diversity than RCM (t ratio = 3.44, p = 0.003) (Figure 3c).

PCA Analysis

A Principal Components Analysis was performed on a subset of 25 OTUs, selected using the criteria of occurrence in 10% of aliquots and 1% average relative abundance in those aliquots

in which they occur. The first axis (PC1) produced by the PCA explained 12% of the variation in aliquot composition and the second (PC2) explained 9% of this variation. PC1 and PC2 captured variation in cultured microbiotas due to the presence of oxygen and the growth medium, respectively (Fig. 4).

PC1 axis was explained by oxygen treatment (i.e. aerobic *vs*. anaerobic conditions; Model 3 v Model 0, ChiSq = 33.7, df = 1, $p < 0.001$), aliquots cultured under aerobic conditions had lower PC1 scores than those cultured under anaerobic conditions (t ratio = -6.40, $p < 0.001$), regardless of growth medium (Fig. 4a and b). Variation in PC2 on the other hand, was explained by medium (Model 4 v Model 0, ChiSq = 7.99, df = 2, $p = 0.018$), with aliquots cultured in RCM having higher PC2 scores than those cultured in R2A (t ratio = 2.59, $p = 0.031$), regardless of oxygen treatment (Fig. 4c and d).

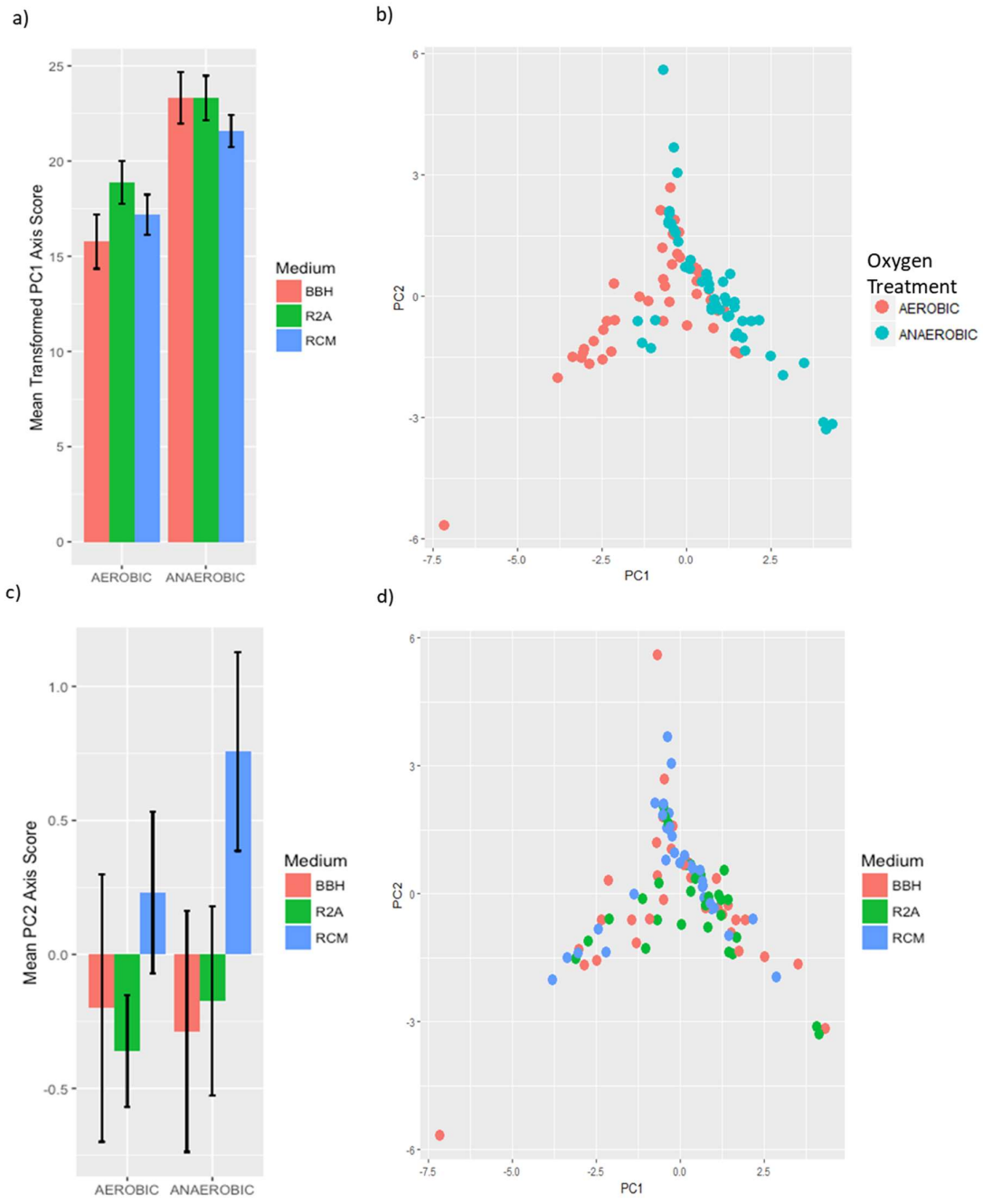


Figure 4 *The first two axes of the PCA by culturing condition. In a) and c) values shown are means of pC1 and PC2 across the aliquots cultured under that condition, and means and standard errors are calculated from data transformed (or not transformed) according to the criteria described in Methods. Plots b) and d) show the position of aliquots within the PCA space created using raw values of PC1 and PC2, and colours indicate treatment.*

OTU Occurrence Profiles

Hierarchical clustering of OTUs and spearman correlations of OTU relative abundances in each aliquot revealed three main clusters, as shown in Fig. 5, judged visually using the branching pattern and colour intensity.

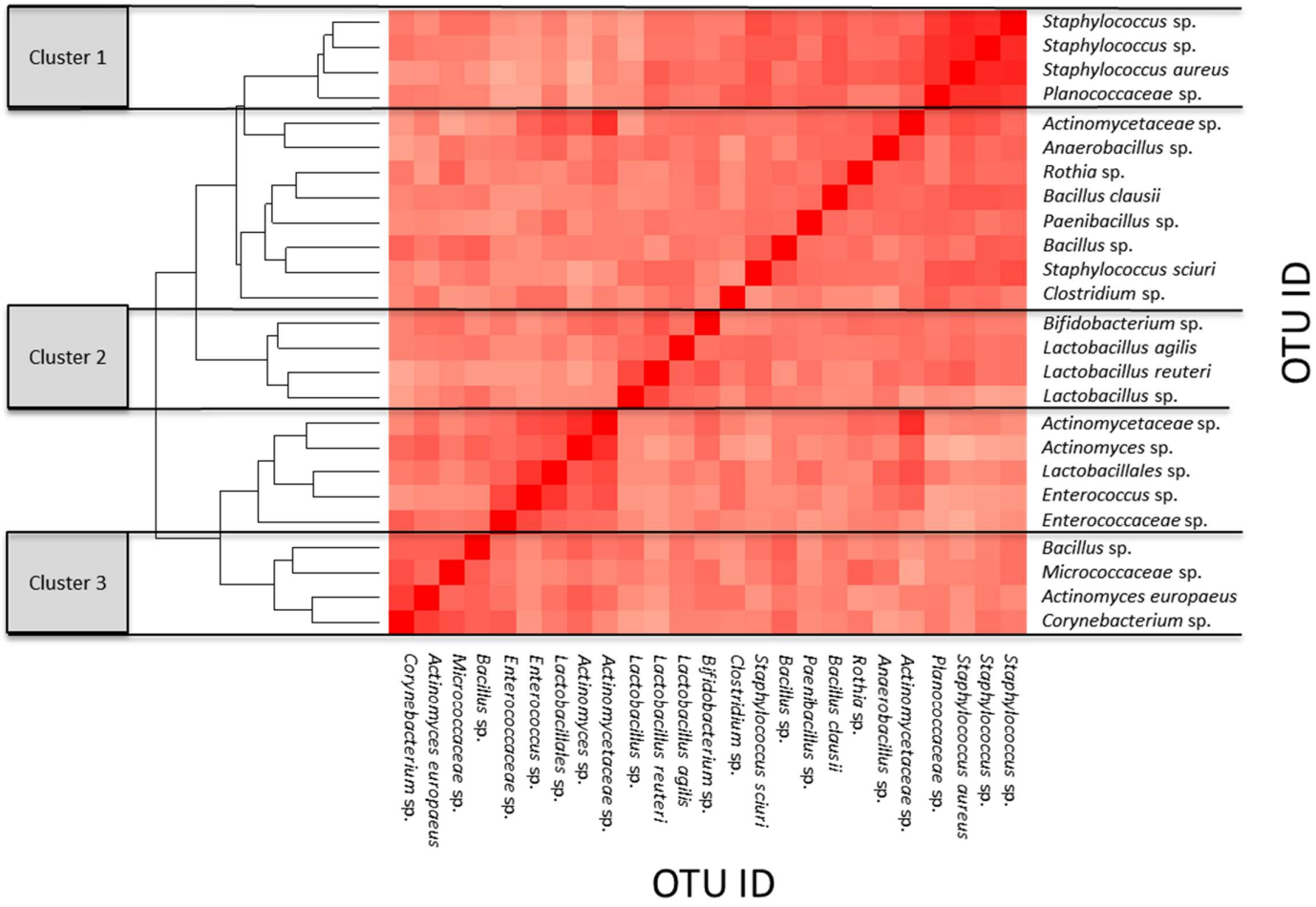


Figure 5 *Dendrogram illustrating clustering of OTUs according to co-occurrence in aliquots at similar relative levels of abundance, and a heatmap illustrating Spearman Correlation of relative abundance of OTUs between aliquots. The three OTU clusters identified through colour intensity and branching are indicated by boxes on RHS, next to the identity of OTUs. The X-axis of the heatmap is not labelled but from left to right follows the same order of OTUs as bottom to top on the Y-axis.*

Cluster 2 groups OTUs whose average per aliquot relative abundances are correlated negatively with sample scores on the PC1 axis, and comprises three *Staphyloccus* species and a *Planacoccaceae* species. Cluster 2 groups OTUs whose average per aliquot relative abundances are correlated positively with aliquot scores on the PC1 axis, and comprises three *Lactobacillus* species and one *Bifidobacterium* species. Cluster 3 groups OTUs whose average per aliquot relative abundances are correlated negatively with aliquot scores on the PC1 axis and comprises a *Bacillus* species, a *Micrococcaceae* species, *Actinomyces europaeus* and *Corynebacterium* species.

Phylum Relative Abundance

To investigate the effect of culturing treatment on the relative abundance of OTUs belonging to each of the nine phyla identified, the average relative abundances of OTUs from each phylum were used as response variables in the models outlined in Table 1. This analysis revealed phylum-specific variation in performance across different culturing conditions (Appendix 3). Variation in three phyla was signifficantly better explained by models including oxygen treatment than the null model (Table 3). The average relative abundance of OTUs belonging to the phylum Actinobacteria was greater in aerobic aliquots, but the average relative abundance of OTUs belonging to the Firmicutes phylum was lower in anaerobic samples (Fig. 6). We do not consider 'Other' any further as this includes OTUs from different phyla who could not be assigned to other phyla, so is uninformative here.

Table 3 *Summary of the outcomes of model comparison analysis to investigate the effect of culturing treatment on the relative abundance of OTUs belonging to each phylum. Only the three phyla affected by culturing treatment are shown, and as only oxygen treatment was shown to effect phylum abundance so the second column shows values relating to comparison of the model including oxygen treatment to the null model, and the third column gives the values for post-hoc comparison of oxygen treatment means.*

Phyla	Model 3 v Model 0	Aerobic v Anaerobic
Firmicutes	ChiSq = 10.6 df = 1 p<0.001	t ratio = -3.32 p = 0.001
Actinobacteria	ChiSq = 11.9 df = 1 p = 0.001	t ratio = 3.53 p < 0.001
Other	ChiSq = 5.74 df = 1 p = 0.017	t ratio = 2.41, p = 0.018

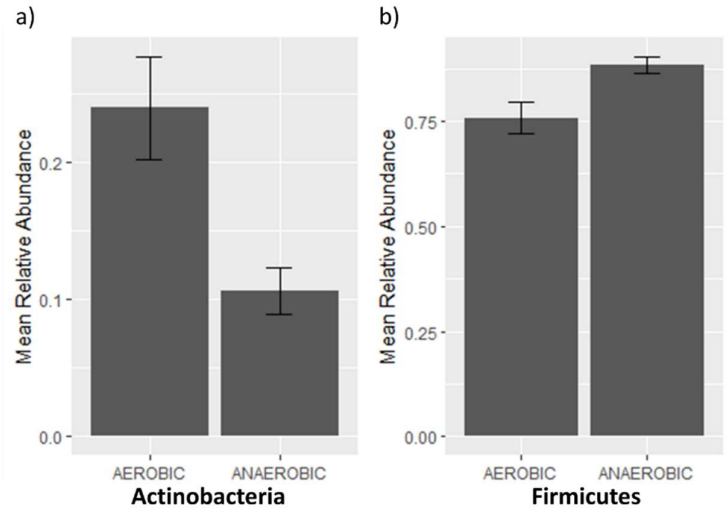


Figure 6 *Mean per aliquot relative abundance of OTUs by oxygen treatment for two phyla which were shown to vary with oxygen treatment. Error bars indicate standard errors.*

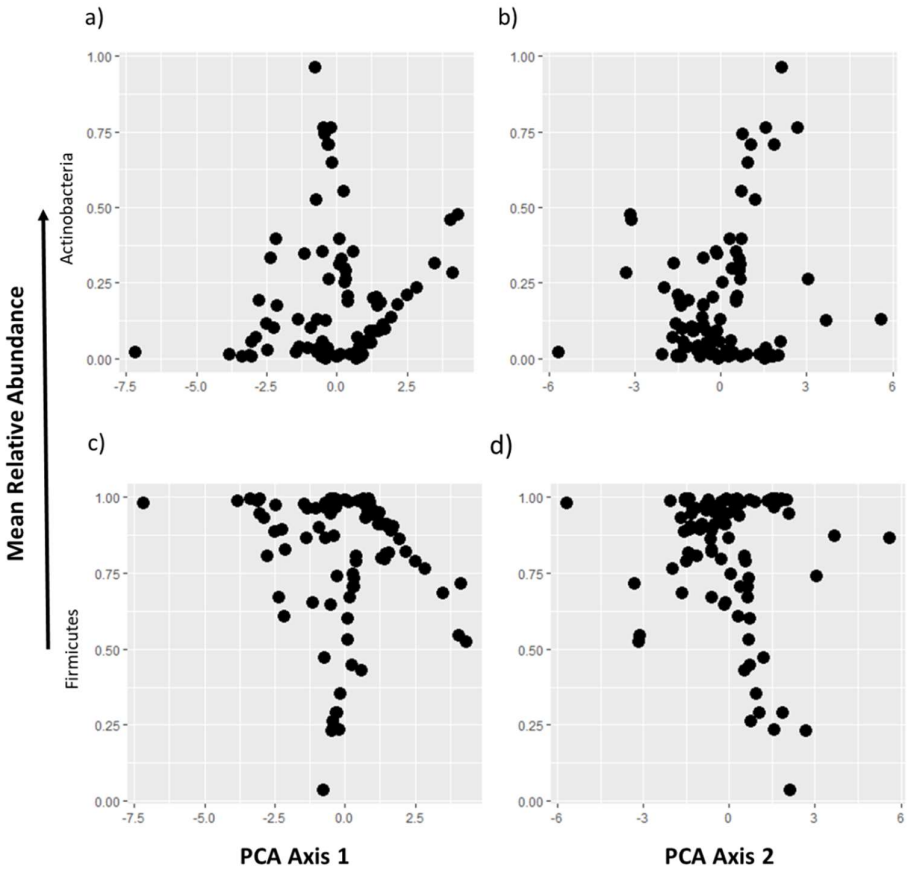


Figure 7 *Mean per aliquot relative abundance of OTUs plotted against aliquot score on PCA Axis 1 (a and c) and PCA Axis 2 (b and d) for two phyla, Actinobacteria (a and b) and Firmicutes (c and d). Values for PCA axis scores are untransformed.*

We followed up these results by visually investigating the relationship between the relative abundance of OTUs belonging to the *Actinobacteria* and *Firmicutes* phyla in an aliquot and scores of an aliquot on the first two PCA axes. At PCA axis 1 and 2 scores below zero there does not seem to be a very obvious relationship with average *Actinobacteria* OTU relative abundance. However at PCA scores greater than zero there appears to be an increase in relative abundance of actinobacteria scores. The reverse is true for *Firmicutes* OTUs, where relative abundance decreases with PCA 1 and 2 scores above zero (Fig. 7).

Presence of Individual OTUs

To investigate the relationship between culturing treatment and the presence of specific OTUs, the presence in an aliquot of each of the 25 OTUs used for the PCA earlier in the chapter was used as a response variable in the models outlined in Table 1. These analyses showed that variation in the presence of eight of these OTUs was explained by culturing treatment (Table 4). Briefly, both medium and oxygen treatment significantly explained occurrence of an *Enterococcus* species (Model 2 v 3 ChiSq = 15.2, df = 2, p=0.006; Model 2 v 4 ChiSq = 14.6, df = 1, p=0.002, Fig. 8a). *Bacillus* species, three *Staphylococcus* species and a species from the *Planococcaceae* family were all significantly more likely to be present in aerobic rather than anaerobic aliquots (Fig. 8b to 8f, Table 4a). Medium explained presence of *Actinomyces europaeus* and an OTU from the *Micrococcaceae* family; the micrococcaceae OTU occurred in significantly fewer R2A aliquots than either BBH or RCM, and *Actinomyces europaeus* was more likely to occur in BBH aliquots than R2A or RCM (Fig. 8g and h, Table 4b).

Table 4 *Summary of the outcomes of model comparison analysis to investigate the effect of culturing treatment on the presence of subset OTUs illustrating a) the OTUs affected by oxygen treatment are shown. The second column shows values relating to comparison of the model including oxygen treatment to the null model, and the third column gives the values for post-hoc comparison of oxygen treatment b) the OTUs affected by medium are shown. The second column shows likelihood ratios (lr) and p-values relating to comparison of the model including medium to the null model, and the third column gives the z ratios and p-values for post-hoc comparisons of the three media.*

a)

OTU	Model 3 v Model 0	Aerobic v Anaerobic
Bacillus sp.	ChiSq = 7.23 df = 1 p=0.045	z ratio = 2.62 p = 0.009
Staphylococcus sp.	ChiSq = 12.3 df = 1 p = 0.012	z ratio = 3.13 p = 0.002
Staphylococcus sciuri	ChiSq = 6.31 df = 1 p = 0.060	z ratio = 2.36 p = 0.018
Staphylococcus sp.	ChiSq = 7.53 df = 1 p = 0.045	z ratio = 2.58 p = 0.010
Planacoccaceae sp.	ChiSq = 7.48 df = 1 p = 0.045	z ratio = 359 p = 0

b)

OTU	Model 4 v Model 0	BBH v R2A	BBH v RCM	R2A v RCM
Micrococcaceae sp.	ChiSq = 20.3 df = 2 p = 0.001	z ratio = 3.10 p = 0.005	z ratio = 0.213, p = 0.975	z ratio = -2.97 p = 0.008
	Model 2 v Model 3			
Actinomyces europaeus	ChiSq = 11.1 df =2 p = 0.032	z ratio = 2.44 p = 0.04	z ratio = 2.65 p = 0.02	z ratio = 0.226, p = 0.972

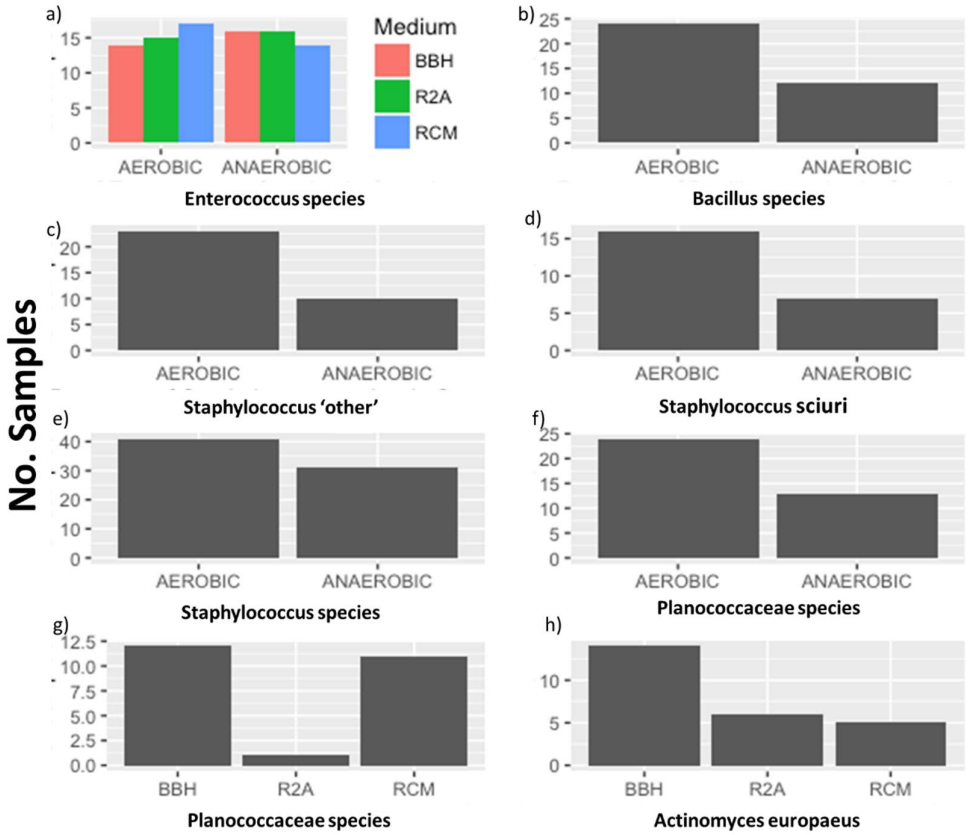


Figure 8 *Presence of OTUs in an aliquot by culturing condition for those eight OTUs the presence of which is associated with culturing treatment. For Enterococcus species both medium and oxygen treatment are shown as both contributed to presence of this OTU.*

Proportion of 'Faecal' OTUs

Both medium and oxygen treatment but not their interaction explained variation in the proportion of faecal OTUS detected in an aliquot (Model 2 v Model 3, ChiSq = 7.58, df = 2, p = 0.023; Model 2 v Model 4, ChiSq = 6.65, df = 1, p = 0.010). This was because BBH aliquots had a lower proportion of faecal OTUs than R2A aliquots (t ratio = -2.80, p = 0.017) and aerobic aliquots had a lower proportion than anaerobic aliquots (t ratio = -2.35, p = 0.021) (Fig. 9a). The same pattern was observed when looking at proportion of sequence reads (Fig. 9b) (Model 2 v Model 3, ChiSq = 8.22, df = 2, p = 0.016; Model 2 v Model 4, ChiSq = 14.2, df = 1, p < 0.001). BBH had a lower proportion of sequence reads of faecal OTUs than R2A (t ratio = -2.71, p = 0.022), and aerobic aliquots a lower proportion than anaerobic (t ratio = -3.86, p < 0.001).

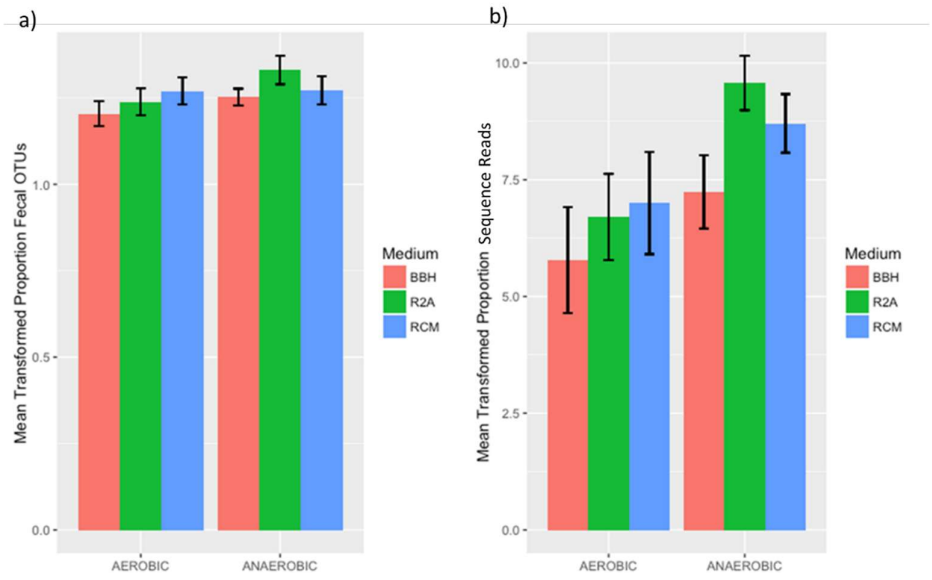


Figure 9 *Mean per aliquot proportion of a) faecal OTUs b) sequences assigned to faecal OTUs by culturing condition. Values shown are mean proportions across the samples in that condition and error bars are standard errors, and means and standard errors are calculated from transformed data.*

Proportion of 'Overall Culture-Only' OTUs

No aspect of the culturing environment explained the proportion of 'overall culture-only' OTUs found in a cultured aliquot (Fig. 10a). However, culturing medium was associated with the proportion of sequence reads of 'overall culture-only' OTUS in an aliquot (model 4 v model 0, ChiSq = 6.53, df = 2, p = 0.038), with RCM aliquots having a much higher proportion of these sequence reads than BBH aliquots (t ratio = -2.34, p = 0.056) (Fig. 10b).

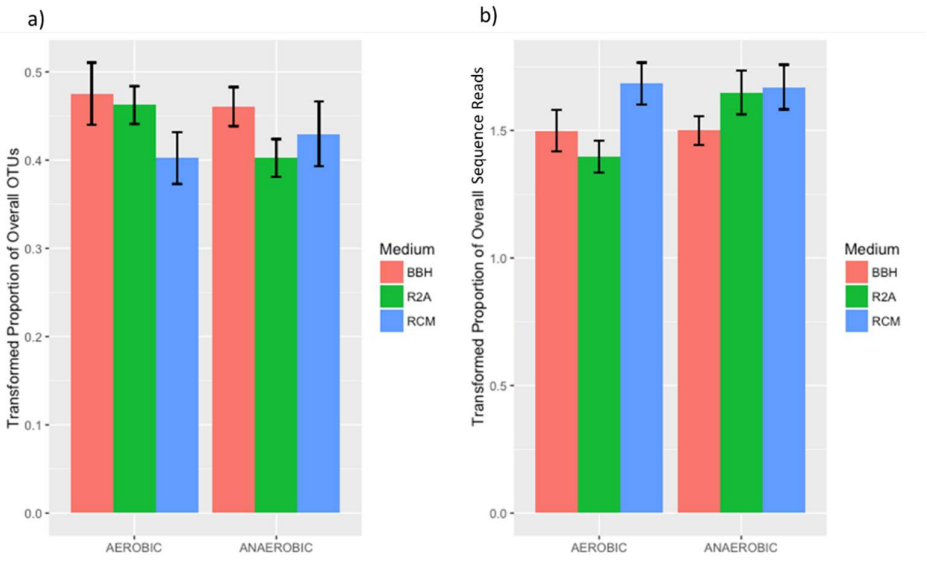


Figure 10 *Mean proportion of a) overall culture only OTUs b) sequence reads assigned to overall culture only OTUs by culturing condition. Values shown are mean proportions across the aliquots in that condition, and error bars show standard errors. Means and standard errors are calculated from transformed data.*

When comparing each of the 8 directly-sequenced aliquots to the matching cultured aliquots from the same ejaculate, on average 19% of OTUs sequenced directly from an ejaculate were found in the matching cultured aliquot. However, there was an average of 67.8% (min = 30.8%, max = 92.9%) of OTUs per cultured aliquot that were not found in the matching directly-sequenced aliquot, making up an average of 61.3% (min = 4.40%, max = 99.6%) of sequence reads per cultured aliquot. Culturing conditions did not explain variation in either proportion of 'matching culture-only' OTUs in cultured aliquots or the proportion of sequence reads from these OTUs (Fig. 11). The same was true for the proportion of OTUs in directly-sequenced aliquots which were also found in the matching cultured aliquots (Fig. 12).

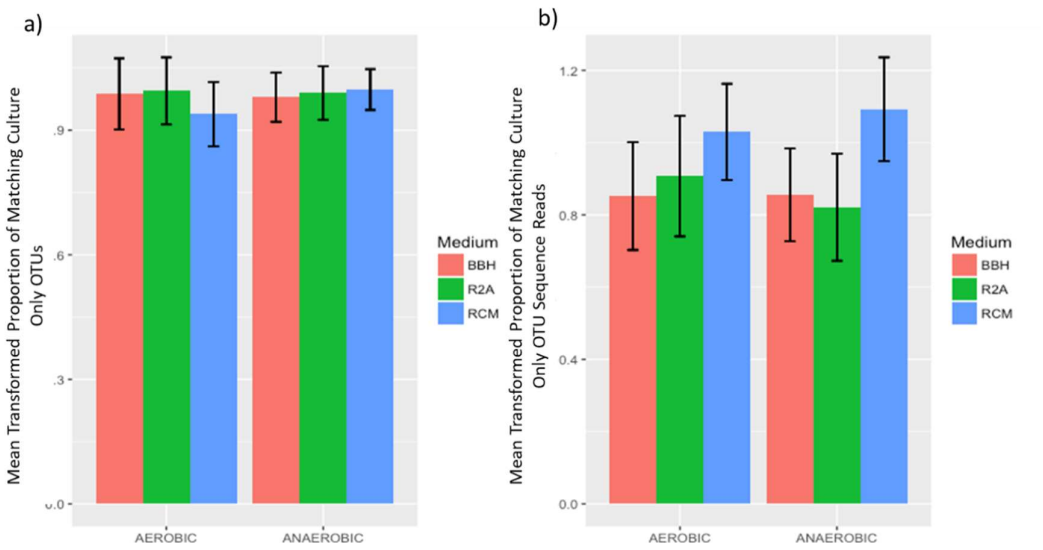


Figure 11 *Mean per sample proportion of a) matching culture only OTUs b) sequence reads assigned to matching culture only OTUs by culturing condition. Values shown are mean proportions across the samples in that condition, and means and standard errors are calculated from data transformed according to the criteria described in Methods.*

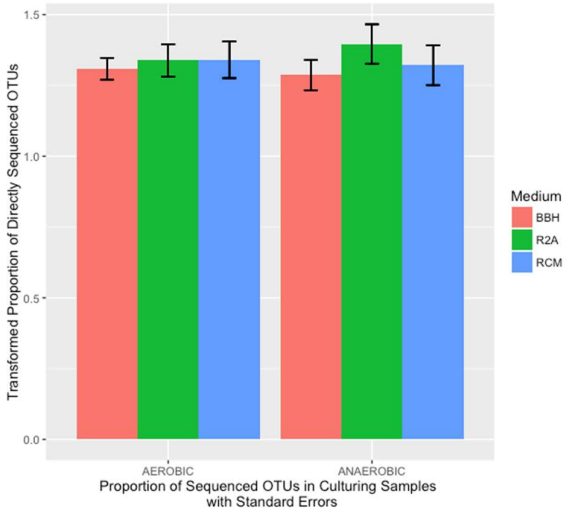


Figure 12 *Mean proportion of OTUs in directly sequenced aliquots which are also found in the matching cultured samples by culturing condition. Values shown are mean proportions across the samples in that condition, and means and standard errors are calculated from data transformed according to the criteria described in Methods.*

DISCUSSION

A set of red junglefowl semen samples were cultured to investigate whether members of the junglefowl seminal microbiota are viable post-ejaculation. We also investigated how the cultured communities relate to those obtained independent of culturing, and explored the effect of culturing condition on community composition and structure. Eighteen semen samples were cultured on agar plates under six different culturing treatments formed from combinations of three different growth media and two oxygen conditions. Growth of bacterial colonies was observed under all treatments, with 217 Operational Taxonomic Units (OTUs) sequenced overall, demonstrating the viability of at least a proportion of the bacteria found in junglefowl semen. These communities differed in the richness and diversity of bacterial OTUs isolated, as well as the abundance and presence of different bacterial groups, when cultured under different conditions. There was a large overlap with faecal samples, with the majority of cultured OTUs also found in faecal samples and again the proportion of OTUs in a sample, which were also found in faecal samples varied with culturing condition. Cultured samples differed in composition from directly sequenced semen samples as there were OTUs in directly sequenced semen samples which were not recovered from cultured samples, and vice versa, and the proportion of sample sequences from OTUs only found in cultures varied with growth medium. The results of this work offer insights into the biology and ecology of the red junglefowl ejaculate microbiota, as well as into the advantages and limitations of some of the methodological approaches employed in characterising and exploring bacterial communities.

Demonstrating Viability of the Seminal Microbiota

This demonstration of viable bacteria in semen is an important addition to work exploring these communities using culture independent, next-generation sequencing methods, which demonstrate the presence of bacterial DNA but cannot tell us whether the DNA is from a live organism or not. Live bacteria in ejaculates means that the male reproductive microbiota transmitted to the female during mating may be able to colonise the female reproductive tract,

leading to disruption, pathogenic colonisation or possibly colonisation by beneficial microbes. Here we cultured a range of bacterial phyla including genera found in the gut of healthy female domestic fowl (Videnska *et al*., 2013) and, when isolated in previous studies of seminal microbiota, have been acknowledged to contain opportunistic pathogens, though it must be acknowledged that a single genus can contain a range of phenotypes. For example, studies sequencing the human seminal fluid microbiota find *Anaerococcus*, *Granulicatella*, *Clostridales* (Hou et al., 2013) and *Lactobacillus* (Weng et al., 2014a), genuses noted by the authors as containing opportunistic pathogens. Westneat & Birch Rambo, (2000) isolate *Enterobacter*, *Acinetobacter* and *Pasteurella* from blackbird ejaculates and cite them as groups containing potential pathogens of humans and mammals. Additionally, two common pathogens of the chicken gut, *Bacillus cereus* and *Staphylococcus aureus* (Wang *et al*., 2014), have also been isolated from the junglefowl samples, though the commercially important pathogens of domestic fowl, *Campylobacte*r and *Salmonella*, previously isolated from fowl ejaculates (Reiber *et al*., 1995), were not cultured in this work.

Equally, members of the *Lactobacillus*, *Enterococcus*, *Faecalibacterium* genera and the *Gallibacterium* family have been cultured here, and these have all been shown to be common in the digestive tract of the healthy female domestic chicken (Videnska *et al*., 2013). In fact, *Lactobacillus* species with demonstrated probiotic properties have been cultured, such as *Lactobacillus reuteri* which is a commonly administered probiotic in chickens (Wang *et al*., 2014) and has the potential to help weight gain in broiler chickens (e.g. Edens *et al*., 1997; Li *et al*., 2007) and to be protective against pathogen induced mortality of chickens through competitive exclusion (Edens *et al*., 1997).

Thus, upon insemination the female red junglefowl is inoculated with a wide range of live bacterial organisms with the potential for a variety of consequences, both neutral, detrimental and even beneficial for the female. Additionally, the wide range of live organisms isolated, and the fact that substantial bacteria growth was observed for all ejaculates under all culturing

conditions, suggests that the junglefowl seminal microbiota is composed of a diverse set of bacteria, able to survive in a wide range of conditions which suggests it originates in the different physiological niches within the red junglefowl reproductive tract, from which originate the different components of the ejaculates (Lake, 1957; Nishiyama, 1955) as well as the digestive tract.

Relationship Between Directly Sequenced and Cultured Communities

As expected not all directly sequenced OTUs are culturable, but a substantial amount are; 60% of the directly sequenced OTUs overall being culturable and an average of 19% of OTUs in a directly sequenced sample also recovered from at least one of the matching six cultured communities, which is a large proportion compared to the Eckburg *et al*. (2005) estimate of 20% gut microbe culturability. Whilst this most reflects the improved scope of culturing methods with a wider range of conditions and more accurate and efficient identification techniques, there remains an uncultured portion, highlighting the advantage of culture-independent techniques in identifying difficult to culture components of the microbiota.

Unexpectedly, however, a substantial number of OTUs are found in cultured, but not directly sequenced samples; almost half of the 217 cultured OTUs are not found in directly sequenced samples, on average about a third of each sample. This is likely due to their existing at such a low number in the uncultured semen that they are missed by the resolution of the sequencing methods, even though this method was shown in chapter two to have achieved a high level of coverage and saturation of the community. Indeed, OTUs dropped from sequenced samples through rarefaction do account for a proportion, though not all, of those found in the cultured communities.

This is a strong reminder that the ejaculate microbiota is a transient community, and that those bacteria present at high levels in the directly sequenced samples though best at reflecting what is present at high abundances in the reproductive tract and ejaculates and thus what is likely to affect the sperm before and during ejaculation, and useful for exploring variation between

males, may not be those bacteria best suited to surviving transmission via the external environment or colonising a female host. This finding is, again, particularly important when considering sexual transmission of the microbiota; it may be that these culturable OTUs, despite existing at very low relative abundance in the ejaculate, survive exposure to the environment upon ejaculation and are successful in colonising the new habitat of the female reproductive tract, or that a perturbation leading to changes in condition in the male reproductive tract or ejaculate environment allow such OTUs to increase in abundance.

Methodologically, this finding of OTUs unique to both culture and directly sequenced samples is further corroboration of the benefits of combining both culturing and next-generation sequencing to fully explore microbial communities; though these OTUs may be rare enough to be missed by direct sequencing and an effectively complete representation of the seminal bacterial community may be captured without them as they are so rare, relative abundances of OTUs may change with a change in condition and those previously rare can increase in abundance. Whilst there was no variation in the number of 'culture only' OTUs per sample, the proportion of sequence reads in a sample assigned to these OTUs was significantly higher in the intermediately nutrient rich medium than the most nutrient rich, perhaps because it is the faster growing but fastidious organisms which dominate the nutrient rich but not the intermediate medium , allowing rare OTUs to colonise further and make up a higher proportion of community abundance, as abundance is shared more equally across OTUs. Thus, the choice of either method, or a complementary use of the two as in Browne *et al*. (2016) for example, as well as the sequencing depth and culturing conditions used needs to be considered with the aim of that specific study in mind i.e. is one interested in considering variation of microbiota in composition the male, focusing on what is transmitted to the female, exploring the sources of the seminal fluid microbiota or investigating the phenotypic properties and potential effects of the organisms present.

Effect of Culturing Conditions

It was found that different components of the diverse seminal microbiota community are selected for under different culturing conditions, with variation in the richness, diversity and composition of the bacterial communities produced between conditions. In both aerobic and anaerobic conditions, the highest bacterial species richness was found in the medium of intermediate nutrient richness, whilst the lowest number of bacterial species are isolated on the most nutrient rich medium. However, the nutrient rich medium had a higher diversity, though only significantly higher in the Shannon Index of diversity and not the Simpson Index. This is interesting as one might expect a lower species richness on a more nutrient rich medium as the fastest growing species can very quickly dominate and exclude slower growing species which are able to establish themselves on the intermediate medium, but one would also expect diversity to decrease with number of species as the indices are a compound of both richness and evenness of species i.e. how evenly abundance is distributed across species. In order for diversity to increase when species are lost, there must be a very high skew of abundances in the more species rich community, which is lost when species richness is reduced, either through the loss of very high and low abundance species, or an increase in rare and a decrease in common species, or a combination of the two. As the Simpson diversity is not significant where Shannon is, we speculate that the changes in diversity are being driven by the loss of very rare species as the Shannon index takes account of rare and abundant species equally where Simpson does not, so any effect of these rare species will be reduced in calculations of the Simpson measure of diversity (Morris *et al*., 2014). What may be happening then is that there are lots of very rare species present in the intermediate medium that are lost in the high nutrient rich medium, perhaps due to exclusion by faster growing species.

We also find that the species composition of a community varies with condition; the Principal Components Analysis axis that explains the highest proportion of variation between communities (PC1), was linked to oxygen treatment, and the second (PC2) with culturing medium. These axes could also be linked to the abundance of members of different taxa and

culturing conditions; *Firmicutes* species being more abundant in the anaerobic condition and *Actinobacteria* species more abundant in the aerobic condition and in turn these phyla seem to be linked to variation along the first axis.

We also see clusters of OTUs, in that groups of OTUs tended to co-occur in the same aliquots and some of these clusters were linked to changes in culturing condition; when clustered by Spearman correlation of their abundance in each aliquot we find several clusters of quite highly correlated OTUs, one of which was correlated negatively with PC1, the other positively with PC1 and a third negatively with PC2. It may be that such clusters are linked to organisms with similar nutrient and oxygen requirements but also there could be more complex processes at work leading to OTUs occurring together. For example, Videnska *et al*. (2014) point out that the presence of *Desulfovibrio* species in only adult chickens is probably due to the increase of *Bacteriodetes* in adult birds, a phylum which contains organisms capable of releasing the sulphate respired by the *Desulfovibrio* species from host products. Indeed, in physiological niches such as the human vagina (Ravel *et al*., 2011) and gut (Arumugam *et al*., 2011) it has been observed that variation between individuals is manifested in distinct clusters of communities, all existing in the same environment and performing the same function through different combinations of species. Such stratifications, as opposed to a continuous spectrum of variation in composition, indicate a more complex, intertwined community with members reliant on each other.

Such interactions could also explain the observed variation in richness and diversity with culturing conditions. A detailed investigation of these dynamics and interactions is beyond the scope of the current study but my findings here serve to highlight the complex dynamics of such microbial communities beyond simply the loss and gain of species independently of one another, and the potential effects that a change in condition may have on community dynamics and, consequently, structure.

What one can conclude is that different communities are obtained dependent on culturing condition, and these are not simply random results of sub-setting of the original species due to

founder effects, but predictable and repeatable communities linked to culturing condition, demonstrating that physiological and environmental conditions can shape microbiota communities.

Also of interest is the fact that these varied communities have all originated from red junglefowl seminal microbiota sample aliquots, and the same patterns are observed across several ejaculate samples. Again, this indicates a phenotypically diverse ejaculate microbiota formed of bacterial species from several different communities, from which different members and communities can be selected for through environmental conditions. This means that variation in the female reproductive tract environment may lead to selection for or against colonisation by different ejaculate members, and the diverse nature of the male ejaculate microbiota makes it more likely for us to expect at least some colonisation of the female reproductive tract, or some prevention mechanism selected for in females.

This variation in communities as a result of culturing condition is also very methodologically relevant, as seen with reference to differences found between cultured and directly sequenced samples, as it is clear that the culturing method used to study a microbiota will lead to the recovery of just a sub-section of the bacterial community present, and differences in community related to culturing condition need to be understood in order for the output of such studies to be properly interpreted and compared to other studies. For instance, one potential source of bacterial contamination of the ejaculate which I did explore in this work was exposure to faecal bacteria when passing through the cloaca and whilst there is a very high proportion of faecal OTUs that are found in cultured aliquots (on average 89% of OTUs per sample, 92% of sequences) the proportion of these OTUs, and sequences assigned to them, varies with culturing condition. The richest medium, BBH samples, have a lower proportion of faecal OTUs, and sequences assigned to these OTUs, than R2A the intermediate nutrient medium, and aerobic samples have a lower proportion of OTUs and these sequences than anaerobic samples.

Variation as a result of culturing condition also provides the opportunity to study the dynamics of the seminal microbiota, exploring variation in response to perturbations that cause a change in environment e.g. infection, dietary change, aging, or environmental changes experienced during the passage through the reproductive tract, ejaculation and once inside the female. Similar to these explorations of faecal work, matching culturing studies with work characterising the microbiota of surgically extracted parts of the male reproductive tract, the male digestive tract and even the female reproductive tract, would reveal the sources of the different species present in ejaculates and would lend context to studies of how the seminal microbiota changes as it passes through the male reproductive tract and is transmitted to and then potentially colonises the female reproductive tract. Considering the sources and conditions most likely to contribute to bacteria that is successfully transmitted to the female also gives some insight as to any aspects of the males' physiology and behaviour which may influence bacterial transmission to the female.

We have demonstrated that a large proportion of the bacteria in the seminal microbiota is live, revealing a transient, complex and phenotypically diverse community. This has a number of both biological and methodological implications and indicates plenty of areas with scope for further exploration.

APPENDIX§1

Table A1 *The make up of each duplicate dataset, indicating which version (1 or 2) of the duplciated samples (A and B) are included.*

Duplicates 1	Duplicates 2	Duplicates 3	Duplicates 4
A1 and B1	A2 and B2	A2 and B1	A1 and B2

Table A2 *Fixed factors of the 'best' models for each community metric for each of the four duplicate datasets.*

	Duplicates 1	Duplicates 2	Duplicates 3	Duplicates 4
SpecRich	Med	Aer * Med	Aer * Med	Med
ShanDiv	Med	Aer * Med	Aer * Med	Med
SimpDiv	Med	Med	Med	Med
PC1	Aer	Aer	Aer	Aer
PC2	Med	Med	Med	Med

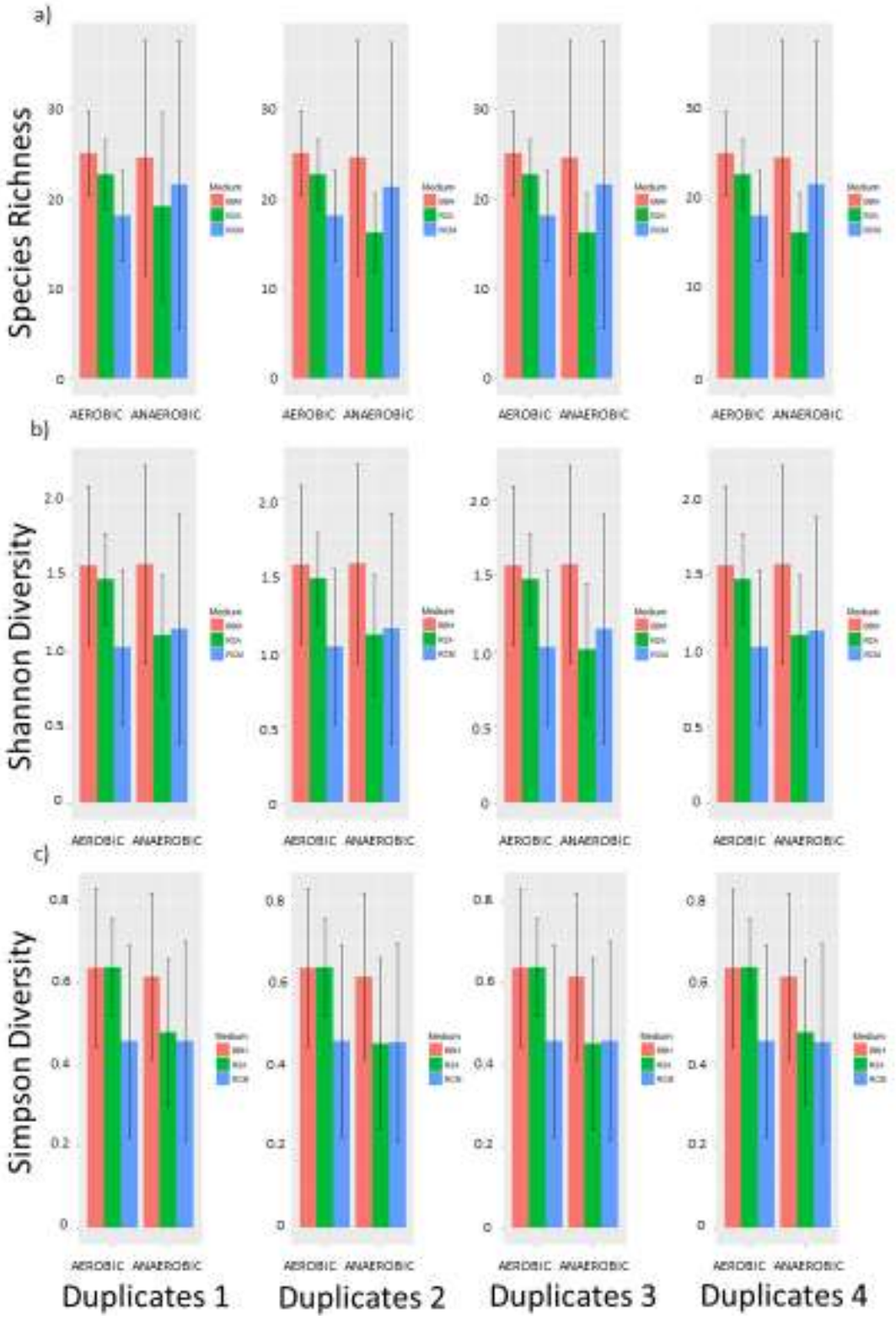


Figure A1 *Figure illustrating the relationships between a) species richness b) Shannon diversity and c) Simpson diversity and the two culturing treatments, oxygen treatment and medium for each of the four duplicate datasets. Values shown are means across the aliquots in that condition, and bars indicate standard errors. Means and standard errors are calculated from data transformed according to the criteria described in methods.*

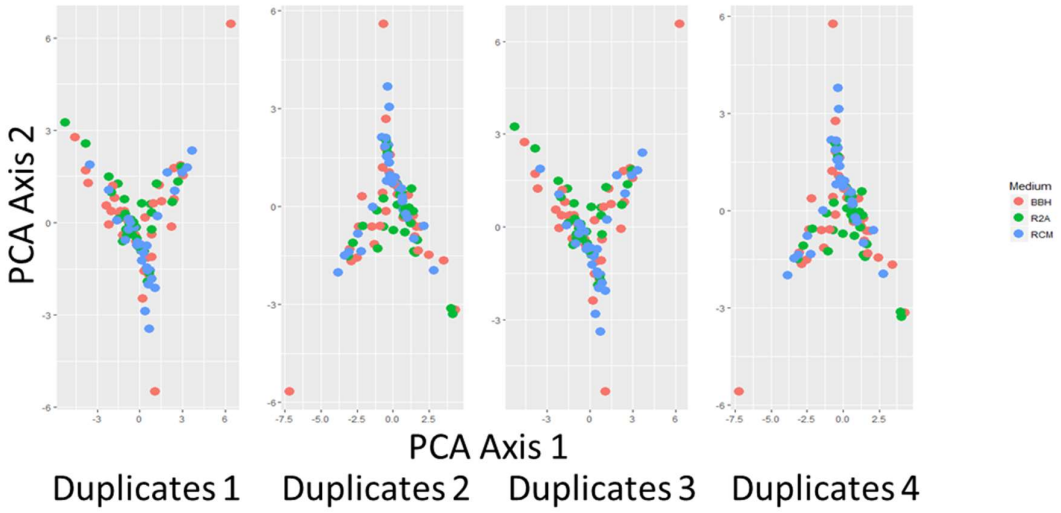


Figure A2 *The position of aliquots within the PCA space created for each duplicate dataset using raw values of PC1 and PC2. Colours indicate treatment medium.*

APPENDIX§2

Table A3 *Fixed factors of the 'best' models for each community metric for each of four duplicate datasets with and without outliers. Removal of outliers only changes output for Duplicate 4 dataset.*

	Dup 1	Dup 1 Outliers	Dup 2	Dup 2 Outliers	Dup 3	Dup 3 Outliers	Dup 4	Dup 4 Outliers
SpecRich	Med	Med	Aer * Med	Aer * Med	Aer * Med	Aer * Med	Aer*Med	Med
ShanDiv	Med	Med	Aer * Med	Aer * Med	Aer * Med	Aer * Med	Aer+ Med	Med
SimpDiv	Med	Med	Med	Med	Med	Med	Aer + Med	Med
PC1	Aer	Aer	Aer	Aer	Aer	Aer	Aer	Aer
PC2	Med	Med	Med	Med	Med	Med	Med	Med

Table A4 *List of OTUs found only in outlier samples, and their relative abundances in each sample. NA indicates that an OTU is not found in that sample. Those in 'OUTLIER' column are found only in that sample and not in other non-outlier samples. Those in 'UNQIUE' column are found only in that outlier sample, not in the other outlier samples nor in non-outlier samples.*

OTUID	X5B_X38_OUTLIEROTUS	X5A_X16_OUTLIEROTUS	X7A_X80_OUTLIEROTUS	X5B_X38_UNIQUEOTUS	X5A_X16_UNIQUEOTUS	X7A_X80_UNIQUEOTUS
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Pseudoalteromonadaceae;Other;Other	0.001912046	0.001912046	NA	NA	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Vibrio;s__metschnikovii	0.001912046	0.001912046	NA	NA	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__Halomonas;s__anticariensis	0.001912046	0.003824092	NA	NA	NA	NA
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Dietziaceae;g__Dietzia;s__	0.001912046	NA	0.001912046	NA	NA	NA
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Intrasporangiaceae;g__Terracoccus;s__	0.001912046	NA	NA	0.001912046	NA	NA
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Micrococcus;Other	0.001912046	NA	NA	0.001912046	NA	NA

k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Micrococcus;s__luteus	0.0019112046	NA	NA	0.0019112046	NA	NA
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Nocardioidaceae;g__Nocardioides;s__	0.0019112046	NA	NA	0.0019112046	NA	NA
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Planococcus;Other	0.0019112046	NA	NA	0.0019112046	NA	NA
k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia;s__	0.0019112046	NA	NA	0.0019112046	NA	NA
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Novosphingobium;s__	0.0019112046	NA	NA	0.0019112046	NA	NA
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingobium;s__xenophagum	0.0019112046	NA	NA	0.0019112046	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Idiomarinaceae;g__;s__	0.0019112046	NA	NA	0.0019112046	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Shewanellaceae;Other;Other	0.0019112046	NA	NA	0.0019112046	NA	NA

k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Enterobacter;Other	0.001912046	NA	NA	0.001912046	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__;s__	0.001912046	NA	NA	0.001912046	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__viridiflava	0.001912046	NA	NA	0.001912046	NA	NA
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__;s__	0.003824092	NA	NA	0.003824092	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__;s__	0.005736138	NA	NA	0.005736138	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__	0.005736138	NA	NA	0.005736138	NA	NA
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingobium;s__	0.013384321	NA	NA	0.013384321	NA	NA
k__Bacteria;p__TM7;c__TM7-3;o__CW040;f__F16;g__;s__	0.034416826	NA	NA	0.034416826	NA	NA

k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;Other;Other	NA	0.001912046	0.001912046	NA	NA	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Lysobacter;s__	NA	0.001912046	0.001912046	NA	NA	NA
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__	NA	0.001912046	NA	NA	0.001912046	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Succinivibrionaceae;Other;Other	NA	0.001912046	NA	NA	0.001912046	NA
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Citrobacter;s__	NA	0.001912046	NA	NA	0.001912046	NA
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Kocuria;s__rhizophila	NA	0.005736138	NA	NA	0.005736138	NA
k__Bacteria;p__Actinobacteria;c__Thermoleophilia;o__Solirubrobacterales;f__;g__;s__	NA	0.005736138	NA	NA	0.005736138	NA
k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae];g__Ornithobacterium;s__	NA	0.015296367	NA	NA	0.015296367	NA
k__Bacteria;p__Acidobacteria;c__Acidobacteria-6;o__iii1-15;f__mb2424;g__;s__	NA	NA	0.001912046	NA	NA	0.001912046

k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Mycetocola;s__	NA	NA	0.001912046	NA	NA	0.001912046
k__Bacteria;p__Bacteroidetes;c__Cytophagia;o__Cytophagales;f__Cyclobacteriaceae;g__;s__	NA	NA	0.001912046	NA	NA	0.001912046
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Tissierella_Soehngenia;s__	NA	NA	0.001912046	NA	NA	0.001912046
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__Devosia;s__	NA	NA	0.001912046	NA	NA	0.001912046
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Vibrio;Other	NA	NA	0.001912046	NA	NA	0.001912046
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Oceanospirillaceae;g__Marinomonas;s__	NA	NA	0.003824092	NA	NA	0.003824092
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Nocardioidaceae;g__Nocardioides;s__plantarum	NA	NA	0.005736138	NA	NA	0.005736138

APPENDIX§3

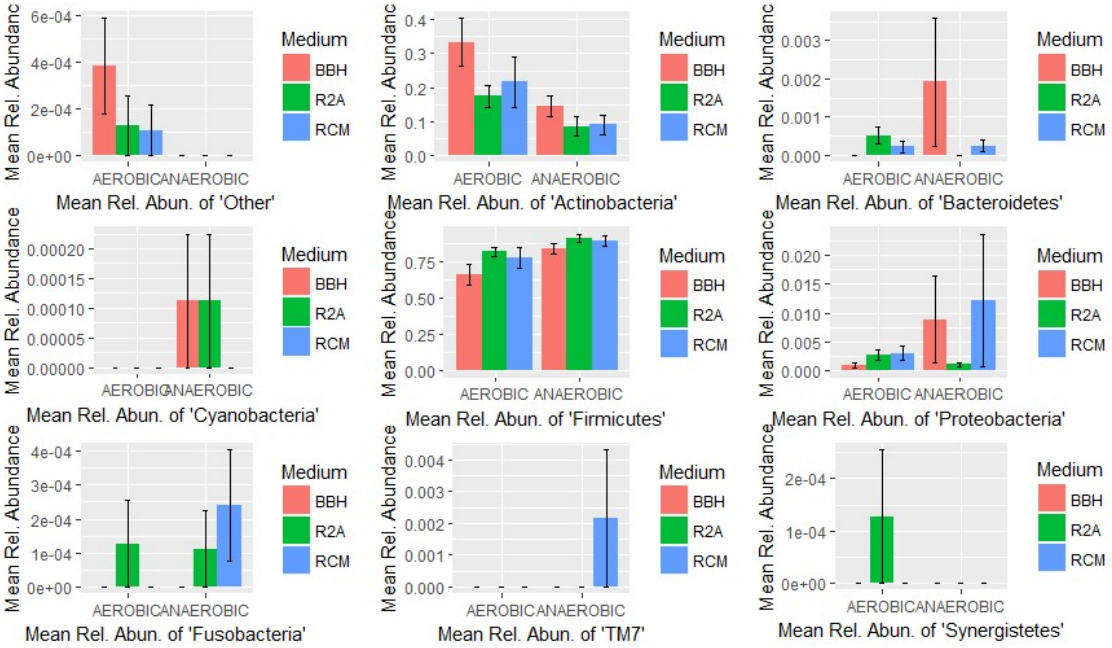


Figure A5 *Mean per sample relative abundance of each phylum for each of the six culturing treatment. Error bars indicate standard errors.*

CHAPTER 4. DRIVERS OF VARIATION IN SEMINAL MICROBIOTAS WITHIN AND ACROSS MALE RED JUNGLEFOWL

Rosie Eccleston[a]; Kiyono Sekii[a,b]; Grant C. McDonald[a]; Lewis Spurgin[a,c]; David S. Richardson[c];

Stephen Preston[d]; Tommaso Pizzari[a];

[a] Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK

[b] Department of Biology, University of Hirosaki, Hirosaki, Japan

[c] School of Biological Sciences, University of East Anglia, Norwich, UK

[d]Department of Zoology, University of Oxford, Oxford, UK

ABSTRACT

The presence of bacteria in ejaculates has been demonstrated and, at least in humans, has been linked to male health, disease transmission and infertility. However, little is known about the patterns of intra- and inter-individual variation and biological drivers underpinning such variation, especially in non-human species. In principle, consistent variation across males can be influenced by the male genotype, especially for loci involved in immune responses such as the Major Histocompatibility Complex (MHC) in vertebrates, while within-male variation can be determined by temporal dynamics at different scales, e.g. ageing, seasonal effects, or the recent mating history of a male. Quantifying these sources of variation is critical to understand the evolutionary ecological role of seminal microbiotas. Here, we present a comprehensive study of variation in the male seminal microbiota of a captive population of red junglefowl, *Gallus gallus*, focusing on two major potential drivers: MHC genotype, and the mating history of males. In an analysis to investigate variation between individual males, as well as seasonal variation within a male, allelic richness and amino acid divergence at an MHC locus were the only factors which significantly explained variation in seminal microbiota composition. Evidence of a genetic determinant of seminal microbiota could allow selection on male microbiotas, and microbiota-mediated means of mate selection and kin recognition. In addition, analysis of samples from males following a mating trial found that experimental groups significantly differed in seminal microbiota composition and structure. Within group variation was linked to the overall pattern of sexual interactions of the male social group, as described by female polyandry and sexual network density, modularity and mean path length. However individual sexual behaviours of males did not explain variation in seminal microbiota across males suggesting that drivers of

group and individual variation may operate at different scales and through different mechanisms. Further work should seek to establish the nature of these effects and discount potential confounds (independent of mating).

INTRODUCTION

Male semen has been demonstrated to harbour communities of bacteria (microbiotas) in a phylogenetically diverse range of species including some insects (Andere *et al*., 2011), mammals (e.g. Javurek *et al*., 2016), birds (e.g. Hupton *et al*., 2003) and reptiles (e.g. Johnston *et al*., 2014). In principle, bacteria can influence the fertilising efficiency of an ejaculate, e.g. by modulating exogenous resources available to spermatozoa in the seminal fluid, or by interacting directly with spermatozoa. This realisation raises the possibility that seminal microbiota may have important and widespread consequences for host fitness. The presence of pathogenic bacteria in semen has been demonstrated in turkeys, *Meleagris gallopavo,* domestic fowl, *Gallus (gallus) domesticus* (Donoghue *et al*., 2004; Reiber *et al*., 1995) and tree swallows, *Tachycineta bicolor* (Lombardo & Thorpe, 2000), as well as evidence of the potential for transmission of such bacteria from males to females via mating (Kulkarni & Heeb, 2007). Further, in a study of red-winged blackbirds, *Agelaius phoeniceus*, repeat ejaculate samples were collected from the same males across two breeding seasons, demonstrating both within and between individual variation in whether bacteria was isolated from a sample, and the species isolated. Samples taken from the same male were no more similar than randomly expected whether taken in different breeding seasons, in the same breeding season or on the same day, though factors driving such variation were not explored. Between male variation remained similarly unexplained, with neither season, geographical location or age significantly predicting bacterial prevalence (Hupton *et al*., 2003).

Evidence suggests that, at least in humans, variation in the nature and composition of the seminal microbiota may be linked with variation in a number of health conditions including chronic prostatitis (Ivanov *et al*., 2009), viral load and shedding of HIV (Liu *et al*., 2014) and male infertility (Weng et al., 2014b). Similarly, the transmission of the seminal microbiota to females at mating may also have repercussions for female fitness if it disrupts the local microbiota of the

female reproductive tract. For example, in human females, changes to the reproductive microbiota associated with dysbiosis as a result of mating (Gajer *et al*., 2012; Ravel *et al*., 2011) suggest that transmission of male microbiota members leads to the disruption of healthy female reproductive microbiotas beyond simply the transmission of pathogenic species. Importantly for female fitness, a deviation away from a stable, *Lactobacilllus*-dominated vaginal microbiota state in human females is associated with reduced prevalence of HIV and other STDs (Borgdorff *et al*., 2014). In principle, ejaculates might also harbour bacteria with beneficial effects on females which could be transmitted to the female via mating (Lombardo *et al*., 1999). Therefore, the male seminal microbiota may represent a neglected but potentially considerable factor in sexual selection, sexual conflict and inter-sexual coevolutionary dynamics in the host.

A critical first step toward developing an understanding of the roles of seminal microbiotas is to quantify patterns of intra- and inter-individual variation in structure and composition of these bacterial communities, and identify biological drivers of such variation. Resolving patterns of intra- and inter-male variation in seminal microbiota and its underpinning drivers would contribute to establishing links with host function and behaviour, and increase our understanding of the ecology of these microbial communities and how they are shaped and develop as a result of the host environment. Such knowledge adds to the increasing body of microbiota research aimed at addressing how micro-organisms interact with each other, their host and their environment. This is considered invaluable in light of the potential value of this research in a variety of applications highlighted by Antwis *et al*. (2017) who discuss how an increased understanding in the ecology, the biodiversity and the dynamics of various microbiota through fundamental research could then be translated into further understanding their role in environmental processes and ecosystem services, and mediators of host evolution and fitness and thus how they could be manipulated and harnessed to solve real-world problems in the fields of human health, food security and climate change resilience.

For instance, much theoretical work has investigated the way in which sexually transmitted microbes may affect the evolution of host mating systems, mating rates and patterns of mate choice, by modulating the cost of mating (e.g. Graves & Duvall, 1995; Thrall *et al*., 1997; Thrall *et al*., 2000; Kokko *et al*., 2002). In the same way, one could imagine that the potential for sexual transmission of seminal microbiota might affect the fitness payoffs (and thus patterns of selection) on mating rates and criteria of mate choice if beneficial microbiota were transferred. For example, inter-individual variation in microbiota has been suggested to inform mate choice, when host phenotypic cues signal health, or genetic quality. For instance, Rowe *et al*. (2011) found bill coloration was associated with anti-bactericidal activity of mallard (*Anas platyrhynchos*) ejaculates, and posit that females could therefore use male bill colour as a cue to whether a male is likely to carry sexually transmitted infections (STIs) or to have bacterially-damaged spermatozoa with lower fertility. Moreover, variation in the microbiota of the ejaculates of different males might in principle enable patterns of cryptic female choice in relation to male genotype or phenotype (Eberhard, 1996; Firman *et al*., 2017). Additionally, inter-individual variation in microbiota may provide a mechanism for kin recognition if a host's microbiota indicates relatedness e.g. through olfactory cues. In spotted hyenas (*Crocuta crocuta*) fermentative bacteria were found in scent glands and the composition of these microbiota co-varied with volatile fatty acid secretions, and with sex and reproductive state (Theis *et al*., 2013). Further, the capacity for *Drosophila melanogaster* to recognise kin varies with ecological context, potentially driven by the gut microbiota (Lizé *et al*., 2014), and in meerkats (*Suricata suricatta*) kin discrimination is mediated by anal gland scent (Leclaire *et al*., 2013) where secreted volatile compounds covary with bacterial communities (Leclaire *et al*., 2017).

The seminal microbiota is a particularly interesting bacterial community to consider given that it is a transient community, likely formed of a collection of bacteria originating from a number of physiological sites and existing together for a short amount of time. Further, because it is necessarily transmitted from one individual to another it is likely to mediate, and be mediated by, interactions between individuals.

Despite the demonstrated potential for the seminal microbiota to impact the biology of both male and female hosts, and the need to understand the implications of variation in microbiota composition for both clinical work and fundamental research, relatively little is known about the patterns of intra- and inter-individual variation, and the forces shaping such variation. Most work has been done in humans with a clinical focus, often limited to certain subsets of a population i.e. those suffering from a particular condition. There is evidence of an effect of HIV infection (Liu *et al*., 2014) where decreased seminal microbiota diversity and richness was observed following infection in men, and varied with certain cytokine levels and semen viral load in infected individuals. Additionally, insertive anal sex has been associated with higher bacterial concentrations and anti-retroviral therapy was associated with lower bacterial concentrations in human semen (Korhonen *et al*., 2017). It is difficult to ascertain the relevance of these clinical human studies for other species.

Both genetic and environmental mechanisms might contribute to inter-male variation in seminal microbiota. For example, microbiotas acquired through shared rearing environment during development may result in little variation within a cohort but substantial variation across individuals from different cohorts. A potentially important source of inter-male variation in seminal microbiota is determined by the host immune-genotype. Whilst the ejaculate is likely a transient community so is not under prolonged selective pressure by the immune system, it likely originates from a variety of sources within the digestive and reproductive tracts where this is more likely to be the case. The vertebrate innate immune system can quickly respond to non-host cells through the recognition of patterns of marker proteins on these cells. The Major Histocompatability Complex (MHC) is a highly polymorphic set of genes encoding which of these molecular patterns a host's immune system recognises and responds to, and thus can be expected to mediate which bacteria are able to inhabit the body without being cleared by the immune system. Such effects might be linked to the diversity and richness of alleles at the MHC as well as to specific alleles, for instance higher richness could lead to lower bacterial diversity as more bacterial molecular patterns can be recognised and responded to by the immune system. We may

also expect microbiota composition to covary with MHC alleles or genotypes such that the presence and absence of given alleles is associated with the presence and absence of given bacterial taxa, and that individuals with a similar MHC genotype have a similar microbiota composition. The MHC therefore provides a potential genetic means of shaping the composition of microbial communities. Studies of the bacterially derived fatty acids in faecal samples of mice, *Mus musculus*, congenic for Major Histocompatibility Complex (MHC) and suggested that composition of the gut microbiota was genetically determined, and MHC alone had an effect (Toivanen *et al*., 2001) In human infants, abundances of bacterial taxa differed between MHC genotypes associated with differing risk of coeliac disease (Palma *et al*., 2010). More specifically, in the three-spined stickleback, *Gasterosteus aculeatus*, Bolnick *et al*. (2014) found that MHC diversity influenced microbial diversity and that there were several pairwise associations between MHC genotypes and the abundance of certain bacterial taxa. A genetic basis of the microbiota is of interest as it may help to explain how geographically separated populations of the same species share the same core microbiota (e.g. Arumugam *et al*., 2011), and how a disrupted microbiota returns to the pre-disrupted state (Dethlefsen *et al*., 2008). Additionally, genetic determinants of microbiota composition provide a mechanistic basis for hypotheses such as identifying genetic quality of mates and recognising kin via microbiota mediated means e.g. odour, as well as a basis of selection for and against beneficial and detrimental microbiota. For instance, in the red junglefowl there is evidence that females favour MHC dissimilar males via cryptic female choice, and there is evidence of MHC-based mating decisions in stickleback (Reusch *et al*., 2001; Milinski *et al*., 2005), badgers, *Meles meles*, (Sin *et al*., 2015) and wolves, *Canis lupus* (Galaverni *et al*., 2016).

Studies of the genetic determinants of seminal microbiota composition are complicated by the fact that seminal microbiotas are also likely to vary dynamically within individual males. These dynamics may reflect different biological processes occurring at different time scales. Microbiotas change and develop with host age, for example the complexity of microbial communities in the chicken gut increases in the first few weeks of life with individuals becoming

more diverse, but more similar to other chickens, as they approach maturity (Crhanova *et al*., 2011; Juricova *et al*., 2013). Moreover, there could be effects as males reach senescence, for example humans over sixty-five years of age had a microbiota distinct from younger subjects, with higher variation between individuals in this older age group (Claesson *et al*., 2012). Microbiota may also fluctuate throughout a season due to behavioural, hormonal changes and dietary changes: the gut microbiota of human females changes from the first to the third trimester of pregnancy, potentially related to pregnancy induced hormonal changes (Koren *et al*., 2012), there is stress-induced changes in microbiota in mice (Tannock & Savage, 1974) and within-species variation in diet associated with different habitats contributed to variation in microbial composition of savannah baboon, *Papio cynocephalus*, fecal samples (Ren *et al*., 2015).

A potentially critical driver of temporal changes within males is determined by patterns of horizontal transmission through social interactions in general, and in the specific case of seminal microbiotas, through sexual interactions. A number of studies have explored the effect of mating on female reproductive microbiota but, perhaps because of the asymmetry of exposure, the effect of mating on male ejaculate microbiota has been neglected; the direction of gamete transfer from male to female means the female is exposed to the male microbiota much longer than the male is exposed to the female's. It has been demonstrated in one bird species, the zebra finch, *Taeniopygia guttata*, that sexual transmission of a pathogen is more likely male to female than female to male (Kulkarni & Heeb, 2007). However, female-to-male transmission also occurs albeit with lower probability, and one can therefore reasonably expect an effect of mating on the male seminal microbiota. Indeed, we see evidence of transmission of bacteria from female to male in longitudinal clinical studies of Sexually Transmitted Diseases (STDs) in humans, for instance in a longitudinal study of Chlamydial infection in men with female sexual partners, reinfection of men following clearance of an initial infection was associated with condom use; those who always use a condom with a new sexual partner are less likely to be re-infected (Dunne *et al*., 2008). Dynamic changes in the seminal microbiota of individual males could therefore reflect temporal patterns in the level of polygamy of a population (i.e. the average number of sexual partners) and

the way in which partners are shared by individuals within the population i.e. the structure of the sexual network (McDonald & Pizzari, 2015). In studies of bacterial transmission through social networks of a social spider, *Stegodyphus dumicola*, structure of the network affects transmission dynamics of the bacteria within the population (Keiser *et al*., 2017), and social group membership predicted gut microbiota similarity in red-bellied lemurs, *Eulemur rubriventer* (Raulo *et al*., 2017). Faecal sample diversity was negatively associated with an individual's sociality according to indices of grooming and huddling behaviours, and lemurs more closely associated through such behaviour had more similar gut microbial composition. Thus we may expect the structure of sexual networks to affect the seminal microbiota composition of network members, through mating associated transmission and disruption.

Studying the way in which dynamic changes in sexual behaviour affect intra-male variation in seminal microbiotas typically requires sequencing the entire microbial community of individual males at multiple time points and collecting fine-grained data on the male's socio-sexual behaviour over the same period of time. Here, we focus on using such data to explore inter- and intra-individual variation in the seminal microbiotas of male red junglefowl, *Gallus gallus*.

Characterisation of the avian seminal microbiota is of particular interest because much of the research investigating the evolution and ecology of mating behaviour and mating systems has focused on birds. Sheldon (1993) highlights the potential advantages of avian systems to investigate the potential for sexually-transmitted microbes to affect the evolution of mating systems. The previous chapters focused on developing a method to characterise the seminal microbiota, and on demonstrating the viability of the bacterial organisms present. This chapter explores the potential sources of variation in these communities within and between male red junglefowl, with a particular focus on immunogenetic-based variation and the effect of a male's recent mating history. The chapter has two broad aims.

First, we quantify population-level patterns of variation independent of mating interactions over two successive breeding seasons, considering the contribution of time in the

season, male age, mating status and the male MHC genotype. The red junglefowl MHC, like that of the chicken, is a minimal essential MHC in that it comprises only two loci, BF and BLB (Kaufman *et al*., 1999; Worley *et al*., 2008). In our study population, BF and BLB loci have nine and ten alleles respectively, BF is the equivalent of the mammalian MHC Class I, and BLB the mammalian MHC Class II (Worley *et al*., 2008). In mammals, MHC II encodes recognition molecules in the immune pathway that samples the extracellular environment and therefore mediates the response to bacteria and is more likely to be involved in determining the composition of the microbiota. MHC I on the other hand, determines the response to intracellular organisms such as viruses. We therefore predict that any association between the red junglefowl MHC and its ejaculate microbiota may be more likely to manifest in BLB, although we investigate both loci here. We consider amino acid diversity (the differences between all alleles at a locus in terms of their amino acid sequences), allelic richness (the number of alleles at each locus) and genotypes at the two loci separately.

Second, having characterised background patterns of variation, we then focus on the potential influence of a male's mating history on seminal microbiota, using fine-grained data on mating behaviour recorded during 10-day experimental mating trials in which groups of males and females interacted freely under standardised conditions. By combining data on mating behaviour recorded during the mating trial with seminal microbiota information obtained by sampling males at the end of a trial, we ask whether males kept in the same mating group develop more similar bacterial communities to each other than males in other groups, whether different aspects of a male and his mate's mating behaviour contributes to inter-male variation in bacterial communities, and whether a male's position in the sexual network contributes to variation in bacterial communities.

METHODS

Study Population

We studied a population (ranging between 200 and 250 adults) of red junglefowl housed in outdoor aviaries at the University of Oxford John Krebs Field Station in Wytham, Oxfordshire. Birds are hatched between summer and autumn and are denoted as one year old in their first breeding season in the following year, two years old in their second breeding season and so on. Males are considered sexually mature from ~6 months of age. Eggs are incubated artificially in batches, and chicks are raised in groups first indoor, and after a few weeks, in small single-sexed groups in outdoor aviaries. Birds are fed on the same diet and are vaccinated against Marek Disease Virus with a live vaccine shortly after hatching.

Over two breeding seasons, 2013 and 2015, we collected and analysed 117 semen samples from 51 sexually mature males, ranging in age from one to five years old. All males were kept in male-only groups, except when exposed to females in experimental mating groups (see below). Males were moved across different male-only groups of varying sizes when not in experimental mating trial groups and the exposure to mating groups throughout their lifetime will differ between males though here we include information only on female mating exposure in the breeding season from which a sample is collected, with 39 of the 51 males sampled having been included in one of the six mating trials carried out (four mating trials in 2013, two in 2015). Samples were collected from May to September at multiple time points during both breeding seasons. As some samples were collected opportunistically when birds were being handled for other experiments, and because semen was not always produced by males when sampling was attempted by abdominal massage, the number of samples from males varied between one and eight per male, and eight males were included in both the 2013 and 2015 season.

From previous work on this population, MHC information was available for 25 of the 51 males sampled, and so this subset of 25 males was used in the first section of the analysis. Sample collection and processing was consistent across seasons and for all birds (see below).

Experimental Design

Firstly, we conducted an analysis to investigate variation between individual males, as well as seasonal variation within males, using 91 semen samples from the 25 males for whom MHC information was available. Samples were collected at multiple time points during both breeding seasons. As some samples were collected opportunistically when birds were being handled for other experiments, and because semen was not always produced by males when sampling was attempted by abdominal massage, the number of samples from males is not standard, with between one and eight repeats per male, and eight males included in both the 2013 and 2015 season. In this analysis we explored the effect of collection date (time in the season), male age, cohort (year of hatching), MHC genotype and metrics and inclusion in a mating group upon microbiota composition. MHC information was obtained from past characterisation efforts for previous work on the study population using the method developed by Worley *et al*., (2008).

In a second set of analyses to specifically consider the effect of mating, 39 samples were obtained, one from each of 39 males which took part in a total of six mating trials. This analysis took advantage of mating experiments run across the 2013 and 2015 breeding season. The methodology of this study is described in detail in McDonald *et al*. (2017). Briefly, each mating trial included a unit of 10 males that were kept in with a unit of 12 females to create a mixed-sex mating group. A unique unit of 10 males was used for each trial. All six mixed-sex mating groups were created from six unique 10-male units paired with four unique 12-female units, such that two male units share the same two female units. Each mixed-sex mating group was studied for 10 consecutive days. On the morning of the first day of each trail all 12 females in the unit were entered into a pen along with all 10 males in the unit. The sexual and social behaviour of these birds was monitored through direct observations as all individuals in the group could be

observed by the experimenter at once, and behavioural observations took place every morning between 0500-0800hr (3hrs) and every evening between 1800-2100hr (3hrs). Because these are the periods when the bulk of mating interactions occur, the study captured the vast majority of all mating activity McDonald *et al*. (2017). All copulation events were recorded; for the purposes of this study a copulation event was defined as when cloacal contact was observed or when the males' tail was lowered over the female's cloaca and contact was assumed to occur. Females were removed from the pen at the end of the trial immediately after the last observation period on the last day of the trial, and males were sampled for their seminal microbiota within two days of completion of a trial, before return to other male-only pens in the general aviary population.

Because only 39 males produced semen samples in response to abdominal massage following a mating trial, this analysis is based only on 39 samples, not 60.

Sample Collection and Processing

Semen samples were obtained using abdominal massage as described in chapter 2, and with the same collection and laboratory processing protocol implemented at the field station to minimize environmental contamination (see chapter 2). Once processed, all samples were immediately placed on dry ice, before transfer to the Department of Zoology and storage at -80°C within the same day of collection for further processing.

Sequences were generated from samples, via the Illumina MiSeq Platform using the pipeline developed and described in chapter 2. Briefly DNA was extracted from the samples using the DNeasy Blood and Tissue Kit (QIAGEN) according to manufacturer's guidelines and adapted to optimize extraction of bacterial DNA from ejaculate samples. The V3 and V4 regions of the bacterial 16S encoding genes were targeted and amplified using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and bacterial primers 347F 803R. Amplification was carried out according to the Phusion polymerase manufacturer's protocol and with reference to the primer melting temperatures. The forward primer, 347F, also has a barcode sequence attached, with a different barcode added to the reaction mix for each individual sample prior to

amplification, to allow the sample of origin of each sequence following sequencing. Again, careful protocols are followed to minimise contamination both from the DNA product of previous amplifications with the same target, environmental contamination from lab space, equipment and reagents as well as cross-contamination between samples.

Once amplification of the target was confirmed by gel electrophoresis and visualization under UV light, the PCR products are pooled into DNA libraries, purified using QIAGEN MiniElute PCR Purification Kit and prepared for sequencing on the Illlumina MiSeq platform, with attachment of the Illumina barcoded index primer Illumina adaptor using the NEBNext Ultra DNA Library Prep Kit before submission to the sequencing facility.

Post-Illumina Processing

Sequences resulting from Illumina MiSeq were analysed, again, according to the chapter 2 protocol, using custom made scripts in R and QIIME and with the same approaches to remove and address whole dataset and per-MiSeq contaminants, to achieve an equal number of sequences per sample with sufficient coverage and to calculate relative abundance. All subsequent analysis was performed using R software version 3.3.2 (R Core Team, 2016).

Statistical Analysis

Measures of seminal microbiota diversity

Three community diversity metrics (species richness, Shannon Diversity Index and Simpson Index of Diversity) were calculated for each sample using the R package 'vegan'. Species richness is simply the number of Operational Taxonomic Units (OTUs; clusters of reads with 97% sequence similarity considered equivalent to bacterial species) found in a sample. Shannon and Simpson are two similar measures of diversity combining richness and evenness, where high numbers indicate a more diverse community, with a higher probability that two individuals randomly selected from a sample will belong to different species in the case of Simpson, and a higher uncertainty in predicting the species identity of an individual that is taken at random from the dataset according to Shannon. Simpson is bounded by zero and one, and the calculation places

more emphasis on more abundant OTUs where the calculation of Shannon weights all OTUs equally.

Further, we performed a Principal Component Analysis (PCA) across all samples using the standard 'stats' R package function prcomp. There are OTUs that occur in samples at very low abundances, or in just a few samples, but could affect sample PCA scores if included in the analysis despite their rarity meaning there is insufficient information on their patterns of abundance. The PCA was therefore performed using a subset of 39 OTUs selected because they occur in at least 10% of samples and at an average relative abundance equal to or greater than 0.1%. The original relative abundances of the 39 remaining OTUs were used as input into the analysis. The first PCA axis explained 13% and the second 8% of variation in relative abundance of OTUs between samples (Fig. 1). We used the first two PCA axes of variation as additional metrics to describe patterns of variation in the microbial communities of semen samples. All five community metrics (species richness, Shannon diversity, Simpson diversity, PC1 and PC2) were used as response variables in statistical models exploring the sources of variation between samples.

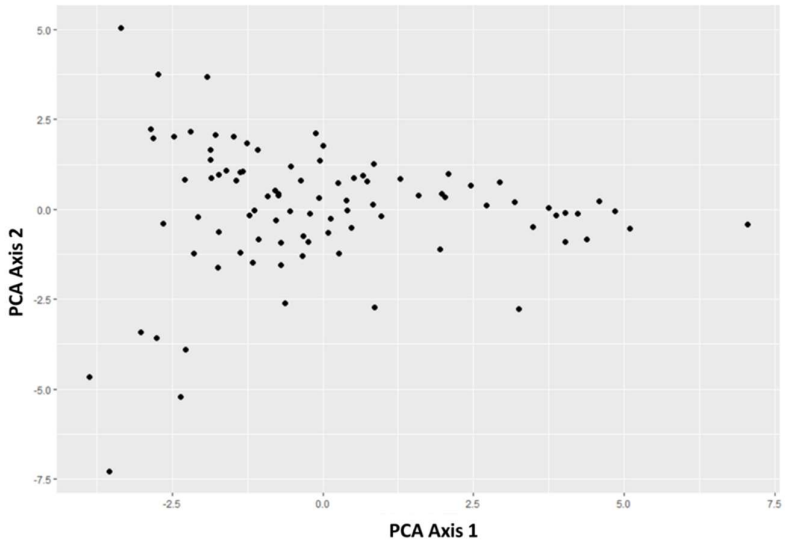


Figure 1 *The first two principal component axes calculated across 91 semen samples from 25 male red junglefowl. Principal components were calculated using a subset of 39 OTUs selected because they occur in at least 10% of samples and at an average relative abundance equal to or greater than 0.1%*

The Relative Influence of Intrinsic and External Drivers of Variation of the Seminal Microbiota

Here potential sources of variation in the composition and structure of the red junglefowl seminal microbiota are simultaneously investigated. These include factors that may result in consistent variation between males – male cohort and the diversity and richness of the male MHC genotype –factors that could lead to differences between samples taken from the same male at different time points, as well as between different males – calendar week, male age and recent inclusion in a mating group.

'Calendar week' is independent of year in order to separate seasonal variation linked to the time of year from differences due simply to temporal separation of samples, and the week containing the first of January is week 1. 'Cohort' groups together birds that hatched in the same year, whereas 'age' is the number of years since a male hatched at the point when the sample was collected. This means samples from the same male taken in different seasons will be categorised as the same cohort, but different male age. Similarly, different males from the same cohort may be sampled at different ages. The distinction between cohort effects and age effects attempts to separate effects relating to the ageing process from stochastic variation associated with year of hatching, e.g. due to the fact that all birds hatched on the same year might have experienced the same developmental conditions.

Though mating effects are considered in more detail in later analysis, we included 'mating treatment' as a categorical variable indicating whether or not a male had been included in a mating group during the breeding season, and prior to, when the sample was collected.

Four MHC metrics were used: for each of the two loci (BF and BLB) allelic richness (the total number of alleles in an individual's genotype) and allelic divergence (the difference between alleles present in a genotype in terms of amino acids) were calculated.

A full linear mixed model was therefore constructed as follows, with the male identity (BirdID) and the date of sample collection (CollectionDate) included as random factors:

Response variable ~ Age*CalendarWeek + Age*CalendarWeek^2 + Cohort*CalendarWeek + Cohort*CalendarWeek^2 + MatingTreatment + BFAllelicRichness + BLBAllelicRichness + BFAminoAcidDivergence + BLBAminoAcidDivergence + (1|BirdID) + (1|CollectionDate)

Given the large number of terms in the full model, we used a Bayesian Information Criterion to select the best model. We retained a term if the model with that term gave a BIC value greater than two lower than the equivalent model without that variable, and if the null model was not within two BIC of the best model. The resultant 'best' model was then run using the 'lme' function in the 'lme4' package to provide coefficients for each term.

Simpson diversity is bounded by zero and one and was square-root, arc-sine transformed. The box cox function in R package 'MASS' was then used to determine the transformation lambda value required to ensure data for the five community composition measures met the assumptions of the linear mixed model, according to diagnostic plots.

Given the range in the scales of values between different variables in the model, they were centred and scaled using the R command 'standardise', and the model set to compare all possible models with all combinations (including absences) of the above variables is generated and models compared using the R command 'dredge'.

The effects of the levels of each significant treatment were then determined through post-hoc comparisons using the lsmeans function in the 'lmerTest' package and inputting the linear mixed model including all significant terms.

To establish whether there are groups of OTUs which occur in the same samples at comparable relative abundances we performed hierarchical clustering using the complete linkage method on Euclidean distances between OTU occurrence profiles (i.e. the relative abundance of OTUs in all the samples). Spearman's correlations between these profiles calculated was then calculated and visualised using a heatmap, with a dendrogram produced from the clustering. Clusters of relevance were then identified using this graphic. We then assessed whether these groups of OTUs may be linked to differentiation of samples along the Principal

Component axes using the correlation of the relative abundances of each of the OTUs in a cluster with the average PCA axis scores of the samples within which they occur, as well as the correlation with average allelic richness and amino acid divergence at each of the two MHC loci.

Mating Behaviour as a Driver of Variation between the Seminal Microbiotas of Males

The experimental mating trials could change the seminal microbiota of a male through non-sexual exposure to females and mating opportunities that create potential for sexual transmission. We use samples collected from 39 males following their inclusion in one of the six trial groups to investigate the effect of inclusion in a given mating trial group, and mating interactions within the group, on male seminal microbiota. We use the same five community metrics as in the previous analysis; species richness, Simpson diversity and Shannon diversity were calculated as before, and a new Principal Components Analysis was performed on a subset of 29 OTUs occurring in at least 10% of these males and at a relative abundance equal to or greater than 0.1%. The first axis (PC1) explained 17% of the variation between samples, and the second axis (PC1) explained 12% of variation (Figure 2). These community metrics were then used as response variables in statistical models. Box cox transformations were performed to ensure these response variables which meet the assumptions of the models used.

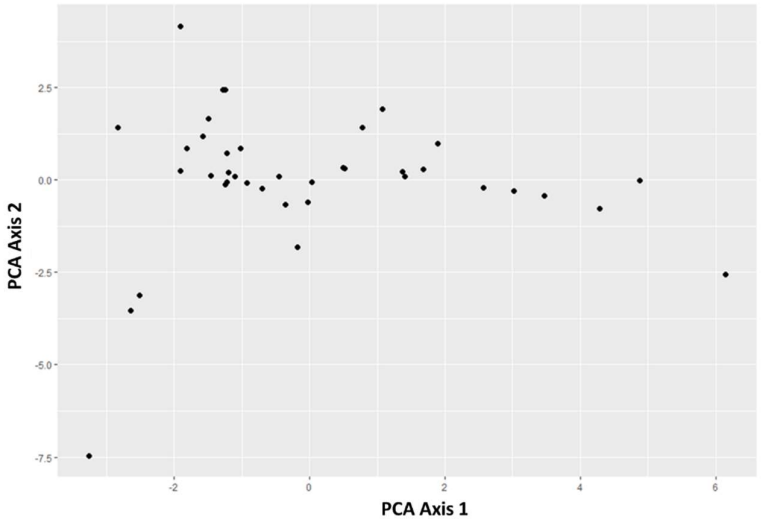


Figure 2 *The first two principal component axes calculated across 39 semen samples from 39 male red junglefowl following inclusion in mating trials. Principal components were calculated using a subset of 29 OTUs selected because they occur in at least 10% of samples and at an average relative abundance equal to or greater than 0.1%*

a) Socio-sexual Environment and Variation in Seminal Microbiota Community Structure and Composition between Males

We investigated whether differences between males could be predicted by inclusion in a mating trial group, either due to exposure to the specific group of females or group inclusion in general though exposure to the same males and females.

We predict that there will be differences between the different groups in the community metrics of samples taken from males, due to co-habitation, transmission across the group, or because of a group characteristic leading to a change in the microbiota community characteristic. These effects could be due to female group or mating trial group.

We investigated whether there was an effect of female unit, and mating group within female unit where the same unit of females was used in two different mating groups. A linear model where the explanatory variable is mating group identity nested within female unit identity was run with the lm function in R, with each of the five community metrics as response variables, and the 'anova' command used to identify variables contributing significantly to variation in these variables.

b) Investigating the effect of Group Sexual Network Structure on Variation in Seminal Microbiota Community Structure and Composition between Males

To investigate whether variation in composition and community metrics between males in different mixed-sex mating groups was predicted by the sexual network characteristics of that group, we calculated spearman rank correlations between group network characteristics and measures of variation in the microbiota community structure and composition within each group, according to the five community metrics calculated and used in earlier analyses, species richness, Shannon diversity, Simpson diversity, PC1 and PC2, and the Bray-Curtis Dissimilarity.

We consider four metrics used to summarise the interactions which comprise a sexual network of a group of individuals identified as useful measures of population social network measures for studying disease ecology (Silk *et al*., 2017):

1. Average polyandry - the average number of unique mates of the females in the group.

2. Modularity – a measure of how much the network is split into smaller subgroups with little interaction between these subgroups.

3. Density - the number of connections between group members as a proportion of the total number of possible connections.

4. Mean path length - the average number of connections required to link any member of the group with any other group member.

We hypothesise that if sexual interaction affects microbiota composition then males in groups with greater average polyandry, lower modularity, higher density and lower mean path length will have more similar microbiotas as they are more likely to have mated with the same females, and there is more opportunity for transmission of microbiota and components influencing microbiota between individuals with more connections and a shorter path length meaning a more direct potential transmission route for any bacteria.

Variation of sample microbiotas within a group was quantified through calculation of the standard deviations of the Principal Component Axes 1 and 2 (PC1 and PC2), the coefficient of variation of species richness, Shannon diversity and Simpson diversity, and the average Bray-Curtis Dissimilarity between pairs of samples. The standard deviation is calculated as the variance divided by the number of data points minus one. The coefficient of variation is calculated by the group standard deviation, divided by the group mean and multiplied by 100 and so represents the variation in community metrics in the group proportional to the average value and expressed as a percentage, used where variation in a metric co-varies with the mean. The Bray-Curtis Dissimilarity quantifies the compositional dissimilarity between two samples based on the

relative abundance of the OTUs shared at both sites (shown to be a robust measure of dissimilarity between ecological sites (Faith *et al*., 1987) and used in studies of microbiota (e.g. Koren *et al*., 2013). Standard deviations and coefficientsof variation in community metrics were calculated using the 'stats' package in R and Bray-Curtis Dissimilarity using the 'vegan' package.

Each group sexual network was represented by an unweighted male-female copulation matrix, which was created for each group from the behavioural observations of copulations observed between all the potential male-female pairs in the group throughout the full 10 days of the trial.

For each group an unweighted matrix was composed of 120 data points, one for each potential male-female pair, zeros and ones indicating whether or not a pair mated at all during the 10-day trial. Only copulations as described above (no other interactions) were used to define the sexual network.

Average group polyandry was calculated from the unweighted matrix by counting the number of males each female in a group had mated with, and then averaging this across all females in a group.

Network density for each group was calculated from the unweighted matrix using the 'graph from edgelist' followed by the 'edge density' function of the 'igraph package' in R.

Mean path length for each group was calculated from the unweighted matrix using the 'mean distance' function of the 'igraph package' in R.

The modularity of each group was calculated from the unweighted matrix using the 'graph from edgelist' function followed by the 'cluster_edge_betweenness' function and then the 'modularity' function, all from the 'igraph' package in R.'

As might be expected there was co-variation between some of these sexual network metrics, with complete correlation of groups ordered by average polyandry and density (Appendix, Fig. A1).

Given there are only six mating groups there is not enough power for a formal significance analysis, but Spearman's Rank Correlations between each of the group level metrics with the six group level community metrics were calculated using the 'cor.test' function in the basic 'stats' package in R, and relationships visualised by plotting using the 'ggplot2' package in R.

c) The effect of Individual Mating Behaviour on Variation in Seminal Microbiota Community Structure and Composition between Males

We investigated the effect of measures of individual mating behaviour on the composition and structure of a male's seminal microbiota, and whether any effect of individual mating behaviour varies within group.

The following mating metrics were calculated for each male:

1. Number of mates (NoMates): the number of females (mating partners) with whom a male mated successfully (i.e. mating success)

2. Unique males (UniqueMales): the number of other males to whom a focal male is exposed by mating with the same females (i.e. the number of males with whom a focal male shares his mating partners). 'Unique' indicates that a male is counted only once even if he has mated with more than one of the focal male's partners.

3. Mates' average matings (MatesAverageMates): the average number of matings of all the mating partners of a focal male (including the focal male himself)

If mating affects microbiota composition, either through transmission or disruption, and these effects are cumulative we might expect variation in community metrics to relate to these measures reflective of the amount of mating related transmission and disruption a male is exposed to either through his own mating patterns, or that of his female partners.

For instance, the higher value of these metrics may result in higher microbiota diversity and richness through being exposed to more OTUs from more individuals, or more disrupted microbiotas.

Similarly, we might expect individuals who have mated less to be more similar to each other than those who have mated more as their microbiotas are less disrupted and represent a 'core' seminal microbiota. Conversely, increased mating, or mating with female who have more mates, may lead to increased similarity if more diverse males become more similar through being more likely to have mated with the same females.

Further, these relationships might vary between mating groups, for instance in a group with high polyandry, relatively more successful males (i.e. more mates) may be much more similar to each other than relatively successful males in weakly polyandrous groups, because they are more likely to be connected through the same females.

These metrics were calculated using custom scripts in R, and calculated to give cumulative values for each bird on each day of the mating trial group. Mating interactions accumulated over time follow a broadly similar cumulative pattern, although the rate at which this happened varied considerably across individuals in a group and across mating groups (Appendix, Fig. A2).

We investigated the significance of each of these individual mating metrics through information-based model selection as described for exploring potential sources of population variation in the previous section, except that model selection is now based on the Aikake Information Criterion, rather than Bayesian, as there are fewer terms in the full model than the previous analysis.

The full model here, with mating group identity (MatingGroup) included as a random effect, is:

ResponseVariable ~ MatesAverageMates + NoMates + UniqueMates + (1|MatingGroup)

An interaction between mating group and each of the three mating metrics, may mask an overall effect of the metrics in the previous model. This was then investigated with an ANOVA of the following linear model including mating group identity as a fixed effect rather than a random effect:

ResponseVariable ~ MatesAverageMates*MatingGroup + NoMates*MatingGroup + UniqueMates*MatingGroup + MatingGroup

However, the diagnostic plots for these revealed an insufficient error structure for confident analysis, presumably due to an insufficient number of males per mating group and thus degrees of freedom. We therefore performed a similar analysis but now only including an interaction with the variable we considered to have the most biological likelihood for an interaction; unique males. The diagnostic plots are now much improved and thus ANOVA was performed on the following model:

ResponseVariable ~ MatesAverageMates + NoMates* + UniqueMates*MatingGroup +MatingGroup

To allow for a potential lag in the effect of mating on microbiota composition two separate analyses are performed, one using cumulative values of the above calculated for day 5 (only matings up to and including day 5) and one using values from day 10 (all matings throughout the duration of the mating trial) as explanatory variables.

d. The Influence of Sharing Mates on Pairwise Compositional Dissimilarities between Male Seminal Microbiotas

We tested the effect of the number of females shared by two males (i.e. mated with both males) on the similarity of these males' seminal microbiotas following the mating trial. If a mating effect is specific to the individual that a male mates with then we would expect males who have mated with the same mates to be more similar to each other than pairs of males who have not shared the same mates.

We investigated this by calculating three compositional dissimilarity metrics for each pair of males within each mating group, and looked for an effect of number of mates shared on these measures. The compositional dissimilarity metrics are the Bray-Curtis Dissimilarity, as previously calculated, and the difference in each of the two PCA axes between two samples.

The number of females shared by each pair of males in a mating group was calculated using male-male matrices created from the unweighted male-female copulation matrices described earlier in these methods. The male-female copulation matrices were converted into male-male matrices using the 'as.one.mode' function from the 'bipartite' package, and were weighted by the number of females shared. These matrices, one for each mating group, illustrated connections between males via a female that they had both mated at some point during the 10-day mating trial, and were composed of 45 data points, one for each potential male-male pair, and the data point for each pair is the number of 'shared' females i.e. the number of females with which they have both mated.

These data points for male pairs from all six mating groups were then added into the dataset of males with compositional dissimilarity metrics, to give a dataset of male pairs with information on females shared and compositional dissimilarity for each pair.

To establish the effect of each of numbers of females shared upon compositional dissimilarity of male samples, models with compositional dissimilarity as the response variable and mating group as a random factor were run using the lme function in the 'lme4' package and compared using the ANOVA function in the standard 'stats' R package. A p-value generated by the function using log-likelihood ratios from two models with and without numbers of shared females as a fixed factor, and as either a linear or quadratic effect to estimate the significance of the extra variation explained by number of shared females.

RESULTS

Variation of The Seminal Microbiota Across a Population of Red Junglefowl

Here potential sources of variation in the composition and structure of the red junglefowl seminal microbiota are simultaneously investigated. All community metrics exhibited substantial variation among males, as well as fluctuation within males across the mating season (Fig. 3).

Clusters of Co-Occurring OTUs

Hierarchical clustering of OTUs by occurrence profiles was visualised using a dendrogram, and Spearman's correlations between these occurrence profiles visualised using a heatmap. This revealed five OTU clusters that could be identified visually, by looking for clusters of darker red, matching clusters of branches on the dendrogram (Fig. 4). Four of these clusters were formed of OTUs with consistent correlation to one or both of the PCA axes, suggesting relative abundance of these co-occurring OTUs is driving the main axes of compositional variation between ejaculate samples.

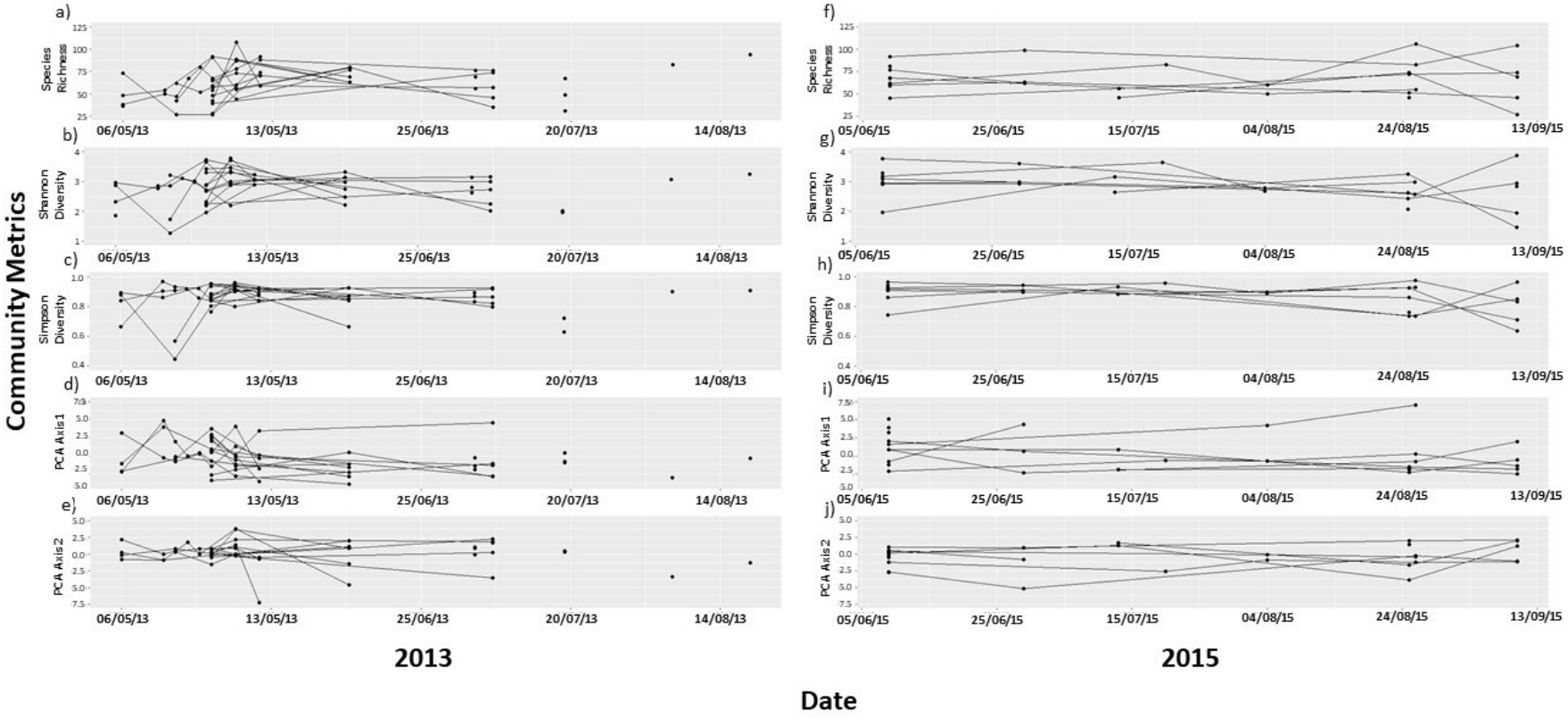


Figure 3 *Community metrics of the seminal microbiota of each male across the 2013 and 2015 breeding seasons showing: a) and f) species richness; b) and g) Shannon diversity; c) and h) Simpson diversity; d) and i) PCA Axis 1 scores; e) and j) PCA Axis 2 scores. Lines connect samples taken from same male at different time points.*

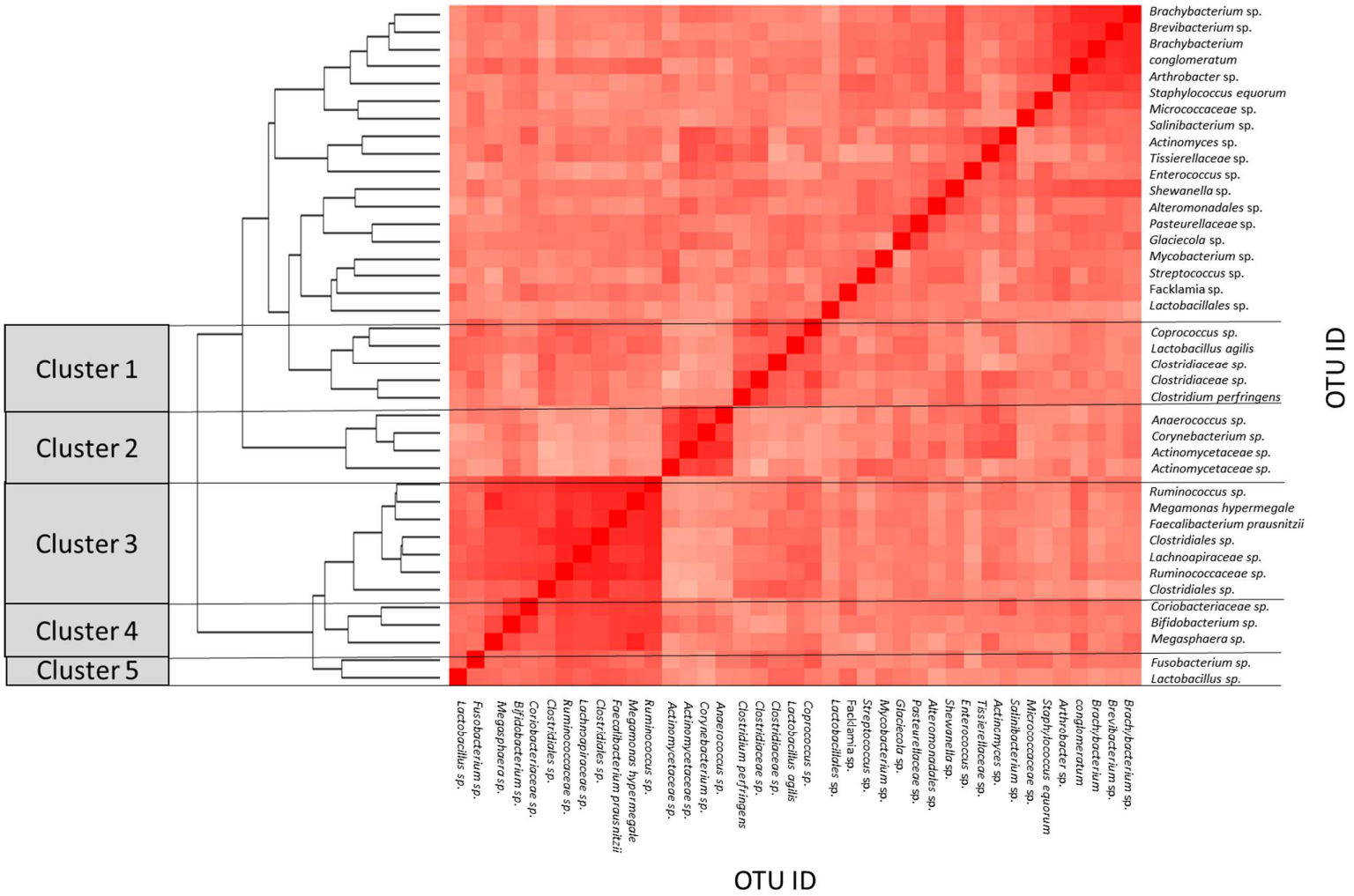


Figure 4 *Dendrogram representing hierarchical Clustering and heat map illustrating Spearman Correlations of OTU occurrence and abundance in samples. Intensity of red demonstrates increased correlation. Clusters are defined by looking for clusters of intense read and matching to branching patterns of the dendrogram. They are arbitrarily numbered.*

Exploring the Relative Influence of Different Potential Drivers of Variation in Seminal Microbiota Structure and Composition

Model comparisons based on the BIC revealed no evidence that any of the explanatory variables influenced species richness, Shannon diversity, Simpson diversity and PC2, as for these variables the null model was within 2 BIC of the best-ranking model (Table 1). However, for PC1 the best model included MHC C2 amino acid divergence and C2 allelic richness (Table 2). PC1 was associated negatively with allelic richness and positively with amino acid divergence (Fig. 5). Regarding random effects, in this model no variation is explained by identity of the male and 6.80% of total variation is explained by collection date.

The association of the BLB MHC locus with PC1 was further explored through correlations of the relative abundance of OTUs identified in clusters in Fig. 4 with PC1 BLB allelic richness and BLB amino acid divergence. The relative abundances of OTUs in cluster 2 were the four most negatively associated with PC1, and were all positively associated with BLB allelic richness. The relative abundances of those OTUs in cluster 3 were all positively associated with PC1, negatively associated with BLB allelic richness and positively associated with BLB amino acid divergence. The relative abundances of OTUs in Cluster 4 were all positively correlated with PC1 and positively associated with BLB allelic richness (Table 1).

Table 1 *The correlation of the relative abundances of the OTUs in each cluster with PCA axis 1, BLB allelic richness and BLB amino acid divergence.' Both' indicates that some correlations are positive and some negative. There is no significance measure of correlations.*

Cluster	No. OTUs in Cluster	Correlation with PCA Axis 1	Correlation with BLB Allelic Richness	Correlation with BLB Amino Acid Divergence
1	5	Both	Both	Both
2	4	Negative	Positive	Both
3	7	Positive	Negative	Positive
4	3	Positive	Both	Positive
5	2	Both	Negative	Negative

Table 2 *The best five models for each community metric according to BIC-based model selection. The models considered to explain significantly more variation in the data than the null are indicated using bold italics.*

Species Richness		Shannon Diversity		Simpson Diversity		PCA Axis 1		PCA Axis 2	
Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC
Null	0	Null	0	Null	0	BLB Amino Acid Divergence + BLB Allelic Richness	0	Null	0
Cohort	2.64	BF Allelic Richness	1.22	BF Allelic Richness	1.21	Cohort + BF Amino Acid Divergence	2.23	Age	1.02
BF Allelic Richness	2.84	Cohort	3.35	BF Amino Acid Divergence	3.13	Cohort + BLB Amino Acid Divergence + BLB Allelic Richness	3.07	BF Allelic Richness	1.90
BLB Allelic Richness	3.73	BF Amino Acid Divergence	3.37	Cohort	3.41	Null	3.55	Cohort	2.49
BLB Amino Acid Divergence	3.73	BLB Allelic Richness	4.09	Calendar Week^2	3.78	BLB Amino Acid Divergence + BLB Allelic Richness + Calendar Week	3.99	BLB Allelic Richness	2.88

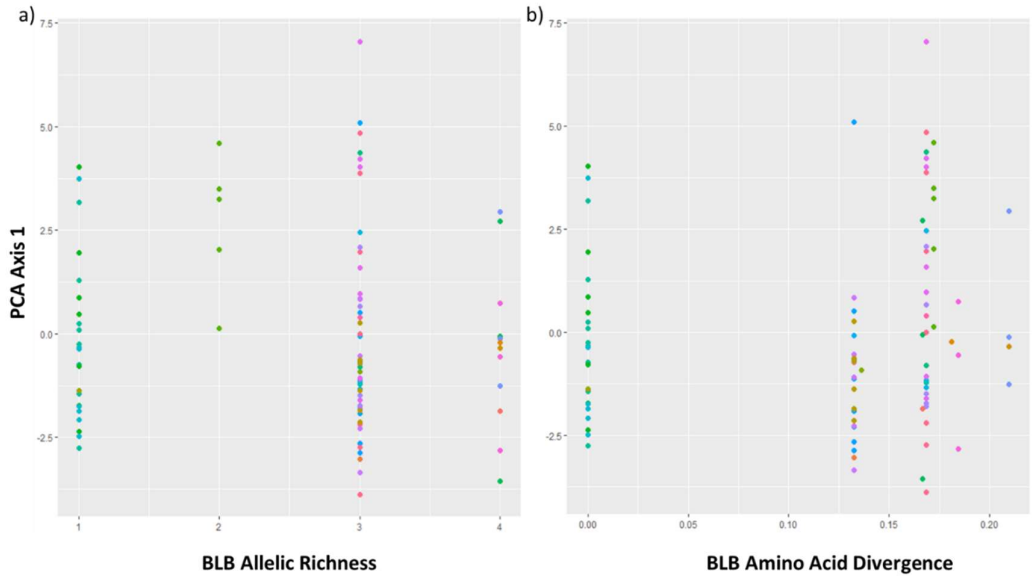


Figure 5 *Variation in un-transformed PC1 values with a) Allelic Richness at the BLB Locus b) Amino Acid Divergence at the BLB Locus. Colours indicate male identity.*

Mating Behaviour as a Driver of Variation between the Seminal Microbiotas of Males

a) Socio-sexual environment and Variation in Seminal Microbiota Community Structure and Composition between Males

We first investigated whether there was an effect of female unit, and mating group within female unit. None of the five microbial community metric response variables were affected by female group. However, variation in Shannon diversity and Simpson diversity were significantly explained by differences across mating groups within female group (Shannon: $F_{(2,33)} = 5.79$, $p = 0.009$; Simpson: $F_{(2,33)} = 6.06$, $p = 0.007$). Males in mating groups 2013 B and 2013 D were kept with the same unit of females (Female Unit F2), but the seminal microbiotas of the males in mating group 2013 B had on average lower Simpson and Shannon diversity than those produced by males in mating group 2013 D (Fig. 6b and 6c).

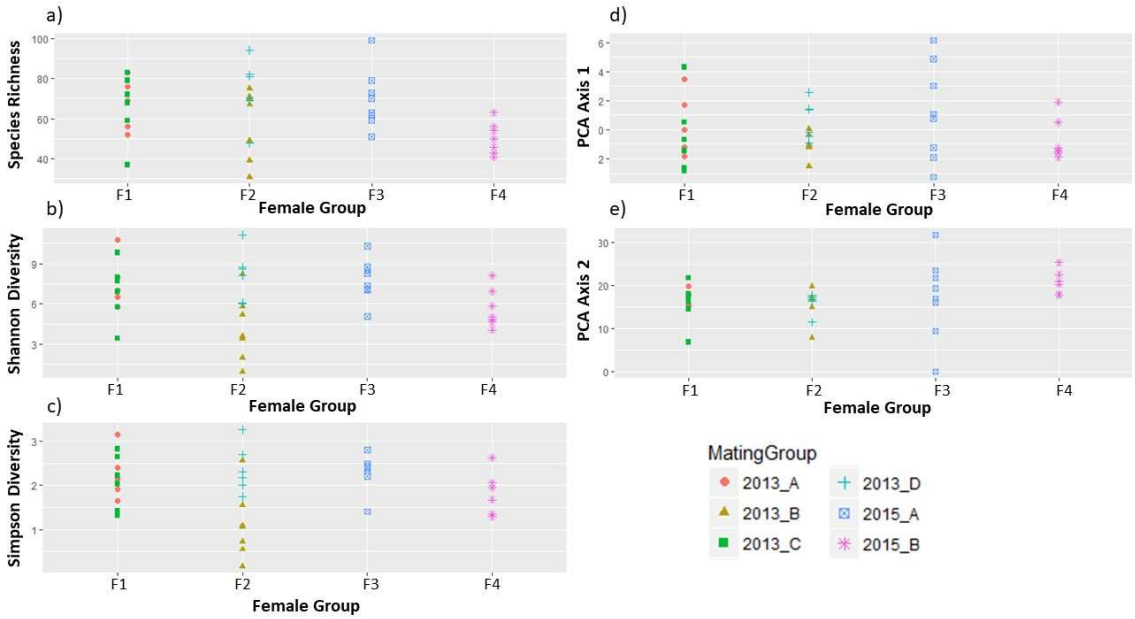


Figure 6 *Community metrics for each male sample grouped by female unit (F1 – F4), with colours and shapes indicating mating group. Transformed values have been plotted for Shannon diversity, Simpson diversity and PCA Axis 2*

b) The effect of Group Sexual Network Structure on Variation in Seminal Microbiota Community Structure and Composition between Males

We investigated whether, within a mating group, the similarity in seminal microbiota composition and community metrics between two males was predicted by the structure of the sexual network of the group, i.e. average group polyandry, network density, network modularity and average path length. There was variation between mating groups in all four network measures. Average group polyandry varied from approximately 4 to approximately 7, though average path length was always very short (between 1 and 2), modularity tended to be low (between 0 and 0.1) and density was also fairly low (between 0.2 and 0.35).

There was no strong evidence of correlation between microbiota community metrics and the structure of the sexual network of the group, though mean path length was moderately positively correlated with species richness (Fig. 7 and Table 3).

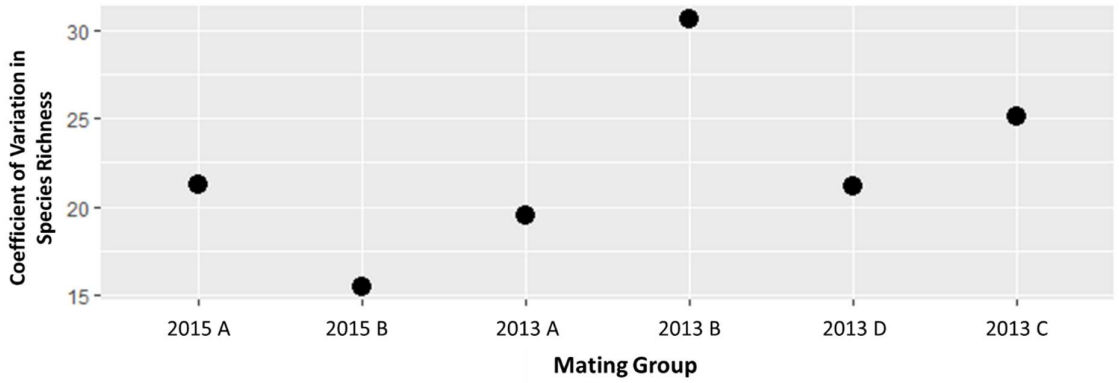


Figure 7 *Group level variation in coefficient of variation in species richness ordered by increasing values of mean path length in each group.*

Table 3 *Correlations of Group Level Community Metrics with Group Level Mating Metrics. Stars highlight those we consider moderate* or strong** correlations. Moderate is correlation >0.5 or <-0.5, strong is correlation >0.8 or <-0.8. No significance is computed for these correlations due to there only being six mating groups.*

	Average Group Polyandry	Network Density	Network Modularity	Network Mean Path Length
Average Bray Curtis Dissimilarity	0.143	0.087	-0.429	0.334
Coefficient of Variation in Species Richness	-0.086	-0.145	-0.257	0.577*
Coefficient of Variation of Shannon Diversity	-0.429	-0.377	-0.029	0.395
Coefficient of Variation in Simpson Diversity	-0.429	-0.377	-0.029	0.395
Standard Deviation in Principal Component Axis 1	0.086	0.116	0.371	-0.213
Standard Deviation in Principal Component Axis 2	-0.257	-0.232	0.200	0.152

c) The Effect of Individual Mating Behaviour on Variation in Seminal Microbiota Community Structure and Composition between Males

According to AICc based model selection, no mating metrics significantly explained variation in any microbial community response variable, as the null model is always within 2 AICc of the best model, if not the best model itself. This was the case for cumulative mating metrics calculated at both day 5 (Table 4) and day 10 (Table 5), and the lack of explanatory power remains even when nine males who did not mate at all are removed from the analysis (Table 6 and Table 7).

Table 4 *The best five models for each community metric according to AICc-based model selection when nine non-mated males are included. The models considered to explain significantly more variation in the data than the null are indicated using bold italics. Day 5. (NoMates = Number of mates; MatesAv = Mates' average matings; Unique = Unique males).*

Species Richness		Shannon Diversity		Simpson Diversity		PCA Axis 1		PCA Axis 2	
Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc
Null	0	Null	0	Null	0	Null	0	NoMates +Unique	0
MatesAv	0.976	MatesAv	1.91	Unique	2.46	MatesAv	2.10	Unique	0.130
Unique	1.00	Unique	1.93	NoMates	2.49	NoMates	2.41	Null	0.574
NoMates	2.13	NoMates	2.37	MatesAv	2.49	Unique	2.49	MatesAv	1.04
MatesAv+No Mates	2.37	NoMates +Unique	4.29	NoMates +Unique	4.92	MatesAv +Unique	3.73	MatesAv +NoMates +Unique	2.70

Table 5 *The best five models for each community metric according to AICc-based model selection when nine non-mated males are included. The models considered to explain significantly more variation in the data than the null are indicated using bold italics. Day 10. (NoMates = Number of mates; MatesAv = Mates' average matings; Unique = Unique males).*

Species Richness		Shannon Diversity		Simpson Diversity		PCA Axis 1		PCA Axis 2	
Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc
Null	0	Null	0	Null	0	Null	0	Unique	0
Unique	0.655	Unique	1.62	Unique	2.39	Unique	2.43	Null	0.349
MatesAv	0.910	MatesAv	1.79	MatesAv	2.44	MatesAv	2.45	MatesAv	0.911
NoMates	2.05	NoMates	2.10	NoMates	2.47	NoMates	2.47	NoMates	1.18
NoMates +Unique	2.28	NoMates +Unique	4.14	NoMates +Unique	4.97	NoMates +Unique	4.57	NoMates +Unique	2.55

Table 6 *The best five models for each community metric according to AICc-based model selection when nine non-mated males are removed. The models considered to explain significantly more variation in the data than the null are indicated using bold italics. Day5. (NoMates = Number of mates; MatesAv = Mates' average matings; Unique = Unique males).*

Species Richness		Shannon Diversity		Simpson Diversity		PCA Axis 1		PCA Axis 2	
Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc
Null	0	Null	0	Null	0	Null	0	NoMates	0
NoMates	2.13	Unique	1.49	Unique	0.499	Unique	0.840	Null	0.989
MatesAv	2.17	NoMates	2.26	MatesAv +Unique	1.77	MatesAv	1.82	NoMates +Unique	1.77
Unique	2.48	MatesAv	2.26	NoMates	1.95	NoMates	2.36	MatesAv +NoMates	2.88
MatesAv +Unique	4.00	MatesAv +Unique	2.38	MatesAv	2.59	MatesAv +Unique	3.70	MatesAv	3.35

Table 7 The best five models for each community metric according to AICc-based model selection when nine non-mated males are removed. The models considered to explain significantly more variation in the data than the null are indicated using bold italics. Day10. (NoMates = Number of mates; MatesAv = Mates' average matings; Unique = Unique males).

Species Richness		Shannon Diversity		Simpson Diversity		PCA Axis 1		PCA Axis 2	
Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc	Top Five Models	Δ AICc
Null	0	Null	0	Null	0	Null	0	Null	0
NoMates	2.50	Unique	2.36	Unique	1.40	Unique	0.541	NoMates	1.56
MatesAv	2.65	NoMates	2.49	NoMates	2.09	NoMates	1.33	MatesAv	2.47
Unique	2.65	MatesAv	2.67	MatesAv	2.68	MatesAv	2.52	Unique	2.66
NoMates +Unique	5.21	MatesAv +Unique	5.01	MatesAv +Unique	3.65	NoMates +Unique	3.18	NoMates +Unique	3.85

We included mating group as a fixed factor in a linear model with the three mating metrics, in an interaction with Unique Males, to allow for any variation between groups in the relationship between Unique Males and microbiota composition. There was no significant interaction between Unique Males and mating group in explaining variation in any of the community metrics, though again there is a significant effect of mating group as a fixed factor.

d) The Influence of Sharing Mates on Pairwise Compositional Dissimilarities between Male Seminal Microbiotas

We next explored the effect of the number of females shared (i.e. mated) by two males on the similarity of these males' seminal microbiotas following the mating trial. ANOVA comparisons based on the log likelihood of linear mixed models with mating group as a random factor and numbers of shared females as a fixed factor revealed no significant effect of the number of female mates shared by a pair of males on any of the dissimilarity measures, as either a linear or quadratic effect. When mating group was included as a fixed factor no interaction was significant but again mating group itself explains significant variation in all three pair dissimilarity measures, Bray Curtis, PC1 difference and PC2 difference (Table 8). As can be seen in Fig. 8. mating Group 2013_B has a significantly higher Bray Curtis Dissimilarity (Fig. 8a), 2013_B and 2015_B a significantly lower PC1 difference (Fig. 8b) and Mating Group 2015_A a significantly higher PC2 difference (Fig. 8c).

Table 8 *Table illustrating the effect of mating group on pairwise compositional dissimilarities between all male pairs in a group, with estimates of those groups who are significantly different.*

Compositional Dissimilarity Measure	Significance of Mating Group	Group Comparisons
Bray-Curtis Dissimilarity	$F_{(5,60)}$ = 3.29 p = 0.011	2013_B est = 0.140, p = 0.044
PC1 Difference	$F_{(5,60)}$ = 8.27 p < 0.001	2013_B est = -0.561, p = 0.002; 2015_B est =-0.403, p = 0.049
PC2 Difference	$F_{(5,60)}$ = 6.05; p < 0.001	2015_A est = 0.546, p < 0.001

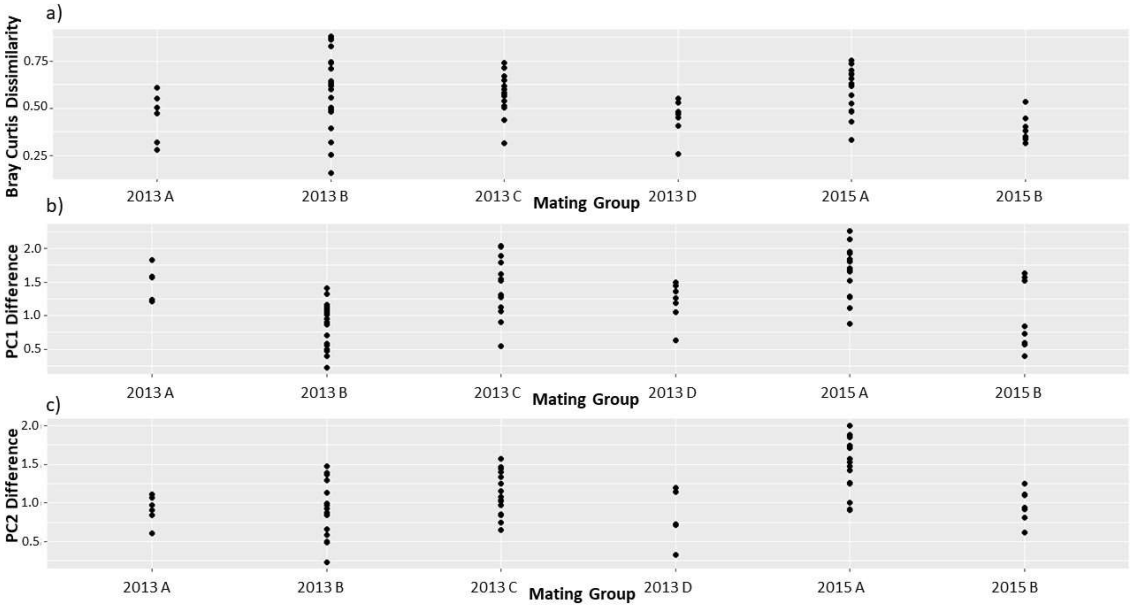


Figure 8 *Between group differences in the three compositional dissimilarity measures of all male pairs in a mating group. by a) Transformed Bray Curtis dissimilarity b) Transformed PC1 Difference c) Transformed PC2 Difference*

DISCUSSION

In the current study we explore sources of inter- and intra- individual variation in the seminal microbiota of the red junglefowl across two breeding seasons, considering a seasonal and temporal effect, age, and a potential immuno-genetic determinant of microbiota composition. We then specifically investigate the potential effect of mating on seminal microbiota using four groups of males used in a mating trial experiment.

We found variation in community structure and composition of an individual's microbiota over time, and between individuals, and there are clusters of co-occurring OTUs that seem to be driving the main axes of compositional variation across semen samples. The immuno-genetic metrics of amino acid diversity and allelic richness at the BLB locus predicted variation along the Principal Component Analysis axis explaining the greatest amount of variation (13 % of total variation) among samples.

In a subset of males subjected to mating trials, diversity of samples was predicted by which mating trial group the male was included in, as was compositional dissimilarity between pairs of males. There was limited evidence that variation in community composition metrics in a group was correlated with the sexual network metrics of a group.

Exploring the Relative Influence of Intrinsic and External Drivers of Variation of the Seminal Microbiota across a Population of Red Junglefowl: A Potential Immunogenetic Determinant of Variation

We found that the model which was significantly best at predicting variation in sample PC1 scores was that just including the immunogenetic metrics, amino acid diversity and allelic richness at the BLB locus. This is consistent with immune-mediated determination of microbiota components whereby the alleles at the BLB locus affect the response of the individual junglefowl's immune response to the presence of microbial taxa present in their body via the immune molecules expressed, allowing some bacterial species to remain and colonise the host whilst clearing others. Further, this effect is restricted to the BLB locus and not the BF locus, which is what we predicted as BLB mediates extracellular pathogens such as bacteria, whereas the BF locus is concerned with intracellular pathogens.

The result supports work by Bolnick *et al*., (2014) who found that MHC Class II diversity and richness predicted gut microbiota richness and diversity in sticklebacks, and that certain microbial taxa were linked to the presence of certain genotypes and alleles, in that we found the junglefowl equivalent of MHC II affects composition. However, beyond the broad conclusion of an

MHC II mediated microbiota composition, our results differed; we did not find an effect upon richness and diversity and there was no effect of genotype. The fact that we detect no effect on diversity and richness of community composition could be due to the fact that the seminal microbiota is likely made up of different microbiota from multiple sources, and so whilst the presence of certain alleles could be as a result of mediation on one or more of the microbiota that are still present and at similar relative abundances once these are combined in the ejaculate, any effect on richness and diversity of composition is lost at this stage, perhaps due to the addition of differing numbers of rarer OTUs. Further, it could even be that MHC mediation only acts on those most common OTUs, and so again any effect is not acting upon rarer OTUs, which do affect measures of richness and diversity. Alternatively, it could be that the MHC mediation results in differing compositions of OTUs, but not diversity or richness, because different OTU configurations can achieve the same levels of diversity (as in vaginal microbiotas, Ravel *et al*., 2011).

Lack of a genotype effect suggests it is diversity rather than specific alleles which lead to selection. This could be because those alleles, or sets of alleles, targeting certain bacteria or groups of bacteria are more likely to occur at higher allelic richness or amino acid diversity. One would of course still expect a genotype effect in this case, but as we have more repeats at the same allelic richness and amino acid diversity, with few repeats of the same genotype, it could be that our data does not provide the power to detect an additional effect of genotype, once variation due to allelic richness and amino acid diversity has been taken into account. Alternatively, it could be that a diverse MHC leads to a stronger immune response towards all alleles, and so allows more of the existing selections for and against microbiota taxa to be realised as diversity increases.

Further work with a larger sample size and more repeats of genotypes might be able to tease apart these complex effects. We were also unable to fully explore individual taxa and allele associations. We did however find that there were clusters of OTUs that were correlated with

MHC metrics further supporting the idea that genetic variation is leading to selection for and against certain bacterial taxa.

It could be that this association between the MHC and microbiota could be reflective of a non-immune genetic determinant of microbiota composition, if variation at the MHC genes is reflective of wider genetic variation. If this is the case, a genetic predictor of microbiota composition is still of interest as it would not negate theories based upon the potential heritability of microbiotas, and would thus still mean that selective pressures could shape microbiota composition over generations, and the microbiota could still act as an indicator of genetic quality or relatedness. The microbiota could still contribute to mechanisms for apparently MHC based mate choice in junglefowl (Løvlie *et al*., 2013) and badgers (Sin *et al*., 2015), add understanding to microbiota based kin recognition (Lizé *et al*., 2014) and the consequences of an effect of microbiota on scent gland secretions (Theis *et al*., 2013; Leclaire *et al*., 2017).

Nevertheless, the fact that MHC Class II, and not Class I, has an effect suggests it is an immune-mediated effect as this is the association we would predict given the role of MHC II, and not MHC I, in recognising extracellular organisms such as bacteria. We know of no reason why wider genetic variation would be reflected in MHC Class II variation but not MHC Class I. Further work clarifying this would be of interest.

This work is also of interest when considering genetic determinants of susceptibility to disease; this could be due to a common genetic determinant could be that link is actually an interplay between immunogenetic effect, the healthy microbiota and the pathogen. A genetic determinant of microbiota also needs to be considered when considering the effect of therapies based on microbiota manipulation such as faecal transplants and probiotics; here the effect may vary or even be negated by population variation in genetic determinants of microbiota composition. Our work may be particularly relevant in the field of fertility treatments, if fertility could be determined by bacteria present, which is genetically determined, explaining compatibility of males and females based on MHC. This may even explain why in badgers (Sin *et*

al., 2015) and wolves (Galaverni *et al*., 2016) MHC-disassortativity was observed between parent pairs despite predictions as to the opposite; more MHC dissimilar pairs may be less fertile if the male introduces bacteria into the female which her immune system reacts to, with harmful effects on sperm, ova or subsequent embryos.

Investigating Mating as a Driver of Variation between the Seminal Microbiotas of Males

We found that both Shannon and Simpson diversity of the seminal microbiota was predicted by which mating group trial the male was a member of, but not the group of females to which the group was exposed. Pairwise differences between male seminal microbiota composition was also predicted by group, including Bray-Curtis dissimilarities and differences in scores on the first two principal component axis. There was a moderate correlation between the group coefficient of variation in species richness and the group mean path length.

These results are not inconsistent with the hypothesis that male seminal microbiotas are disrupted by mating, potentially as a result of transmission of microbiota members, with sexual transmission of bacteria from males to females in avian species already demonstrated in the zebra finch (Kulkarni & Heeb, 2007).

Similarities in community structure, including ecological diversity could be as a result of a mating effect for a number of reasons; firstly because of transmission of either the community from one male to another via a female, or a factor leading to that diversity. Transmission from the female herself could also be a possibility, but given that female group does not predict variation, it is not solely exposure to the same females that would lead to male similarity, but it could be exposure to the same females at the same time, if female reproductive microbiotas fluctuate, or a combination of exposure to the same males and females. However, it does not seem to be the case that it is simply transmission of microbiota members that results in this similarity; variation along the PC axes, which captures variation in microbiota members, is not predicted by mating group. Further, similarity of microbiota composition between males is not predicted by males that have

mated with the same females a similar number of times neither in terms of diversity nor in terms of compositional measures.

Alternatively, then, an effect of mating group could be due to mating affecting male microbiota independent of who they have mated with, or who their mates have mated with, but such that the nature or amount of mating rather than identity of the mate is what is important.

This could lead to differences in groups as a result of variation in the nature and amount of mating between groups. Indeed, species richness was positively correlated with mean path length. However, the number of females a male mates with, the number of males he is exposed to via these females, or the average number of matings that females had did not affect community structure or composition even when group effects were taken into account.

An effect of network structure on bacteria reflects a study considering transmission dynamics of a single bacteria in social spiders (Keiser *et al*., 2017). It was found that larger groups of spiders, with a lower density and more modularity, had increased transmission of the species, as did groups with a lower mean path length. However, as in this study, behavioural traits of individuals do not seem to affect transmission; they conclude that accounting for traits of a single exposed host has less power in predicting transmission dynamics than the resultant group level network structure. Despite this encouraging consistency with our results, it is still unclear as to how a relationship between network structure and the similarity of seminal microbiota of males within groups would be realised through sexual transmission, without the detection of individual mating metric effects. It must also be pointed out that changes and similarities in the composition of whole microbiotas are likely to be more complex than the transmission of a single organism.

Similarities between males in a mating group trial could also be as a result of co-habitation, meaning they share the same environment, eat the same food and come into close contact thus the microbiota could be transmitted through shared diet, ingestion of faecal matter or through non-mating social contact i.e. not due to mating induced transmission or disruption. Social group membership predicted microbiota similarity in red-bellied lemurs, *Eulemur*

rubriventer (Raulo *et al*., 2017) and savannah baboons, *Papio cynocephalus* (Tung *et al*., 2015). In this current work, though, the fact that mating group variation correlated with sexual network metrics does suggest that group differences are linked to mating activity of the group. However, this could be a via a non-mating driver of microbiota composition that correlates with the sexual network metrics of a group. For instance, general activity of the group, male condition or average hormone levels of males. Further, variation in male sexual ornaments and characteristics that influence female mating preferences has been associated with parasitic infection (Zuk *et al*., 1990) and ejaculate antibacterial activity (Rowe *et al*., 2011), so perhaps groups where males have differing levels of sexual ornamentation, accompanied by differing microbiota composition, would exhibit differing levels of overall mating activity.

The lack of an association between microbiota composition and mating behaviour on an individual level could also be due to individual differences in microbiota at the beginning of the experiment masking later change, perhaps due to an immuno-genetic effect. If males initially very dissimilar to each other in seminal microbiota composition have the same mating activity, they still may not be more similar to each other than to other males with differing mating activity but who were initially very similar in their seminal microbial composition. Similar work with samples collected both before and after inclusion in the mating trial, would add granularity to the effect of all these mating variables on males. Further, an increase in sample size would add power to our ability to detect effects, and information as to microbiota of females would allow investigation of the effect of mate identity as well as to the potential for transmission.

In summary we have one of the first explorations of the potential drivers of variation in male ejaculate seminal microbiota. We have found evidence of an effect of an immune-genetic driver on seminal microbiota composition, and an effect of inclusion in a mating trial group. However, with so many potential co-variates, the exact nature of these complex effects are difficult to tease apart. This study sets the stage for further work investigating these as yet unanswered questions.

APPENDIX§1

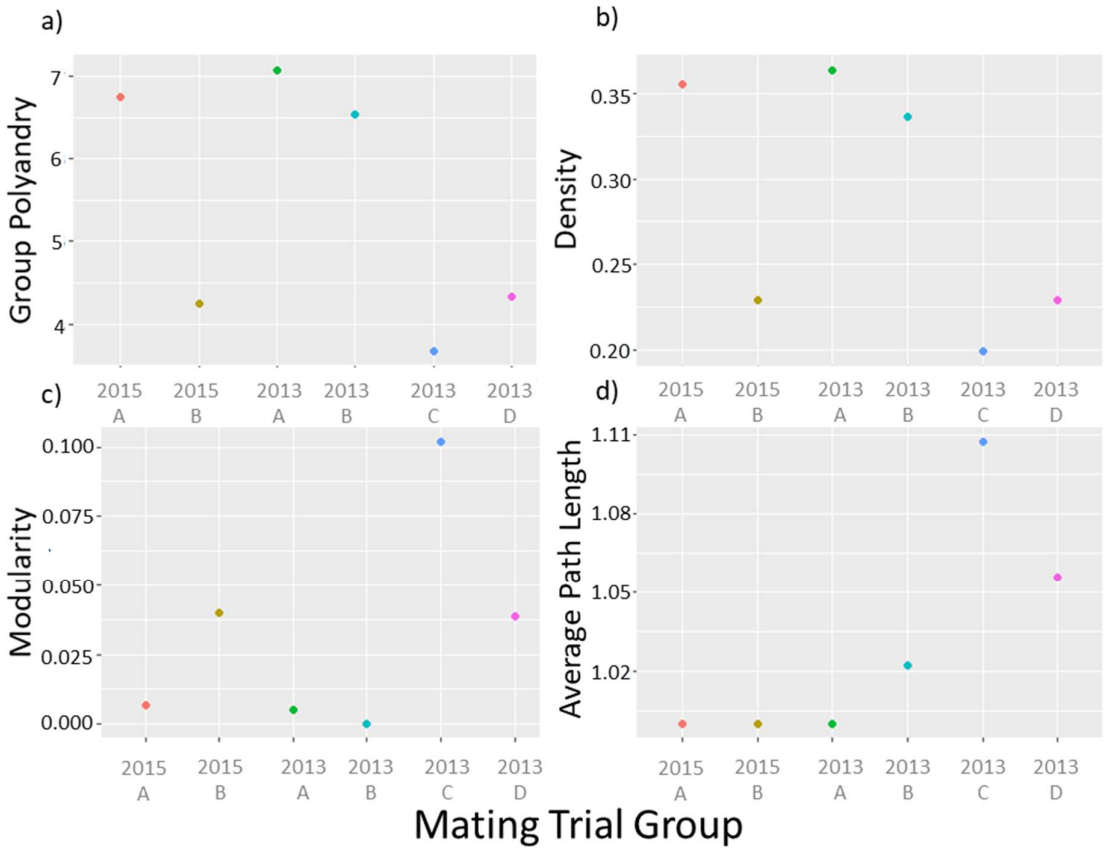


Figure A1 *Values of each of the four network metrics for each mating group, illustrating a correlation between average group polyandry, and sexual network density, and a slight correlation between modularity and mean path length*

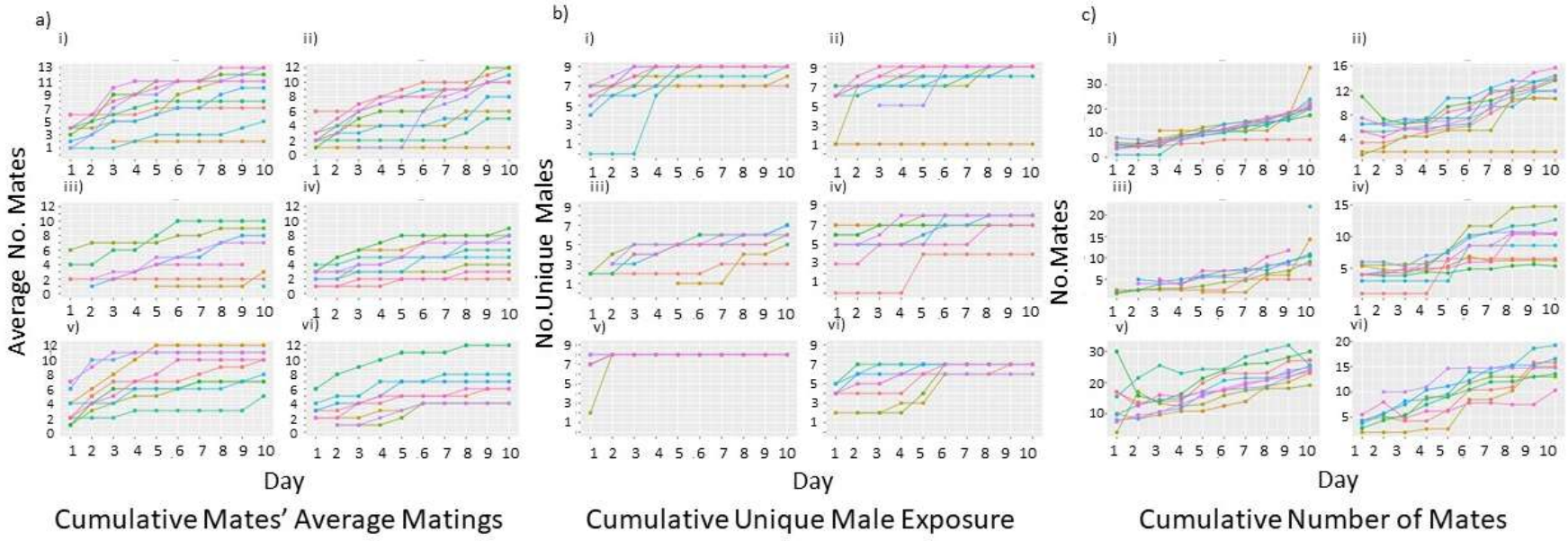


Figure A2 Cumulative mating metrics of each male for each day of a mating trial, collated by mating group: i) 2013_A ii) 2013_B iii) 2013_C iv) 2013_D v) 2015_A vi) 2015_D. Colours indicate male within each group. Where there are not points for every day for a male it is because calculations only began following a male's first mating.

CHAPTER 5. INSEMINATION AS A DRIVER OF VARIATION IN CLOACAL MICROBIOTAS WITHIN AND ACROSS FEMALE RED JUNGLEFOWL

Rosie Eccleston[a]; Kiyono Sekii[a,b]; Cedric Tan [a,c]; Grant C. McDonald[a]; Stephen Preston[d];

Tommaso Pizzari[a]

[a] Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK

[b] Department of Biology, University of Hirosaki, Hirosaki, Japan

[c] Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Oxford, UK

[d]Department of Zoology, University of Oxford, Oxford, UK

ABSTRACT

The community of bacteria living in the lower part of the female reproductive tract can play fundamental roles in female health, reproductive function and reproductive success. Evidence suggests that this community can be disrupted as a result of mating and the insemination of an ejaculate, which may affect costs of mating, with potentially important consequences for selection on female mating rates, female mating decisions, female mate choice, and thus patterns of sexual selection on males. Empirical evidence for insemination-mediated changes in female reproductive microbiotas however is currently limited and largely indirect because studies are often unable to: (i) control for confounding effects of co-habitation of breeding partners, (ii) distinguish the effects of mating per se from those of the insemination of an ejaculate, or (iii) isolate the signature of insemination from standing within- and between-female variation in microbiota composition. Here, we seek to address these limitations experimentally, by analysing temporal changes in the microbiota of the cloaca of female red junglefowl, *Gallus gallus*, over a series of artificial inseminations and following cessation of inseminations. We investigate whether the composition and structure of female reproductive microbiota changes in response to insemination, whether it becomes more similar to the seminal microbiota of the inseminating male, and whether there is transmission of bacterial species from males to females. Samples did not become less similar to virgin samples with repeated inseminations, or more similar to the inseminating male sample, with repeated inseminations. However, insemination did seem to have an effect on community composition and this effect is due to both the physical act of insemination and introduction of an ejaculate, and the effect is greatest when inseminating virgin females. Additionally, community composition changed

following the cessation of artificial insemination, with female samples becoming more similar to their virgin sample, and decreasing in the proportion of their OTUs shared with the last inseminating male ejaculate sample. These results are consistent with an effect of mating on the female reproductive microbiota and loss of male OTUs following mating cessation found by previous work in the field, and expand on these results demonstrating a greater effect of virgin inseminations, confirming an effect of both physical insemination and introduction of an ejaculate, and demonstrating a return towards a virgin state following mating cessation.

INTRODUCTION

The distal part of the female reproductive tract and the female genitalia host a community of bacterial species which has been shown to have a range of functional consequences for the host. For instance, evidence suggests that such microbiota can modulate protection of the female reproductive tract against disease. In women, a vaginal microbiota dominated by either *Lactobacillus crispatus* or *Lactobacillus iners* was associated with lower viral Sexually Transmitted Infection load and reduced HIV-1 viral RNA, compared with more diverse microbiota dominated by several anaerobes (Borgdorff *et al*., 2014). This could be as a result of a protective effect of a less diverse, *Lactobacillus*-dominated microbiota, potentially as a more stable community is less prone to disruption – Gajer *et al*., (2012) demonstrated stability of the vaginal microbiota of women was predicted by community composition. The reduced infection load could also be through more direct action of microbiota members; for example, *Lactobacillus crispatus* could prevent HIV colonisation through anti-inflammatory activity (Rose *et al*., 2012), whilst *Gardnerella vaginalis*, a bacterium often present in the more diverse, non-*Lactobacillus*-dominated communities, can activate HIV expression (Hashemi *et al*., 1999). Further, in dairy cows, *Bos taurus*, disruption of the normal microflora at parturition was followed by an increase in pathogenic bacteria and the subsequent development of metritis (Wang *et al*., 2013). Effects on fertility and gestation have also been demonstrated in humans, as an increase in total bacterial abundance in vaginal samples, a decrease in relative abundance of *Lactobacillus* species and symptoms associated with bacterial dysbiosis have all been linked to decreased likelihood of conception and increased likelihood of miscarriage (Eckert *et al*., 2003; Nelson *et al*., 2015). The

effect of a mother's vaginal microbiota can also persist post-partum through vertical transmission to offspring; vaginally delivered babies, as opposed to those delivered by caesarean, share a significantly higher proportion of gut microbiota species with their own mother compared with other mothers for the first two years of their life, while caesarean-delivered babies have a lower gut microbial diversity, are colonised later by usual members of the human gut microbiota and have differing immune response pathways (Jakobsson *et al*., 2014).

Mating and insemination can disrupt the female reproductive microbiota and possibly lead to the colonisation of the female tract by bacterial species from the male seminal microbiota through horizontal transmission. In humans, a change in composition of the vaginal microbiota was observed by Borovkova *et al*. (2011), with some females losing species, and some gaining species following intercourse, and mating was also a determinant in the temporal destabilisation of vaginal microbiotas (Gajer *et al*., 2012). Some of these community-level effects of mating are driven by individual species; *Lactobacillus crispatus* decreased in relative abundance post-mating (Mändar *et al*., 2015), whilst *Gardnerella vaginalis* increased in abundance in individuals following sexual debut (Mitchell *et al*., 2012), and changes in the frequency of *Lactobacillus iners* have also been associated with mating activity (Vodstrcil *et al*., 2017). There is also some evidence, albeit indirect, suggesting that some of these changes associated with mating may be caused by the transmission of bacteria or other bacteria-influencing factors, from males to females, rather than by the physical disturbance associated with mating and intromission. For example, a greater change in vaginal microbiota was detected in females whose partners had inflammatory prostatitis (Borovkova *et al*., 2011), which can be associated with changes in the relative abundance of different bacterial species in the male ejaculate (Ivanov *et al*., 2009). Additionally, Mändar et al. (2015) found higher similarity in community composition between couple members according to Principal Co-ordinate Analysis, and male community composition was a predictor of the magnitude of change in a partner's microbiota following intercourse. An effect on vaginal microbiota of mating is also observed in non-human species; mating activity was associated with vaginal microbiota composition in yellow baboons, *Papio anubis*, with higher

compositional similarity of this microbiota between individuals with more shared partners (Miller *et al*., 2017), again providing indirect evidence of an effect upon female microbiota other than simply mechanical disruption. Similarly, MacManes (2011) suggested that a difference in vaginal microbiota diversity between two sympatric species of closely related deer mice, *Peromyscus californicus* and *P. maniculatus*, may reflect species-specific differences in the levels of mating rates (i.e. sexual promiscuity). Such inter-specific differences in the female reproductive microbiota could either represent a result of different mating rates or as a protective adaptation to a harmful effect of mating. White *et al*., (2011) also posit that a positive correlation between cloacal bacterial diversity of the female common lizard, *Zootoca vivipara*, with multiple paternity of a brood as evidence that mating affects the female reproductive microbiota in this species.

As well as affecting health and reproductive success of a female host, the microbiota of the female genital tract is also likely to play an important role in determining the outcome of sperm competition, in species where females mate with multiple males (Parker, 1970). For instance, the female reproductive microbiota could be harmful to the spermatozoa transferred, either directly through bacterial products or through eliciting a host immune response, and if the ejaculates of certain male were more or less susceptible to these effects the microbiota could bias the probability of fertilisation in favour of certain male partners of the female (Firman *et al*., 2017). Studying the female reproductive microbiota is therefore not only crucial to the better prevention and treatment of sexually transmitted diseases and reproductive disorders, but also to the understanding of more fundamental aspects of sexual conflict, sexual selection, and mating system evolution. For example, the transmission of microbes (pathogenic or beneficial) or disruption of a healthy and functioning microbiota associated with mating can influence female costs of mating, inform criteria for female mate choice (Graves & Duvall, 1995) and affect the evolution of mating systems (e.g. Thrall, 1997; Thrall *et al*., 2000; Kokko *et al*., 2002).

Demonstrating the effect of mating on female reproductive microbiotas hinges on parsing out these effects from other sources of inter- and intra-individual variation. Empirical evidence

indicates that female reproductive microbiotas can vary substantially amongst females and dynamically within females. In olive baboons, *Papio anubis*, the diversity of the female reproductive microbiota declines with female age (Uchihashi *et al*., 2015), whilst in wild yellow baboons, *Papio cynocephalus*, variation in the relative abundance of certain bacterial taxa in the vaginal microbiota changed within females with a female's ovarian cycling and reproductive state (Miller *et al*., 2017). Similar patterns are found in humans, where vaginal microbiota changes with menstrual cycle (Gajer *et al*., 2012) and varies across ethnicities (Ravel *et al*., 2011; Vodstrcil *et al*., 2017). In addition to these general potentially confounding effects, studying the effect of mating on female reproductive microbiotas also requires controlling for factors specifically associated with mating behaviour, such as differential rates of re-mating across pairs or individuals. Crucially, mating individuals may spend more time in close proximity of each other than individuals that do not mate with each other. An increased similarity between the reproductive microbiotas of mating partners over time might therefore reflect co-infection due to shared environment, rather than necessarily the effect of the introduction of an ejaculate into the female.

The goal of this chapter is to investigate the effect of insemination on the microbiota of the female cloaca in a population of birds, using an experimental approach to control for potential confounds. In birds, the cloaca is the common opening for the digestive, reproductive and urinary tracts in both males and females, and we consider the cloacal microbiome as the avian female reproductive microbiota. There is therefore increased opportunity for digestive and urinary bacteria to be incorporated into, or influenced by, the reproductive microbiota, making the avian female microbiota a particularly interesting system in which to study the interplay between mating, microbiota and host function. The lack of intromission in most bird species limits the potential effect of physical contact between the male and female reproductive systems, indicating that any mating-related effects may be more confidently attributed to the insemination of a male ejaculate. Kulkarni & Heeb (2007) demonstrated sexual transmission in zebra finches, *Taeniopygia guttata*, from males to females and there is evidence of mating-induced changes in

the avian cloacal microbiota of females; in black-legged kittiwakes, *Rissa tridactyla*, breeding pair members shared more species with each other than with other individuals, and diversity of the female cloacal microbiota decreased and lost species shared with the male following prevention of insemination (White *et al*., 2010). Cloacal samples collected during the breeding season from house sparrows, *Passer domesticus*, produced growth such that colony counts of seven of eight cultured microbe types were significantly positively correlated between breeding pairs (Stewart & Rambo, 2000), which could be as a result of sexual transmission of microbiota members between pairs. Similarly, relative abundances of taxa were more positively correlated between European barn swallow, *Hirundo rustica*, breeding pairs following the breeding season than those of random pairs of population members (Kreisinger *et al*., 2015), again potentially as a result of sexual transmission.

Collectively, this previous work suggests that insemination might have important effects on the female reproductive microbiota of several bird species. This suggestion however, awaits conclusive evidence, through the experimental elimination of potential confounds. For example, the lack of female and male samples prior to mating, the lack of entirely virgin samples for comparison, and the conflation of mating with shared environment mean that previous studies often cannot separate mating effects from intra- and inter-female variation and other transmission routes. In this study, we attempt to address these issues by conducting an artificial insemination experiment in a captive population of red junglefowl, *Gallus gallus*. This species is an established system for research for studies of sexual selection and sperm competition; work has demonstrated both pre- and post-copulatory sexual selection on males through male behaviour and sperm characteristics (e.g. McDonald *et al*., 2017; Tan *et al*., 2017), selection against inbreeding (Pizzari *et al*., 2004) and cryptic preferences for more MHC-dissimilar mates in both males (Gillingham *et al*., 2009) and females (Løvlie *et al*., 2013). Understanding of the effect of mating on the red junglefowl female microbiota will therefore add greatly to the context of this work, and similar work in other avian species, for instance in understanding how the reproductive microbiota may mediate sexual selection and competition through providing cues

of mate information, affecting costs of mating, or mediating post-copulatory processes. We use next-generation sequencing to investigate patterns of variation in the microbiota of the female cloaca in relation to exposure to repeated inseminations. Virgin females were subjected to a series of artificial inseminations (up to six over a period of two weeks), sampled as virgins and then prior to each new insemination (giving an 'after' sample for the previous insemination), with the microbiota of the semen used at several insemination events being characterised as well. This generated a repeatedly sampled group of females of known mating history inseminated with ejaculates of known composition. Through the use of artificial inseminations, we were able to isolate the effect of insemination from co-habitation and other means of transmission. We investigate the following predictions arising from the hypothesis that the female cloacal microbiota is affected by inseminations:

1. The composition and structure of a female cloacal microbiota will change in response to insemination, and any such effect will increase (with an already disrupted microbiota more susceptible to change) or diminish (with subsequent disruptions having less of an impact on an already disrupted microbiota) in magnitude with repeated inseminations over time.

2. The cloacal microbiota of a female will become more similar to the seminal microbiota of the inseminating male, and more similar to that of other females inseminated by the same male, due to the sexual transmission of species, or transmission of factors favouring certain species.

3. There will be a change in the composition and structure of the female cloacal microbiota following the cessation of inseminations.

4. This change will represent a return towards the virgin state, suggesting this virgin state is as a result of selection within the female.

MATERIALS & METHODS

Study Population

We studied a captive population of red junglefowl at the John Krebs Field Station of the University of Oxford, in Wytham, Oxfordshire, which has been used in many studies of sexual selection, mating systems and sperm competition. The study subjects were six males and seventeen females used in a series of systematic sampling and artificial insemination experiments in the 2013 breeding season. Seven females were only sampled once, as virgins before any artificial inseminations, and the other ten females were sampled as the series of artificial inseminations continued. Males and females belonged to two cohorts hatched in 2012 and were all one year old and thus sexually mature at the time of the experiment. All females were virgin at the start of the experiment.

Experimental Design

Over a period of twelve weeks females were exposed to three separate series of artificial insemination events. A series of artificial insemination events comprised repeated events with two to three days between events, and at each event the cloacal microbiota of a subset of females was sampled as described below, and then the female inseminated with the ejaculate of a given male.

Between each series, there were gaps of between one and two weeks where females were not inseminated, but the cloacal microbiota continued to be sampled.

In the first series of inseminations there were fourteen artificial inseminations across approximately six weeks, the second two inseminations across three days and the third four inseminations across a week (Figure 1).

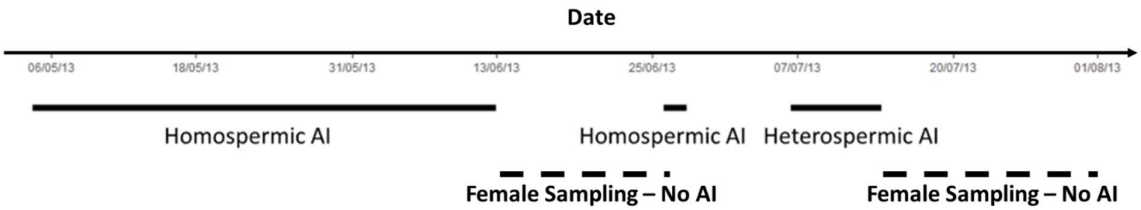


Figure 1 *Timeline of three insemination series, consisting of artificial insemination (AI) events two to five days apart, and two series of sampling events two to five days apart following the cessation of series of artificial inseminations*

To investigate the effect of artificial insemination, two sets of female samples were used: those samples collected during the artificial insemination events of an artificial insemination series (Table 1a), and those samples collected following the cessation of an artificial insemination series (Table 1b).

In this study seventeen females were exposed to repeated artificial insemination events, and ten of these were sampled at each artificial insemination event and in gaps between artificial insemination series, immediately prior to insemination, with the remaining seven only sampled as virgins prior to any artificial inseminations. These ten were randomly selected through Bird IDs. One of these ten females was a control, inseminated identically to other females, except with sterile phosphate buffered saline (PBS) rather than male semen. Figure 2. illustrates a single artificial insemination event. Each male was used as the semen donor to inseminate a pair of females, and the same male was used to inseminate the same two females each time. Semen samples were obtained from a male through abdominal massage, as outlined in chapter 2, and were inseminated approximately 2 cm into the vagina from the cloaca (prior to the sperm storage tubules) of females, using a 1000uL Gilson pipette. The volume inseminated was not standardised, as semen samples were calibrated to standardise for the number of sperm, but each was approximately 150uL which is within the range of natural ejaculates.

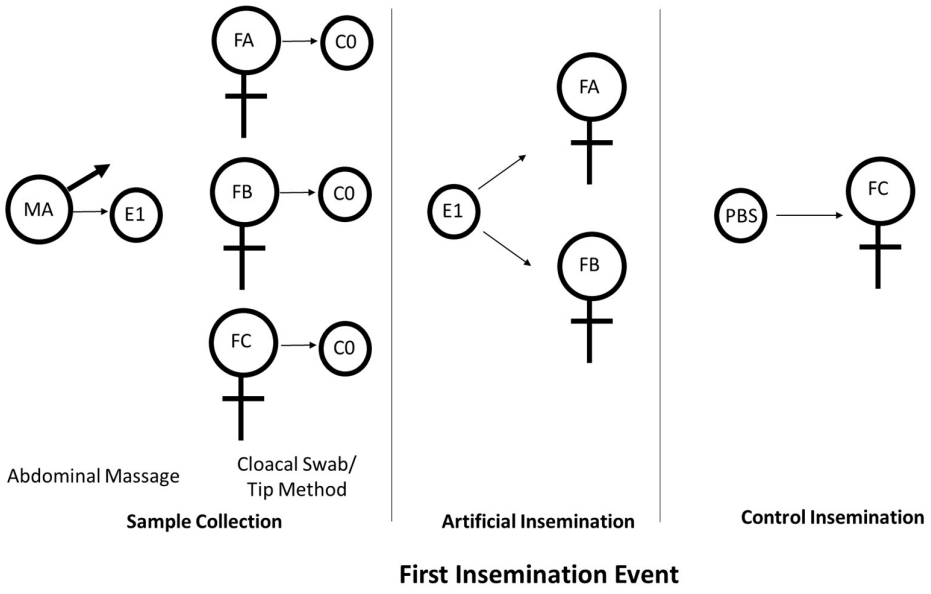


Figure 2 *Diagrammatic representation of a single insemination event where there is sample collection from males and females to collect a cloacal sample (C0) and semen sample (E1). This is followed by artificial insemination of females A (FA) and B (FB), with a single male semen sample from male A (MA) being used to inseminate both females. Control inseminations follow the same protocol as artificial inseminations, except that the control female (FC) is inseminated with sterile Phosphate Buffered Saline (PBS) and not a male ejaculate. All subsequent insemination events follow the same process, and after cessation of inseminations, there is no more insemination or collection of male samples, but sampling of females proceeds in the same way at sampling events. All events in a series occur between two and five days apart.*

Table 1 *The samples collected and used for each artificial insemination event in a series, and collected at sampling events following cessation of inseminations*

a)

Insemination Event	Male Sample Collected & Used for AI	Female Sample Collected
First AI	E1	C0
Second AI	E2	C1
Third AI	E3	C2
Fourth AI	E4	C3
Etc		

b)

Collection Event	Male Sample Collected & Used for AI	Female Sample Collected
Last AI of Series	Last E	C0 (not used)
First Post AI Collection	-	C1
Second Post AI Collection	-	C2
Third Post AI Collection	-	C3
Etc		

Females were sampled using two methods, 'tip' and 'plate' over the 12 weeks, with the method being switched from tip to plate after three weeks part way through the first series of inseminations, due to considerations for the females' welfare. The effect of switching sampling methods is considered in analysis stages. On certain insemination events, the semen used to inseminate females were also sampled for microbiota, providing 12 insemination events at which we have sample information for at least some of the males, seven of which events are used in this work. Due to the aims of a concurrent study using the same birds and artificial insemination events, the third series of inseminations was different to earlier inseminations as these were 'heterospermic', rather than 'homospermic' inseminations. This means that semen samples from two males were combined and used to inseminate a female.

Sample Collection and Processing

Female cloacal samples were taken through one of two methods. Plate samples were taken by applying gentle pressure to evert the cloaca and a sterile petri dish was pressed quickly and gently against the exposed cloaca, and the dish lid closed immediately. The part where the cloaca had touched was marked on the back of the petri dish and the dish placed immediately on dry ice for transportation. The time that the petri dish was open was minimised in order to avoid exposure to environmental contamination. Tip samples were taken through the insertion of a sterile pipette tip into the cloaca, and then washed in sterile PBS. Both types of sample were then placed on dry ice within five minutes of collection before storage at -80°C. Semen samples for microbiota analysis were aliquots from the same samples obtained through abdominal massage for artificial insemination (see above). Once processed, all samples were immediately placed on dry ice, before transfer to the Department of Zoology and storage at -80°C within the same day of collection for further processing via the Illumina MiSeq Platform using the pipeline developed and described in chapter 2, modified for the collection of DNA from tips or plates in the case of female samples. All handling of the birds, samples and equipment at the field station was done with the same careful collection and laboratory processing protocol implemented to minimize environmental contamination outlined in chapter 2.

DNA Extraction

DNA was extracted and purified from both male and female samples using the DNeasy Blood and Tissue Kit (QIAGEN) and according to the manufacturers guidelines, with temperatures, timings and volumes optimised for bacterial DNA extraction for cloacal and ejaculate samples. For tip cloacal samples, 200uL lysis buffer mix and 190uL water were added to 10uL of the PBS wash then incubated at 37°C for 10 minutes. For plate cloacal samples, 200ul lysis buffer and 190uL water were added to the petri dish and the sample scratched in a hash pattern with the pipette tip, then placed in an oven at 37°C for incubation. The liquid was then all collected in a new sterile 1.5ml eppendorf and extraction continued in the same way for both sample types. 50ul Proteinase K is added followed by 400uL Buffer AL then incubated at 56°C for 70 minutes. 8uL RNase was then added and the mix incubated at 37°C for 10 minutes. DNA was purified by precipitation with 750uL of 100% ethanol and passed through a DNEasy spin column, then washed with 500uL Buffer AW1 then again with 500uL of Buffer AW2 and eluted with 60uL molecular biology reagent water. Purified DNA was frozen and stored at -20°C. The process of extraction from male semen samples was the same as for tip cloacal samples, with the water and lysis buffer mix added to a 10uL aliquot of semen.

PCR Amplification, Library Preparation and Sequencing

Following DNA extraction, we proceeded with PCR amplification of the DNA product in the same way as amplification of ejaculate and culture samples in previous chapters. Briefly, amplification of the V3 and V4 regions of the bacterial 16S encoding genes was carried out using Phusion High-Fidelity DNA Polymerase according to the Phusion polymerase manufacturer's protocol and with reference to the melting temperature of primers 347F and 803R (New England Biolabs). A barcode sequence attached to the forward primer, 347F, allows identification of the sample of origin of the amplicons following sequencing, as a different barcode is added to the reaction mix for each individual sample prior to amplification on a 96well Veriti Thermal Cycler (Applied Biosystems). As usual, we were careful to minimise contamination of samples from the laboratory environment and other DNA products, following the previously described protocols.

We confirmed amplification of the desired target by gel electrophoresis and visualization under UV light, and the PCR products were then pooled into DNA libraries, purified using QIAGEN MiniElute PCR Purification Kit and prepared for sequencing on the Illlumina MiSeq platform, with attachment of the Illumina barcoded index primer Illumina adaptor using the NEBNext Ultra DNA Library Prep Kit before submission to the sequencing facility.

Post-Illumina Processing

Sequences resulting from Illumina MiSeq were analysed, again, according to the chapter two protocol, using custom made scripts in R and QIIME and with the same approaches to remove and address whole dataset and per-MiSeq contaminants, to achieve an equal number of sequences per sample with sufficient coverage and to calculate relative abundance. Subsequent analysis was performed using R software version 3.3.2 (R Core Team, 2016).

Statistical Analysis

We investigated the effect of mating upon the female microbiota, and the potential for transmission of members of the male ejaculate microbiota to the female using a series of artificial inseminations and sampling events. Initially we performed an exploratory analysis to ask which aspects of insemination, and potentially confounding methodological factors, may lead to changes in community richness, diversity and composition. This was followed by more focused tests of key predictions generated by the hypothesis that the female cloacal microbiota is affected by inseminations.

Exploratory Analysis: Identifying Sources of Variation in Community Composition and Structure

We first explored insemination as a source of variation using 187 female samples collected from 16 females across twelve weeks, including two samples per female on the day where both tip and plate samples were collected, to allow a direct comparison of these methods. Additionally, a further 19 samples were collected from the 'Control' female. These samples are not included in formal statistical analysis as we have only one control, but are plotted alongside

other females in graphs, illustrating change in community structure over time for visual comparison.

We calculated three community diversity metrics (species richness, Shannon diversity and Simpson diversity) for each sample using the R package 'vegan', and we performed a Principal Component Analysis (PCA) across all samples using the standard 'stats' R package function 'prcomp'. Species richness is simply the number of Operational Taxonomic Units (OTUs; clusters of reads with 97% sequence similarity considered equivalent to bacterial species) found in a sample. Shannon and Simpson are two similar measures of diversity combining richness and evenness, where high numbers indicate a more diverse community. The PCA was performed using a subset of 54 OTUs, that occur in at least 10% of samples and at an average relative abundance equal to or greater than 0.1%. This excludes OTUs that occur in samples at very low abundances, or in just a few samples, which could affect sample PCA scores if included in the analysis as their rarity means there is insufficient information to estimate their patterns of abundance. The relative abundances of these OTUs, as calculated prior to exclusion of those OTUs which did not meet the inclusion criteria, were used as input into the analysis. The first PCA axis explained 10.6% and the second 7.4% of variation in relative abundance of OTUs between samples (Fig 3.). These two axes were used in further analysis to give five community metrics (species richness, Shannon diversity, Simpson diversity, PC1 and PC2) used as response variables in statistical models exploring the sources of variation between samples.

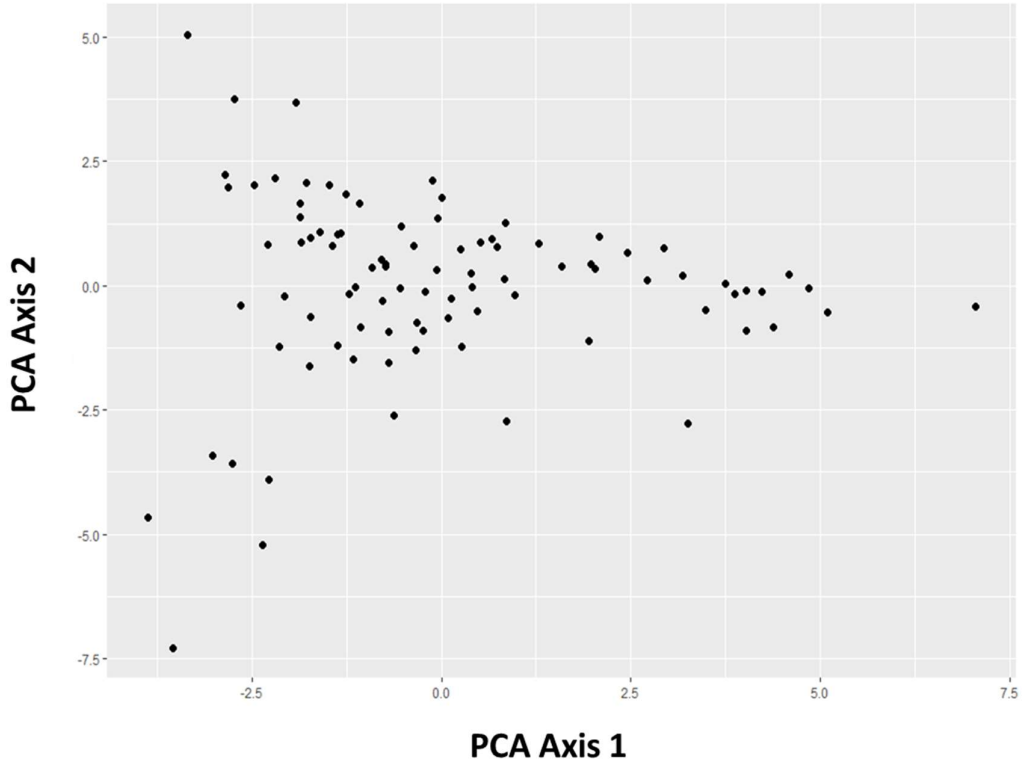


Figure 3 *Variation of samples along the two first two axes of the Principal Component Analysis*

We investigated different aspects of the artificial insemination experiment which may contribute to variation between samples across the twelve-week experimental period. These include factors relating to an effect of artificial insemination – virgin status, numbers of artificial insemination in a series - as well as more methodological sources of variation – collection method, artificial insemination type and collection date.

Virgin status simply indicates whether a female has been previously inseminated or not. Artificial insemination number is the number of artificial inseminations a female has undergone in that series of inseminations and is zero if a female has not been inseminated in the last week, as we are more interested in if there are cumulative effects of inseminations in relatively quick succession rather than total inseminations in a whole season. A previous study of avian cloacal microbiota suggests effect of insemination to be much reduced within a week or two of insemination cessation (White *et al.*, 2010). Insemination type indicates whether the insemination is heterospermic or homospermic, and is an artefact of the dual purpose of this

series of inseminations and not a focus for our analysis, but is included as a variable here. Insemination with one or two ejaculates could affect the magnitude of any effect of any insemination, and an interaction with insemination number is included in case insemination type also affects the nature of any effect of insemination number.

A full linear mixed model was therefore constructed as follows, with Female ID included as a random factor, and the five community metrics as response variables:

Response variable ~ Virgin Status + Insemination Number*InseminationType + Collection Date + Collection Method + (1| Female ID)

Simpson diversity is bounded by zero and one and was therefore square-root, arc-sine transformed. The box cox function in R package 'MASS' was then used to determine the transformation lambda value required to ensure data for the five community composition measures met the assumptions of the linear mixed model, according to diagnostic plots.

Variables were standardised using the R command 'standardise', and the model set to compare all possible models with all combinations (including absences) of the above variables from the null model to the full model is generated and models compared using the R command 'dredge'.

Given the large number of terms in the full model, we used a Bayesian Information Criterion to select the best model. If the null model was not within two BIC of the best model, all terms in models within two BIC values of the best model were retained. The resultant 'best' model for each response variable was then run using the 'lme' function in the 'lme4' package to provide coefficients for each term. If the null model was within two BIC of the best model, no terms were considered significant.

Investigating the Effect of Artificial Inseminations upon Female Cloacal Microbiota

To more specifically explore the effect of artificial inseminations we selected a section of the first insemination series, beginning with the virgin female cloacal samples and ending with

samples collected after 6 artificial inseminations. This comprised 53 female cloacal samples from 9 females, plus 6 from one control female, and 14 male semen samples from 5 males overall, where we can look at the change in a virgin female in response to initial insemination and five subsequent inseminations. We do not look any further ahead as the samples transition to a different collection method, and thus would introduce a confounding factor. Unfortunately, no sample was collected after the first insemination, so we can only compare virgin samples to samples collected after the female has been subjected to two or more inseminations.

We also look at the effect of cessation of artificial inseminations by selecting two series of four samples collected after the last of a series of homospermic and a series of heterospermic inseminations.

The homospermic dataset comprises 6 male samples from 6 males and 31 female samples from 9 females, plus 3 samples from one control female.

The heterospermic dataset comprises 25 female samples from 9 females plus 3 from one control female. No male samples were used here due to incomplete information regarding the identity of the second inseminating males.

In all these sub-datasets, samples have been lost at the various stages of collection and processing, and thus there is not a sample for every bird at every time point. We therefore don't have complete coverage - particularly as most measures investigated here require two samples - but coverage is still sufficient to observe and analyse patterns emerging.

The control female (inseminated with sterile PBS, not an ejaculate), was not included in formal statistical analysis, but is included in graphical representations of the data to at least visualise the dynamics of this microbiota.

We therefore used these sub-datasets to further explore the effect of inseminations on the female microbiota, investigating:

a) Whether there is a change in community composition and structure of the female reproductive microbiota in response to repeated inseminations

b) Whether any change is as a result of becoming more similar to the inseminating male i.e. due to the transmission of OTUs or factors favouring certain OTUs

c) Whether there is a change in community composition and structure of the female reproductive microbiota following the cessation of inseminations

d) Whether any change represent a return towards virgin community composition

e) Whether any change is a result of becoming less similar to the male i.e. due to the loss of the source/exclusion of male OTUs or factors favouring certain OTUs

i) Effect of Repeated Artificial Inseminations on Community Composition and Structure of Female Reproductive Microbiota

Here we investigated whether insemination results in a change in community composition of the post-insemination female sample compared with that of the virgin sample, in terms of the magnitude of change (either an increase or decrease) in species richness, Shannon diversity or Simpson diversity; the proportion of OTUs shared with the virgin sample and the Bray-Curtis Dissimilarity to the virgin sample. Absolute magnitudes of changes were calculated from richness and diversity measures calculated in previous analysis, with the virgin female sample's value being substracted from the focal sample's value, and converted to an absolute value using 'abs' function in the R basic stats package. Bray-Curtis Dissimilarity is calculated in R with the 'vegdist' function of the 'vegan' package, and proportion of shared OTUs is the number of OTUs present in the focal female sample and the virgin sample as a proportion of the total number of OTUs in the focal female sample. 'Number of Inseminations' is the number of inseminations that the focal female has been subjected to before the focal sample was collected.

The following full model is therefore created, with Female ID and ID of the inseminating male included as random factors:

$$\text{Response Variable} \sim \text{No. Inseminations} + \text{No. Inseminations}^2 + (1|\text{FemaleID}) + (1|\text{MaleID})$$

To establish the effect of each of the two fixed factors, the full model was simplified, running less parametrised models (Table 2) using the lme function in the 'lme4' package and comparing models using the ANOVA function in the standard 'stats' R package. A p-value generated by the function using log-likelihood ratios from two models with and without the focal model component was utilised to estimate the significance of the extra variation explained by each model component.

Table 2 *Structure of the four models compared to establish the contribution of number of inseminations to variation in response variables across female samples*

Species Richness		Shannon Diversity		Simpson Diversity		PCA Axis 1		PCA Axis 2	
Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC
Virgin Status	0	Null	0	AI Type	0	Collection Method	0	Collection Method	0
Null	4.05	AI Type	0.90	Null	0.46	AI Type + Collection Method	4.34	AI Type + Collection Method	3.46
AI Type + Virgin Status	4.41	Collection Date	1.76	Collection Date	0.89	Collection Method + Collection Date	4.89	Collection Method + AI Number	4.24
Collection Method + Virgin Status	4.80	Collection Method	1.89	Collection Method	1.70	Collection Method + AI Number	5.01	Collection Method + Virgin Status	4.94
Virgin Status + Collection Date	4.87	Virgin Status + Collection Date	2.73	Virgin Status + Collection Date	3.17	Collection Method + Virgin Status	5.13	Collection Method + Collection Date	5.15

The nature of the relationship between any significant factors and the response variables were then determined through running the linear mixed model including all significant terms to obtain an estimate coefficient.

We then investigated whether samples from the same female collected before and after an artificial insemination event become more or less similar to each other as inseminations progress. We consider similarity both between samples before and after a single artificial insemination and also before and after two artificial inseminations, in order to be able to comparably include the similarity of virgin samples to the next female sample collected; as there was no sample taken at the second insemination event i.e. after one insemination, we could not compare the virgin sample, or the sample taken after two inseminations, to a sample taken after just one insemination. So first we investigated samples after three inseminations compared with samples collected after two, samples after four compared to those after three etc. In the second analysis we compared samples collected after two inseminations with the virgin sample, four samples with two, five with three etc.

The focal female sample is the latter of the two samples being compared. Number of inseminations in both cases is the number of inseminations that the focal female has been subjected to before the focal sample was collected, so when number of inseminations is four in the 'one AI apart' analysis, the female sample collected after four inseminations is being compared to the female sample collected after three inseminations, but in the 'two AI apart' analysis the female sample collected after four inseminations is being compared to the female sample collected after two inseminations.

Bray-Curtis Dissimilarity is calculated as before and proportion of shared OTUs is the number of OTUs present in the focal female sample and the female sample one or two inseminations previous as a proportion of the total number of OTUs in the focal female sample.

The full model is therefore created as before, with Female ID and ID of the inseminating male included as random factors, and the significance of the number of inseminations as a fixed

factor established, and the nature of relationship with the response variable calculated, as previously described.

ii) Change in Community Composition and Structure of Female Reproductive Microbiota in Relation to the Inseminating Male

Here we investigate whether a female sample becomes more similar to the male semen used as inseminations progress in terms of the proportion of OTUs shared with the male sample and the Bray-Curtis Dissimilarity to the male sample. We consider similarity both to the male sample immediately previous and also to the male sample from two previous inseminations, again in order to be able to comparably include the similarity of the virgin female.

Number of inseminations in both cases is the number of inseminations that the focal female has been subjected to before the focal sample was collected, so when number of inseminations is four in the previous male sample analysis, the female sample collected after four inseminations is being compared to the male sample used for the fourth insemination, but in the two male samples previous analysis the female sample collected after four inseminations is being compared to the male sample used for the third insemination.

Bray-Curtis Dissimilarity is calculated as before and proportion of male OTUs is the number of OTUs present in the focal female sample and the male sample as a proportion of the total number of OTUs in the focal female sample.

The full model was created as before, with Female ID and ID of the inseminating male included as random factors, and the significance of the different factors established, and the nature of relationship with the response variable calculated, as previously described.

We also investigated whether samples from females inseminated with the same ejaculate become more similar to each other as inseminations progress in terms of the proportion of OTUs shared and Bray-Curtis Dissimilarity. Bray-Curtis Dissimilarity is calculated as before and proportion of shared OTUs is the number of OTUs present in both female samples and as a

proportion of the total number of OTUs found in either of the two samples. Number of Inseminations is the number of inseminations that the females have been subjected to before the focal samples were collected.

The full model was therefore created, with ID of the inseminating male included as a random factor, and the significance of number of inseminations established and the nature of relationship with the response variable calculated as previously described.

We also investigated whether OTUs found in the male sample used to inseminate a virgin female, but not in that female's virgin sample, are found in that female following insemination and whether they make up an increasing proportion of the female's OTUs as inseminations continue.

A list of these OTUs is generated for each male and overlap with each sample of the females that male inseminates at each point is calculated as a proportion of total OTUs in each female sample.

To distinguish between OTUs found in the female due to transmission of OTUs from the male to the female and simply as a result of turnover of bacteria within the female microbiota overtime, male only OTUs of randomly picked females, and control females were also calculated. Pseudo-random pairs were generated from the males and females such that each female was paired with a male whose ejaculate was not used to inseminate her. A new set of male-only OTUS was then generated for each random pair and the proportion in each female at each time point calculated. The same procedure was repeated for pairs created by pairing each of the males with the PBS control female.

Whether the relationship between male-only OTUs and number of inseminations differs between original, random and control pairs is then tested by including a 'Pair Type' term in the full linear mixed model, and an interaction between this and number of inseminations. Number of inseminations is the number of inseminations that the focal female has been subjected to before the focal sample was collected.

The full model is therefore created, with Female ID and ID of the inseminating male included as random factors, and the significance of the number of inseminations established and the nature of relationship with the response variable calculated as previously described.

iii) Effect of Cessation of Artificial Inseminations on Community Composition and Structure of Female Reproductive Microbiota

Here we investigated whether following the cessation of a series of inseminations there is a change in the female microbiota, becoming less similar to the first female sample taken after the last insemination. We considered two series of samples taken after two different series of inseminations, one homospermic and one heterospermic, and so AI Type was included as a fixed factor in the full model.

We considered change in community composition in terms of the magnitude of change (either an increase or decrease) in species richness, Shannon diversity or Simpson diversity; the proportion of OTUs shared with the first post-AI sample and the Bray-Curtis Dissimilarity to the first post-AI sample. Absolute magnitudes of changes, Bray-Curtis Dissimilarity and proportion of shared OTUs are calculated as described previously. Number of samples is the number of sampling events since the last artificial insemination of the females.

The following full model was therefore created, with Female ID and ID of the inseminating male included as random factors:

Response Variable ~ No. Inseminations*AI Type + No. Inseminations^2*AI Type+ (1|FemaleID) + (1|MaleID)

The effect of each of the fixed factors and the interaction between them was established using model simplification as described previously.

iv) Return to Virgin Community Composition and Structure Following Cessation of Artificial Inseminations

Here we investigated whether insemination results in a change in community composition of the post-insemination female as a return towards that of the virgin sample, in terms of the alpha diversity, proportion of focal sample OTUs shared with the virgin sample and the Bray-Curtis Dissimilarity to the virgin sample, both calculated as before. Number of Samples is the number of sampling events since the last artificial insemination of the females.

The full model was therefore created, including an interaction with AI Type, with Female ID and ID of the inseminating male included as random factors, and the significance of the effect of each of the fixed factors and the interaction between them established as described previously.

We then investigated whether a female sample becomes less similar to the male semen used in the last artificial insemination before cessation of the series, in terms of the proportion of OTUs shared with the male sample and the Bray-Curtis Dissimilarity to the male sample. Number of Samples is the number of sampling events since the last artificial insemination of the females. Bray-Curtis Dissimilarity and proportion of male OTUs are calculated as before.

The full model was therefore created, with Pair ID (combining Female ID and ID of the inseminating male) included as a random factor, the significance of fixed factors and interactions established and the nature of the relationship with the response variable calculated as previously described. Pair ID was used due to the lower number of samples meaning insufficient degrees of freedom for two random factors. Graphs suggested there is variation between females, but not necessarily an effect of inseminating male.

We were unable to compare similarity of females inseminated by the same ejaculate because there are not enough pairs of samples from females inseminated by the same ejaculate.

RESULTS

Exploratory Analysis: Identifying Sources of Variation in Community Composition and Structure

All community metrics (i.e. species richness, Shannon diversity, Simpson diversity, PC1 and PC2) exhibited substantial variation among females, as well as temporal fluctuations within females across the experimental period and the course of inseminations (Fig. 4).

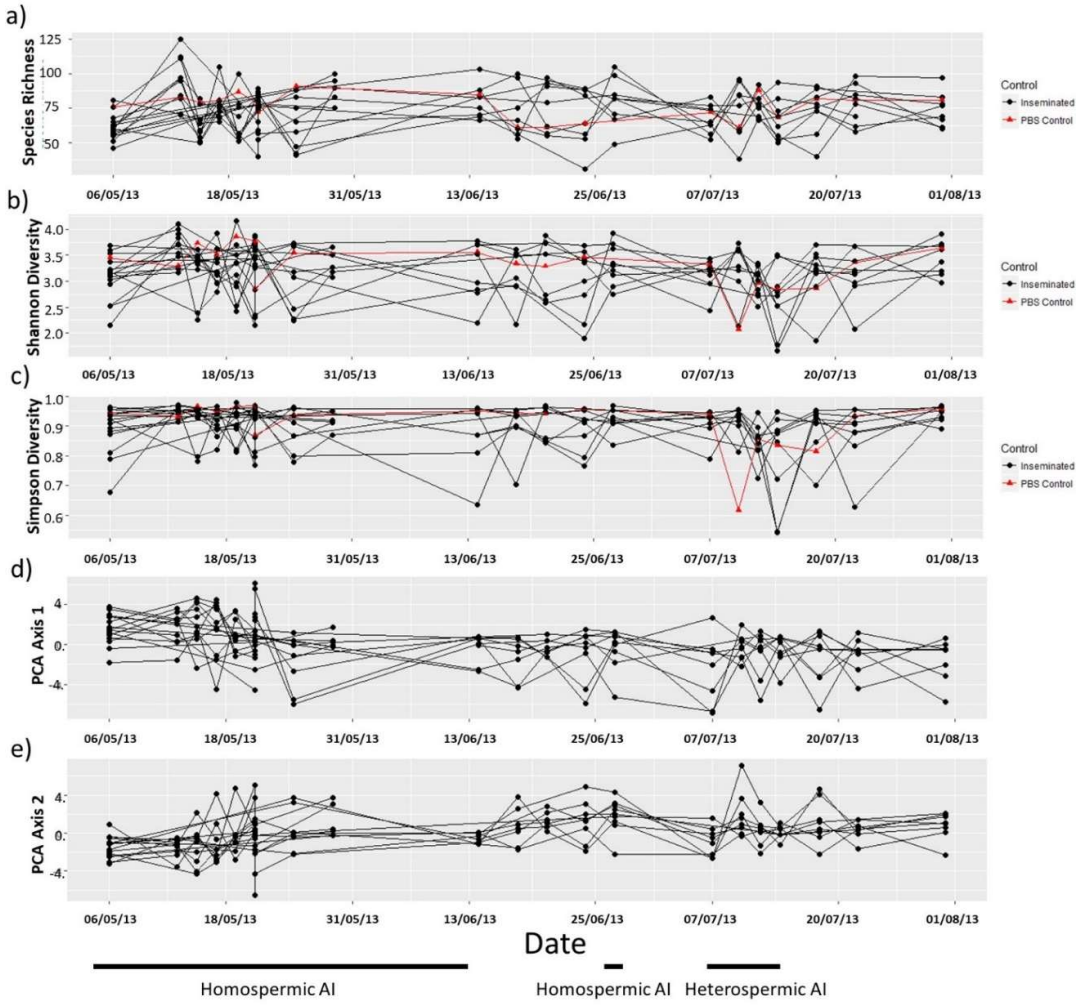


Figure 4 *Time series demonstrating temporal variation in five microbiota community measures across the twelve weeks of the experimental period: a) Species richness b) Shannon diversity c) Simpson diversity d) PCA Axis 1 e) PCA Axis 2. Thick black lines at the bottom of the fifth graph correspond to all graphs and show a series of 14 homospermic inseminations 2-4 days apart, a series of two homospermic inseminations over three days and a series of four heterospermic inseminations 2-3 days apart. The graphs are all aligned with each other by the x-axis so that the thick black lines are applicable to all and indicate the same dates. Graphs a) to c) include a control female inseminated with sterile PBS, differentiated from samples from experimental females inseminated with ejaculates by shape and colour. As there was only one control, this sample was not included in statistical analysis and therefore was not included in the Principal Component Analysis and is not shown in the bottom two graphs.*

Variation in species richness was predicted by the 'Virgin' status of a female: samples from virgin females had lower species richness than non-virgin samples (Estimate = -5.69; Standard Error = 1.85; t-value -3.07; Table 3; Fig. 5c). However, for both Shannon and Simpson diversity no model was a significantly better predictor of variation than the null model (Table 2).

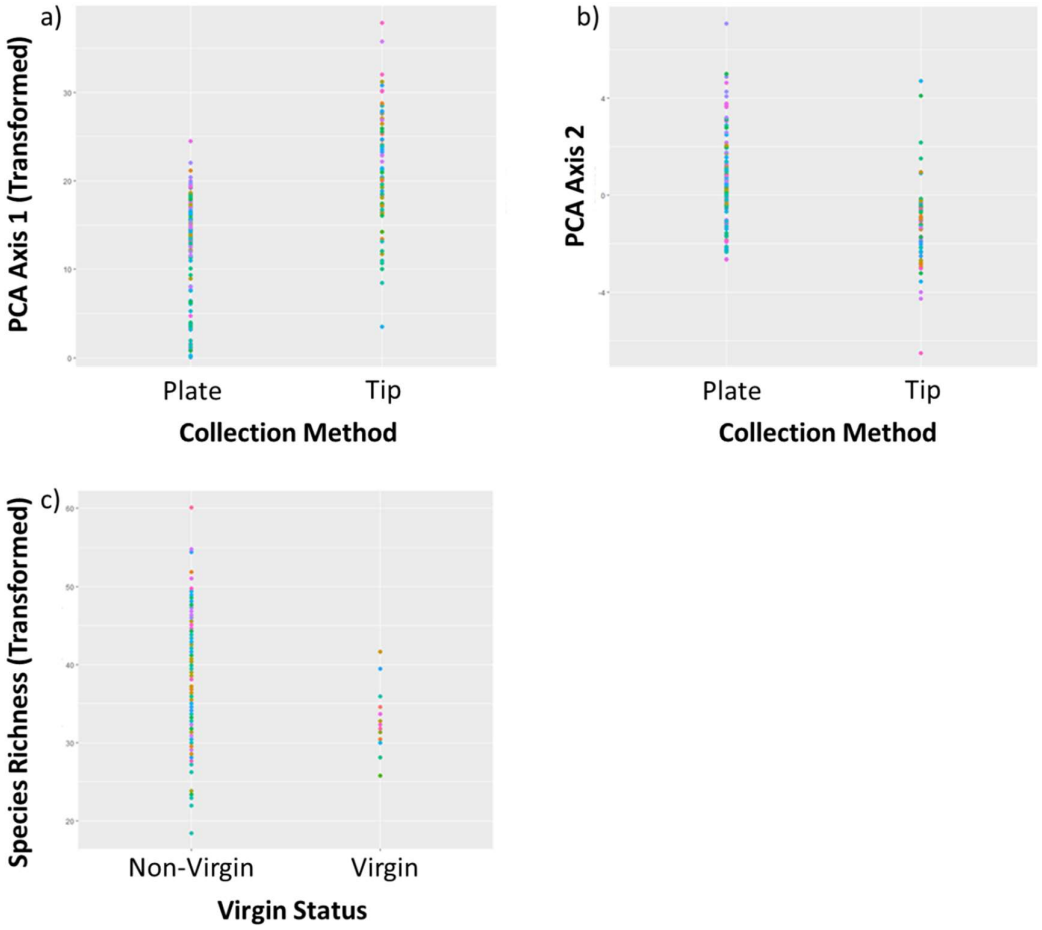


Figure 5 *Sources of variation in female reproductive microbiota communities. Colour indicates female.*

The best predictor of variation across both PCA axes was Collection Method with Tip samples having higher values of PC1 than Plate samples (Estimate =7.81; Standard Error = 0.92; t-value = 8.47; Table 3; Fig. 5a) and Plate samples a higher value than Tip samples for PC2 (Estimate =-1.78; Standard Error =0.28; t-value = -6.38; Table 3; Fig. 5b).

Table 3 *The best five models for each community metric according to BIC-based model selection. The models that are significantly better than the null (the null is not within two BIC of the top model) are indicated using bold italics. The left-hand column for each community metric indicates fixed factors included in the model, whilst Collection Date and Bird ID are included in all as random factors.*

Species Richness		Shannon Diversity		Simpson Diversity		PCA Axis 1		PCA Axis 2	
Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC	Top Five Models	Δ BIC
Virgin Status	0	Null	0	AI Type	0	Collection Method	0	Collection Method	0
Null	4.05	AI Type	0.90	Null	0.46	AI Type + Collection Method	4.34	AI Type + Collection Method	3.46
AI Type + Virgin Status	4.41	Collection Date	1.76	Collection Date	0.89	Collection Method + Collection Date	4.89	Collection Method + AI Number	4.24
Collection Method + Virgin Status	4.80	Collection Method	1.89	Collection Method	1.70	Collection Method + AI Number	5.01	Collection Method + Virgin Status	4.94
Virgin Status + Collection Date	4.87	Virgin Status + Collection Date	2.73	Virgin Status + Collection Date	3.17	Collection Method + Virgin Status	5.13	Collection Method + Collection Date	5.15

Investigating the Effect of Artificial Inseminations

To more specifically explore the effect of artificial inseminations we selected a section of the first insemination series where we can look at the change in a virgin female in response to initial insemination and five subsequent inseminations. We do not look any further ahead as the samples transition to a different collection method, and thus would introduce a confounding factor. We also look at the effect of cessation of artificial inseminations by selecting two series of four samples collected after the last of a series of homospermic and a heterospermic inseminations.

i) Effect of Repeated Artificial Inseminations on Community Composition and Structure of Female Reproductive Microbiota

Here we investigate whether insemination results in a change in community composition of the post-insemination female sample compared with that of the virgin sample, in terms of the magnitude of change (either an increase or decrease) in species richness, Shannon diversity or Simpson diversity; the proportion of OTUs shared with the virgin sample and the Bray-Curtis Dissimilarity to the virgin sample.

According to comparisons of linear mixed models based on the log-likelihood ratio, there is no evidence that number of artificial inseminations influenced variation in the similarity of female samples to virgin samples in any of the measures, as inclusion of number of artificial inseminations did not improve explanatory powered compared with the null model, either as a linear or quadratic effect.

According to comparisons of linear mixed models based on the log-likelihood ratio, there is no evidence that number of artificial inseminations influenced variation in the Bray-Curtis Dissimilarity of female samples to a sample collected either one or two artificial insemination events previously, as inclusion of number of artificial inseminations did not improve explanatory powered compared with the null model, either as a linear nor quadratic effect.

However, there was a linear effect of the number of artificial inseminations on the proportion of OTUs shared by female samples with both one and two previous samples, so as the number of inseminations increases. Inclusion of the number of artificial inseminations significantly improved explanatory power of the proportion of OTUs as both a linear (ChiSq = 17.0, df =1, p-value < 0.001) and quadratic effect (ChiSq = 17.0, df =1, p-value < 0.001). However, the quadratic effect was not an improvement on the linear effect. Along with graphical representations of untransformed values, this suggests that after the second artificial inseminations, samples either side of a focal artificial insemination become increasingly dissimilar with repeated artificial inseminations (Fig. 6). The control female does not appear to follow the same trajectory as the ejaculate inseminated females.

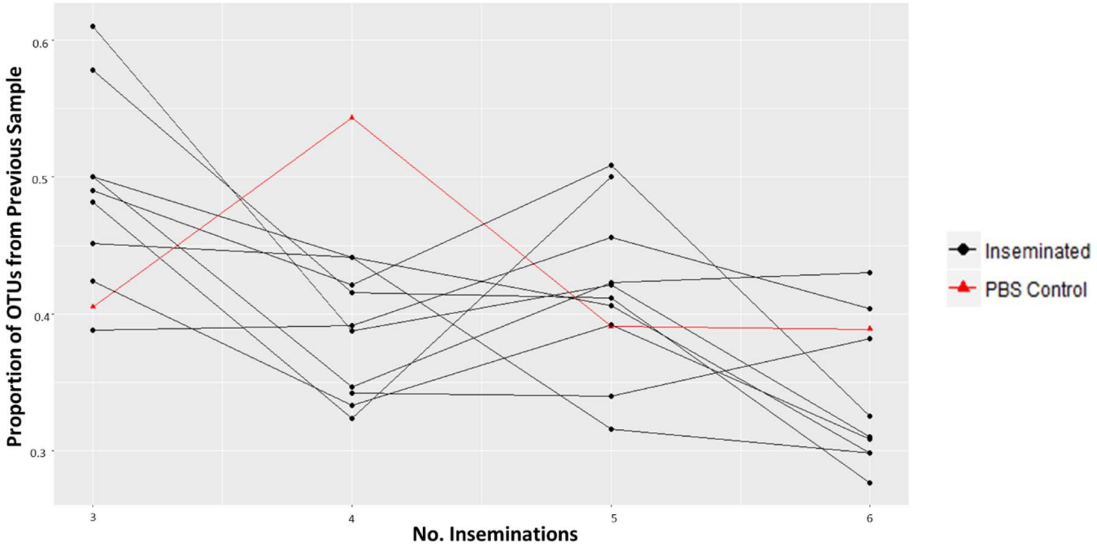


Figure 6 *Figure illustrating the proportion of OTUs in a sample that are also present in the previous sample (one insemination before) from that female. As there is no sample after the first insemination, we can only compare samples after the second insemination. Here, where 'No Inseminations' is 3, this is showing the proportion of OTUs in samples collected after the third insemination that are also found in the sample collected after the second insemination. Other missing points for a female represent a case where either the focal female sample or the sample from before the previous inseminations, or both, are not present in the dataset.*

The same statistical pattern was present for two previous samples; inclusion of the number of artificial inseminations significantly improved explanatory power of the proportion of OTUs as both a linear effect (ChiSq = 11.7, df = 1, p-value = 0.001) and a quadratic effect (ChiSq = 11.7, df = 1, p-value = 0.001). However, the quadratic effect was not an improvement on the linear

effect. Similarly, graphical representations of un-transformed values indicate that following the initial two artificial inseminations samples from the same female separated by two artificial inseminations become increasingly dissimilar as the number of repeated inseminations increases (Fig. 7). However, the samples sharing the smallest proportion of OTUS are those taken from the same female before and after the initial two inseminations, suggesting a large effect of insemination of a virgin female. Further the control sample appears to now have different dynamics from the ejaculate inseminated samples; unlike ejaculate inseminated samples there is no increase in proportion of OTUs shared between the fourth and second samples, compared to the second and virgin samples. The control has one of the highest amounts of OTUs shared between second and virgin samples, and then one of the lowest in the other three comparisons.

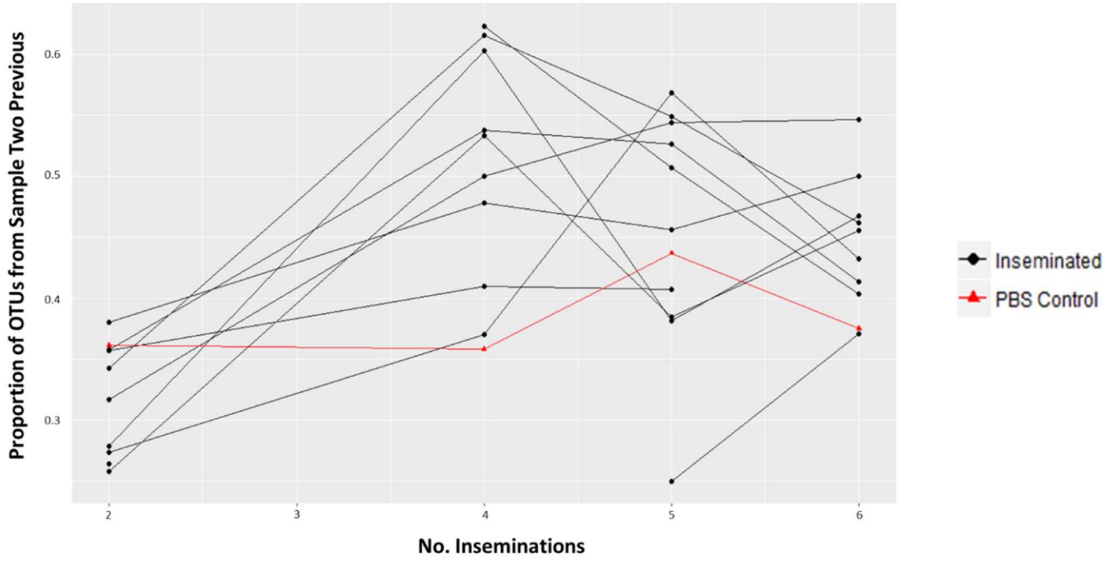


Figure 7 *Figure illustrating the proportion of OTUs in a sample that are also present in the sample two previous (two inseminations before the focal sample) from that female Here, where 'No Inseminations' is 2, this is showing the proportion of OTUs in samples collected after the second insemination that are also found in the virgin sample i.e. collected after zero inseminations. As there is no sample after the first insemination, there are no points where 'No. Inseminations' is 3. Other missing points for a female represent a case where either the focal female sample or the sample from two previous inseminations, or both, are not present in the dataset.*

ii) Change in Community Composition and Structure of Female Reproductive Microbiota in Relation to the Inseminating Male

According to comparisons of linear mixed models based on the log-likelihood ratio, there is no evidence that number of artificial inseminations influenced variation in the proportion of a

female samples OTUs that are shared with the sample used to inseminate them either one or two artificial insemination events previously, as inclusion of number of artificial inseminations did not improve explanatory powered compared with the null model, either as a linear nor quadratic effect.

However, there was a linear effect of the number of artificial inseminations on the Bray Curtis Dissimilarity of female samples to the male sample used to inseminate them two inseminations before (though not the male sample used for the previous insemination) Inclusion of the number of artificial inseminations significantly improved explanatory power of the model as both a linear (ChiSq = 4.43, df =1, p-value = 0.035) and quadratic effect (ChiSq = 4.43, df =1, p-value = 0.035). However, the quadratic effect was not an improvement on the linear effect.

Graphical representations of un-transformed values indicate that Bray-Curtis Dissimilarity of the female to the male sample used two previously increases with an increased number of inseminations (Fig. 8).

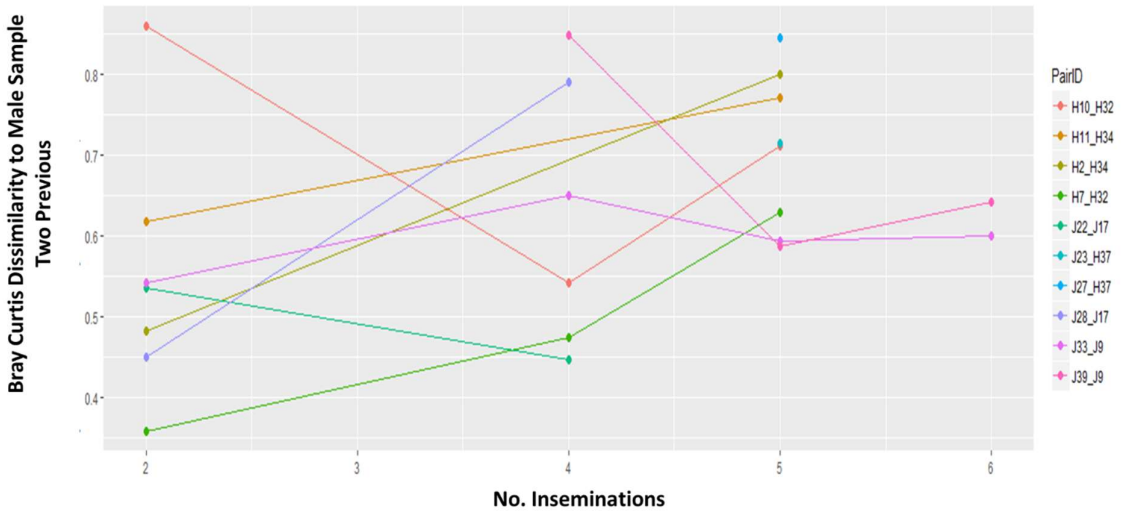


Figure 8 *Figure showing Bray Curtis Dissimilarity of female samples to the male sample used to inseminate her two inseminations previously. At 'No. Insemination' is 2, the female sample collected after two inseminations is compared to the male sample used to inseminate her as a virgin i.e. the sample used after zero inseminations. As no male or female samples were collected after the first insemination there are no points at 'No Inseminations' is 3 as there is no data on the male sample used in the first insemination for the female sample after three inseminations to be compared to. Missing points for a pair represent a case where either the female sample or the matching male sample from two previous inseminations, or both, are not present in the dataset.*

According to comparisons of linear mixed models based on the log-likelihood ratio, there is no evidence that number of artificial inseminations influenced variation in the proportion of OTUs shared by two females inseminated by the same male, or in Bray-Curtis Dissimilarity between the two females, as inclusion of number of artificial inseminations did not improve explanatory powered compared with the null models, either as a linear or quadratic effect.

However, graphs suggest that there is an effect, but that it is not captured by the formal analysis as it is only for the first three inseminations, followed by fluctuations from this point on (Fig. 9 and 10).

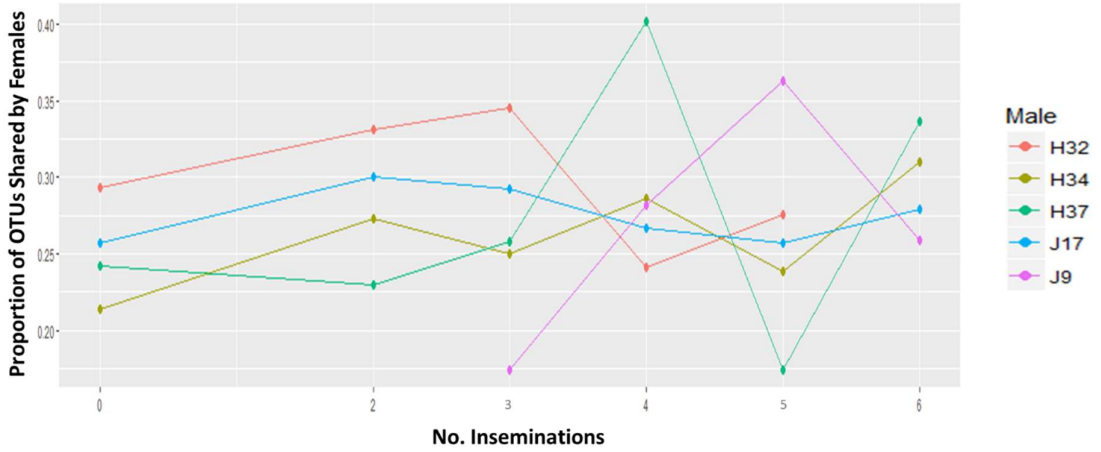


Figure 9 *Proportion of total number OTUs found in both females inseminated by the same male that are found in both females.*

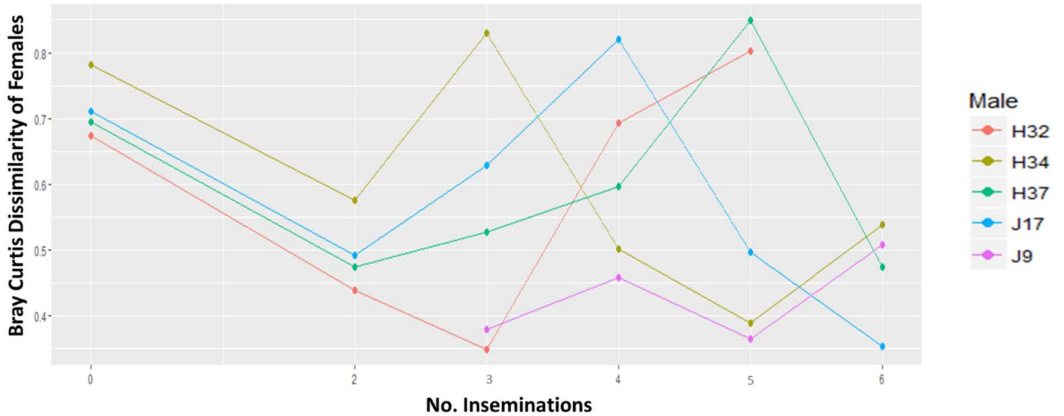


Figure 10 *Bray-Curtis Dissimilarity between the samples of two female inseminated by the same male*

According to comparisons of linear mixed models based on the log-likelihood ratio, there is an effect of number of inseminations on the proportion of male only OTUs in females, but this effect is the same whether the female has been inseminated with that male's semen sample, another male's semen sample or sterile PBS.

Inclusion of the number of artificial inseminations significantly improved explanatory power of the model as both a linear (ChiSq = 8.02, df =2, p-value = 0.018) and quadratic effect (ChiSq = 8.02, df =2, p-value = 0.018). However, the quadratic effect was not an improvement on the linear effect, and inclusion of 'Comparison' i.e. whether pairs are original, random or control, and its interaction with number of inseminations did not improve models either.

Graphical representations of un-transformed values indicate that there is a decrease in proportion of male only OTUs as insemination increases, though there are lots of fluctuations (Fig. 11).

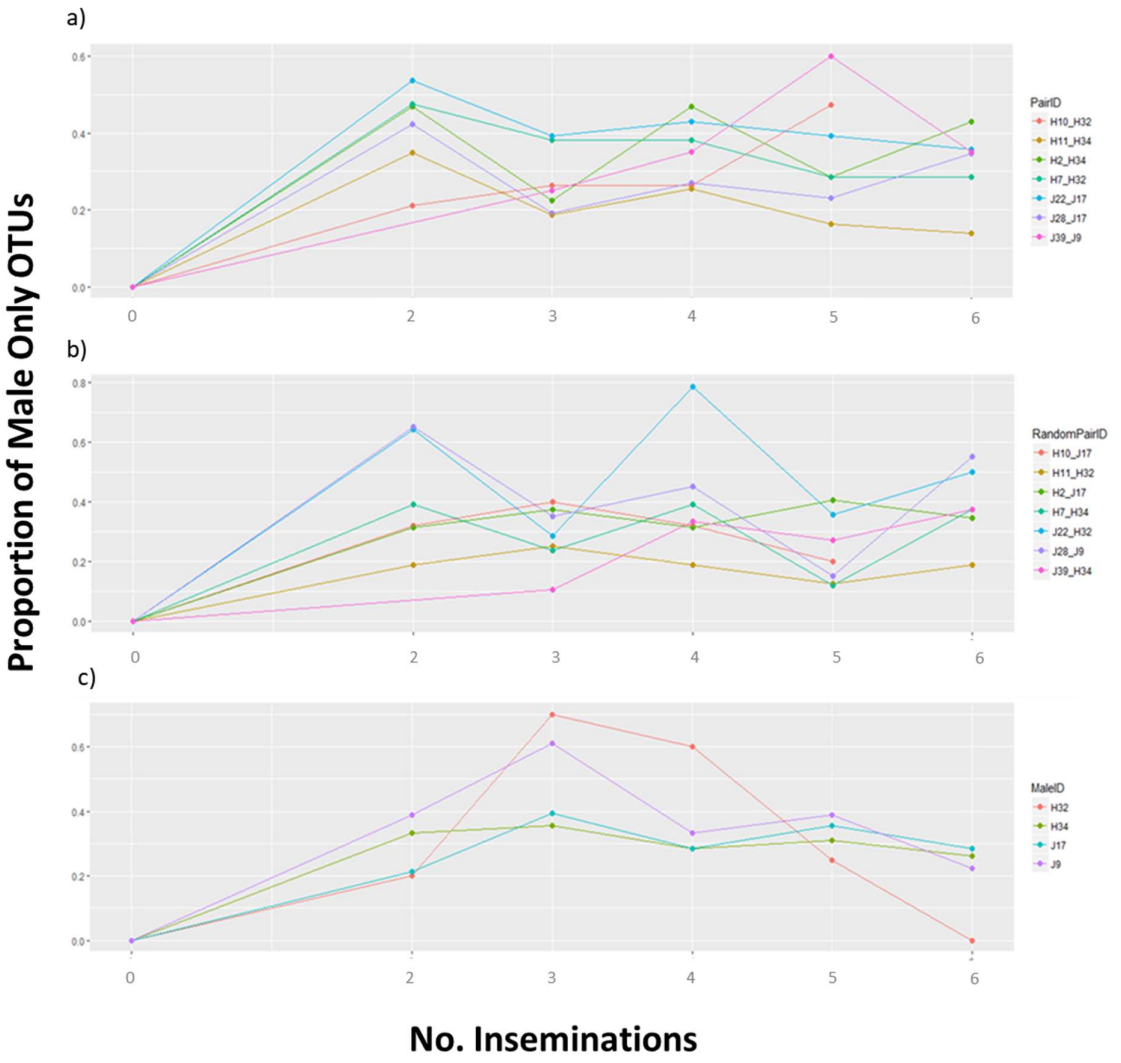


Figure 11 *Graphs illustrating the proportion of Male Only OTUs that are found in a paired female after each insemination event. Each graph represents one of three different kinds of pairs: a) Original pairs where females are compared to a male whose semen samples are used to inseminate them b) Random pairs where semen-inseminated females are compared to a male whose semen was not used to inseminate them and c) Control Pairs where the control female inseminated with sterile PBS is compared to each of the males used to inseminate other females. No Inseminations' indicates the number of inseminations the female has received before the sample used for comparison was collected. Male Only OTUs are those OTUs found in the male sample used to inseminate virgin females, but not in virgin females i.e. is always zero at zero inseminations. Missing points represent a point where the female sample for that No Inseminations is not in the dataset.*

iii) Effect of Cessation of Artificial Inseminations on Community Composition and Structure of Female Reproductive Microbiota

According to comparisons of linear mixed models based on the log-likelihood ratio, there is no evidence that number of artificial inseminations influenced variation in the proportion of a female's samples OTUs shared with the first Post-AI female sample. This is the case when considering the difference in species richness, Shannon diversity or Simpson diversity measures

between a focal female sample and the first post-AI sample of that female, as well as the Bray Curtis Dissimilarity between a focal female sample and the first post AI sample of that female. This is true following either homospermic or heterospermic inseminations. Inclusion of number of artificial inseminations, either as a linear nor quadratic effect. or AI type, either separately or together, interacting, did not improve explanatory power compared with the null models.

iv) Return to Virgin Community Composition and Structure Following Cessation of Artificial Inseminations

According to comparisons of linear mixed models based on the log-likelihood ratio, there is no evidence that number of samples since last insemination (as a measure of time elapsed) influenced variation in the proportion of a female samples OTUs that are shared with the virgin sample of that female. Inclusion of number of samples since last insemination did not improve explanatory power compared with the null model, either as a linear nor quadratic effect.

However, there was a linear effect of the number of samples since last insemination on the Bray Curtis Dissimilarity of female samples to the virgin sample. Inclusion of the number of artificial inseminations significantly improved explanatory power of the model as both a linear (ChiSq = 17.9, df = 2, p-value < 0.001) and quadratic effect (ChiSq = 17.9, df = 2, p-value < 0.001). However, the quadratic effect was not an improvement on the linear effect. Female ID as a random effect explained a lot of variation in both these models (71.8% in both).

Graphical representations of un-transformed values indicate that with an increased number of inseminations, Bray-Curtis Dissimilarity of the female to the virgin sample decreased, though this effect appears slight and to be driven by a few females (Fig. 12).

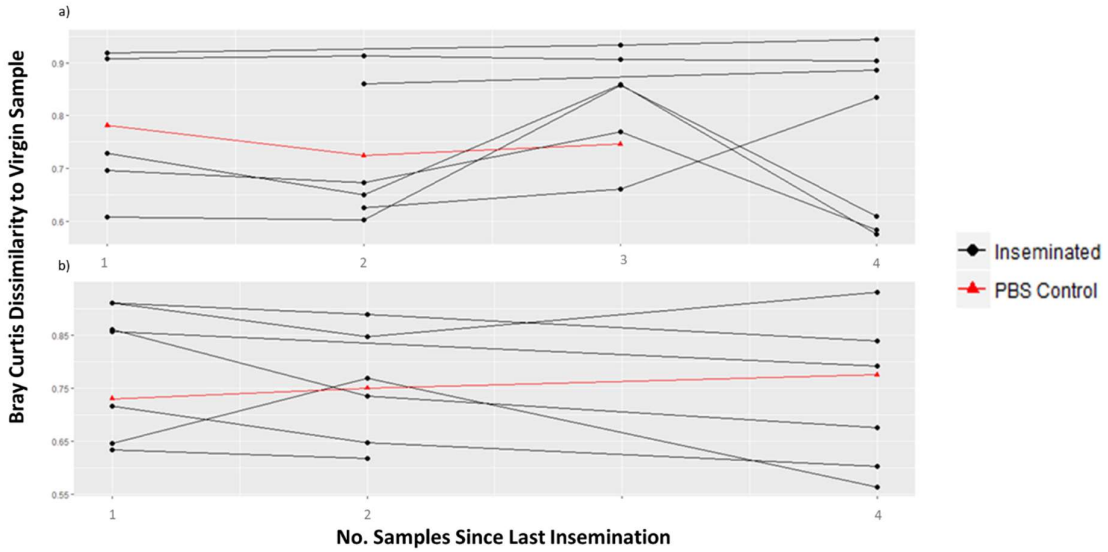


Figure 12 *Graphs illustrating the Bray Curtis Dissimilarity of a female sample to her virgin sample following cessation of a series of a) homospermic inseminations and b) heterospermic inseminations. Samples Since Last Insemination indicate the number of samples that have been collected since the cessation of artificial insemination so where 'Samples Since Last Insemination' is 1, this is the first sample collected since the last insemination of the series.*

According to comparisons of linear mixed models based on the log-likelihood ratio, there is no evidence that number of samples since last insemination (as a measure of time elapsed) influenced variation in Bray Curtis Dissimilarity of a female sample with the male sample last used to inseminate her before cessation of Artificial Inseminations, for both heterospermic and homospermic inseminations. The number of samples since last insemination did not improve explanatory power compared with the null model, either as a linear nor quadratic effect, and nor did an inclusion of an interaction with AI Type.

However, there was a linear effect of the number of samples since last insemination on proportion of OTUs of female samples to the last inseminated male sample, following both types of insemination. Inclusion of the number of samples since last insemination significantly improved explanatory power of the model as both a linear (ChiSq = 5.16, df = 1,p-value = 0.023) and quadratic effect(ChiSq = 5.16, df = 1,p-value = 0.023). However, the quadratic effect was not an improvement on the linear effect, and nor was the inclusion of AI Type. Pair ID as a random effect explained over half the variation in both these models (54.1%).

Graphical representations of untransformed values indicate that with an increased number of samples since last insemination, proportion of female sample OTUs that are also found in the last inseminated male decreases (Fig. 13).

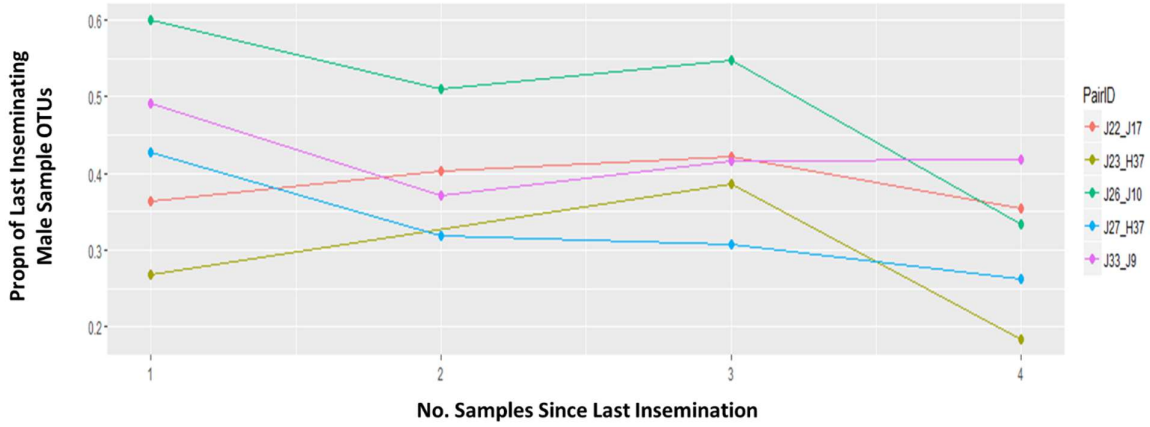


Figure 13 *Graph illustrating the proportion of OTUs in a female sample that are also found in the male sample used for the last insemination of a series of heterospermic inseminations. 'Samples Since Final Insemination of Series' indicate the number of samples that have been collected since the cessation of artificial insemination so where 'Samples Since Final Insemination of Series' is 1, this is the first sample collected since the last insemination of the series.*

DISCUSSION

In this study we explore the potential for insemination with a male ejaculate to affect the reproductive microbiota of the female red junglefowl through investigating changes in community composition and structure of the cloacal microbiota during a series of artificial inseminations, and following their cessation.

We first used all female cloacal samples collected across a twelve-week experimental period, which included three series of artificial inseminations with male semen, though one control female is inseminated with sterile PBS, to look for signatures of different aspects of insemination (virgin and non-virgins, number of inseminations, whether an insemination was with a single ejaculate or a combination of two ejaculates) as well as methodological confounds (collection method and collection date).

We found that there was an effect of virgin status on species richness, with virgin samples having a lower species richness than non-virgin samples, and that in both Principal Component Analysis axes investigated, collection method i.e. 'Tip' or 'Plate', explained variation in scores between samples (Fig. 5).

We then isolated subsets of samples from this dataset, along with corresponding male samples used to inseminate females, looking at specific parts of the insemination series to investigate more closely the effect of artificial inseminations, asking the following questions:

1. Is there a change in community composition and structure in response to insemination, and does any effect increase or diminish with repeated inseminations?
2. Is any effect as a result of becoming more similar to the inseminating male i.e. due to transmission of OTUs, or transmission of factors favouring certain OTUs, or just a disruption of the existing female microbiota?
3. Is there a change in community composition and structure following the cessation of inseminations?
4. Does any change represent a return towards the virgin community composition or structure?

We found that, whilst samples did not become less similar to virgin samples with repeated inseminations, or indeed more similar to the inseminating male sample, with repeated inseminations, insemination did seem to have an effect on community composition and that it seems that this effect is due to both insemination itself (with PBS or semen) and introduction of an ejaculate, and that the effect is greatest when inseminating virgin females. We also found that community composition changed following the cessation of artificial insemination, with female samples becoming more similar to their virgin sample and decreasing in the proportion of their OTUs shared with the last inseminating male ejaculate sample.

Consistent with previous studies of the avian reproductive microbiota (e.g. Kreisinger *et al*., 2015; White *et al*., 2010; Stewart & Rambo, 2000), we find evidence of a change in cloacal microbiota of the female in response to insemination, and it appears that this change is mostly due to the initial insemination of a virgin female. In the first analysis, virgin status, but not number of inseminations, predicted species richness (Fig. 5c). Further, in the second part of the analysis when we consider the proportion of OTUs of a female sample shared with her samples two inseminations previously, the proportion of OTUs shared between the virgin sample and the sample after two inseminations is much lower than the proportion shared between those taken after two and four inseminations, three and five inseminations and six and four inseminations (Fig. 7). This suggests that there is disruption or transmission of microbiota after the first insemination of a female that is greater than any disruption as a result of subsequent inseminations. The fact that following this first virgin insemination, samples share a decreasing proportion of OTUs with the previous female sample (Fig. 6) suggests a continued disruption with a cumulative effect due to mating, though not directionally as there was no evidence of becoming increasingly dissimilar to the virgin sample with a repeated number of inseminations, as one would expect with a directional, 'additive' effect of insemination. This lack of directionality could be due to fluctuations of microbiota within the inseminating male, such that a different microbiota is inseminated into the female each time.

These changes with number of inseminations could be due to an effect of insemination, rather than introduction of semen specifically, or a seasonal change in microbiota unrelated to inseminations. However, graphical comparisons of semen-inseminated females with a control female inseminated with sterile PBS (Fig. 6 and 7) shows that the control female remains constantly similar to the previous sample and the sample two previous and is the most similar to the virgin sample following the first two inseminations, suggesting that not only is the effect not due to seasonal effects, but is specifically due to introduction of semen and not simply insemination that is causing changes in OTU proportions.

These results are consistent with mating effect evidenced in the previously cited avian studies in providing evidence of change in the female microbiota in response to insemination, but they also differentiate between the first insemination of a virgin and subsequent inseminations. Differences in the effect of a mating when a female is a virgin is relevant to considerations of a mating effect, as the cost or benefit of a mating to a female could depend on whether or not she has been previously mated, and parallels could be drawn between this and the cost of mating following infection; if you are already disrupted then the cost of subsequent matings is reduced if each mating does not further disrupt, just as if you are already infected further mating with infected individuals will not carry any further cost in terms of exposure to infection. It also highlights the relevance of including virgin samples, as it could mean that in such a system, investigation of non-virgin samples may not reveal a cost of mating.

Further this study isolates sexual transmission specifically due to insemination from an overall effect of co-habitation or of mating, as in White *et al*. (2010), but also distinguishes between an effect of insemination and effect of insemination with semen. This is relevant, as if it is only insemination i.e. the introduction of any foreign matter, and not the ejaculate introduction, that is causing change in the female microbiota then variation in e.g. bacterial load of semen which has been shown to vary across human males (e.g. Korhonen *et al*., 2017) is not relevant to the predictions or conclusions a study might make regarding the effect of insemination with different ejaculates.

It does not appear that this change in proportion of shared OTUs is due simply to transmission of OTUs from male to female as there is no increase in the proportion of OTUs shared with the male inseminating semen sample, or in the proportion of 'Male Only OTUs' (i.e. those found in the first inseminating male sample, but not the virgin sample) as inseminations progress to explain the patterns of proportions of OTUs shared with previous female samples. This lack of change could be due to the fact that the ejaculate from the same male is different at each insemination, and so the female is not repeatedly being inoculated with the same microbiota,

rather than there being a lack of transmission. However, there is no statistical evidence that pairs of females inseminated by the same male semen sample become more similar to each other, suggesting that there is no transmission, though graphical visualisations of the data do suggest an initial increase in similarity (Fig. 9 and 10). In fact, there is an increase in Bray-Curtis Dissimilarity between female samples and the ejaculate used to inseminate her two inseminations previously as inseminations progress (Fig. 11), suggesting that disruption as a result of inseminations is leading to the female becoming less similar to the inseminating male sample, perhaps due to an increase in disruption preventing the effect of insemination being detected. Similarly, there is actually a decrease in the proportion of the Male Only OTUs found in a female with an increased number of inseminations (Fig. 12), and this is the case whether we are considering Male Only OTUs from semen used to inseminate that female, or Male Only OTUs when the female is compared to another male who did not inseminate her, and considering change in the control female. This suggests that the presence and subsequent decrease of Male Only OTUs in later female samples is not due to transmission, but due to a natural turnover of OTUs meaning that as inseminations progress, the proportion of any given set of OTUs found in the cloacal microbiota decreases, perhaps due to an increased species turnover as a result of continuing inseminations. In this case it is more difficult to discount a seasonal effect, but we can conclude that if it is due to insemination it is due to the insemination effect itself as well as introduction of an ejaculate given that the control female follows the same pattern as ejaculate inseminated females (Fig. 11c). Other studies have found evidence suggesting increased similarity in males to females following mating. Kreisinger *et al*. (2015) demonstrated increased similarity of reproductive microbiota in breeding pairs of barn swallows compared with pairs of birds from different breeding pairs, but there is no separation of the effect of mating and co-habitation, and also no information regarding the similarity of breeding pairs prior to the breeding season. Similarly, in humans Mändar et al. (2015) showed that similarity between reproductive microbiota of mates increased following mating, but cannot distinguish between the act of mating and insemination. However, in this study there is no separation of mating and insemination.

It is important to highlight here that being unable to differentiate between the different components of mating that could affect the microbiota (i.e. co-habitation, mating, insemination, introduction of a specific ejaculate or seminal factor) does not negate conclusions regarding the effect of mating; there is much more to mating than just the transfer of an ejaculate and any component contributing to variation in the reproductive microbiota is of interest. However, it is important to appreciate that predictions and conclusions surrounding the relationship between mating and the reproductive microbiota will usually be at least partly dependent upon which component is driving variation. For instance, if similarity between the reproductive microbiota of breeding pairs is driven by co-habitation and not mating itself, then the role of extra-pair copulation in driving variation in a female's microbiota is likely to be less important. Similarly, if it is the disruptive effect of mating and insemination, rather than transmission of an ejaculate and its contents, that results in a mating-induced change then when predicting an effect of mating then it would be more relevant to consider the rate of female mating when investigating the female microbiota in that system, rather than the identity of her mates. This understanding then also reinforces the idea that it can be important to consider dynamics of the microbiota as a community, and not just as individual sexually transmitted micro-organisms, when considering the cost and benefits of mating events.

Interestingly, dynamics of the female reproductive microbiota following the cessation of artificial inseminations seem to complement, and in fact expand upon, the results discussed so far. There is a slight, but significant, decrease in Bray-Curtis Dissimilarity to the virgin sample following cessation (Fig. 12). This suggests that following cessation of inseminations this change is lost. However, it does not seem to be the case that the proportion of virgin OTUs in a female increases with number of samples following insemination cessation, but it could be that OTUs that dominate in a virgin female sample, begin to increase in relative abundance, or OTUs which are gained or increase in relative abundance as a result of insemination decrease in relative abundance, but are not lost. Due to the variation between females, it is difficult to conclude from the graphical comparison with the control as to whether this is an effect of insemination with

ejaculate or simply insemination. However, it is unlikely to be a seasonal effect as we see the same pattern, and similar levels of dissimilarity, across two sets of samples collected at different time points following homospermic (Fig. 12a) and heterospermic (Fig. 12b) inseminations.

Further, we found a decrease in the proportion of OTUs found in the female also found in the inseminating male following artificial insemination cessation (Fig. 13), consistent with White *et al*. (2011), who found females lost OTUs also present in their male partner at a faster rate than other OTUs when insemination was prevented. This is unlikely to be simply a temporal effect resulting in loss of any set of OTUs, as we do not see a similar decrease in the proportion of OTUs shared with the female's first post AI sample. This therefore suggests that the change we observe in females following insemination as a virgin is due in part to accumulation of OTUs through insemination, though this seems to contradict results showing a decrease in Male Only OTUs and proportion of OTUs from previously inseminating ejaculate. This again is perhaps due to a large effect of the first insemination on a non-virgin sample, and no effect of subsequent disruption in increasing proportion of male OTUs. It could also be that the multiple coinciding effects of insemination leading to disruption of a community, as well as gain of OTUs, obscures the identification of all these changes in isolation. In contrast, following cessation of inseminations, change is not due to repeated disruptions but simply the increase, decrease and loss of OTUs and thus the signal of changes in OTU proportions is clearer. It is also strange that we do not see this change reflected in the proportion of OTUs shared with the first post-AI female sample. This could perhaps be that the female does lose male OTUs that are shared with first post AI-sample, but also regains other OTUs that are lost and re-occur in cloacal samples in underlying fluctuations of the female reproductive microbiota.

The complementary patterns demonstrated here of increasing similarity to virgin samples and loss of male OTUs further adds to the conclusions of White *et al*. (2010), who posit that loss of male OTUs represents a return to a stable, functional virgin state. Other studies have also talked of a 'stable' virgin state including Ravel *et al*. (2011), who interpret a mating-induced

change to the human female microbiota as a deviation away from a functional, protective microbiota, but as neither study have investigated virgin samples this remains inexplicitly tested. The pattern here is quite weak, and the nature and extent of the change unclear; similar work, expanding on the current study is required. This result is also of interest in the more general context of host intrinsic, or even microbiota intrinsic, drivers of community structure and composition; a microbiota returning to a pre-disruption state could represent selection by host factors to achieve a functional microbiota, or a more intrinsic change in microbiota composition towards a more stable bacterial ecosystem.

This study, whilst expanding on previous work through the combination of virgin samples, matching male samples, repeated inseminations and samples post insemination, also raises further questions as to the nature and cause of these changes, presenting more avenues for exploration, and there are limitations to our approach. Firstly, whilst the breadth of this dataset provides samples across several weeks with a number of series of insemination, cessations of inseminations and types of artificial insemination means we are able to ask a number of different questions, that fact that we are missing a number of samples through loss at various stages in the process, may mean we do not have the power to detect patterns in and isolate sources in female microbiota variation. More focused datasets dedicated to more thoroughly investigating individually the number of interesting dynamics surrounding insemination highlighted in this study are now needed.

Additionally, whilst inseminations are closely controlled, there is potential for variation from other sources of variation that we cannot account for here. For example, in humans (Ravel *et al*., 2011) and baboons (Miller *et al*., 2017) menstrual or ovarian cycle and reproductive status have been shown to affect the composition of reproductive microbiota. It is therefore conceivable that fluctuations in female physiology or behaviour due to her reproductive cycle could produce variation in the cloacal microbiota and confound or obscure insemination-associated change. Further, as in the junglefowl the cloaca is the common opening for the reproductive, digestive and

urinary tract this may also provide fluctuations in the female microbiota associated with, for instance, how recently they have eaten or defecated. Studies building on the findings here, then, could benefit from consideration of such potential sources of variation, through measurement or manipulation, to add granularity to the exploring the relative effect of mating in driving fluctuations in the red junglefowl female reproductive microbiota.

In conclusion, through the use of virgin samples, matching male samples and repeated sampling of females as well as isolation of insemination as opposed to mating have produced results consistent with, and expanding on, previous work in the field. We have demonstrated an effect of insemination, as opposed to mating, upon the reproductive microbiota of the female red junglefowl, particularly the initial insemination of a virgin female, and demonstrated a potential return towards the virgin state with loss of male OTUs following the cessation of inseminations. The effect however is not a simple case of a consistent directional change of each female in the same direction, or of simply becoming the same as the inseminating male seminal microbiota, highlighting the complexity of these effects and thus the requirements of carefully controlled experiments to disentangle and understand the multi-faceted effect of insemination. Additionally, the effect of collection method, and differing conclusions from the different aspects of community structure and composition highlight the need for careful consideration of the method and metrics used both in planning in experiments, and in the comparison to and interpretation of existing studies.

CHAPTER 6. DISCUSSION

Recent theoretical and empirical work suggests that microbiotas can play important roles in host health (Cho & Blaser, 2012), reproductive function (e.g. Nelson *et al*., 2015) and reproductive success (e.g. Thrall *et al*., 2000), with potentially fundamental consequences of reproductive microbiotas for mating strategies, sexual conflict, sexual selection on males, and the evolutionary ecology of mating systems (Sheldon, 1993, Kokko *et al*., 2002). However, our knowledge of reproductive microbiotas is currently patchy, and largely limited to clinical studies in humans.

This thesis sought to develop a next-generation sequencing approach to describe reproductive microbiotas in a sexually promiscuous population of red junglefowl, characterise patterns of intra- and inter-individual variation, and explore drivers of such variation. Most of this work focused on the male seminal microbiotas, although in the final data chapter, I expand this work to study the microbiotas of the female cloaca. The work started by describing and comparing communities of bacteria present in male semen, which enabled me to study how the structure and composition of seminal microbiotas changed within an individual male over time, and between individual males in relation to their age, cohort, immuno-genetic make-up, and mating history. I combined the next-generation sequencing approach with *in vitro* culturing techniques, to demonstrate that the bacteria contained in red junglefowl semen are alive, and explore the way culturing environmental conditions affect the composition and structure of the cultured community. The work then extended the next-generation sequencing approach to analyse cloacal samples taken from females, to investigate how inseminations affect the female reproductive microbiota, and if there was evidence of sexual transmission of microbiota members from the male to the female. This allowed me to establish the potential for reproductive microbiotas to influence the evolutionary ecology of mating systems.

Main Findings

Our next-generation sequencing approach addressed a number of methodological issues, which affect microbiota studies including different sources of contamination, amplification bias and uneven sampling depth. Whilst we were unable to eliminate these challenges entirely, the newly developed pipeline offers a methodology to reduce environmental contamination, identify residual (likely reagent-associated) contaminants and deal with them analytically, as well as providing a protocol to systematically quantify issues related with amplification bias and uneven sampling depth, and deal with these analytically. This is a promising development, which enabled us to describe distinct reproductive bacterial communities and explore variation between them in order to carry out the aims of the thesis. This work indicates that natural ejaculates contain diverse bacterial communities, which are to some degree repeatable across different males and largely reflect gut microbiotas, as suggested by complete overlap with faecal microbiota.

Through the complementary use of *in vitro* culturing techniques, we found that a large proportion - 60% - of directly-sequenced (Operational Taxonomic Units) OTUs could be identified through culturing. Thus, a large proportion of the bacteria making up the red junglefowl seminal microbiota is alive, which has implications both for the potential effect of the seminal microbiota on males and their spermatozoa, but also for the female to whom this live bacteria is being transmitted during insemination. However, it was also revealed that 49% of the cultured OTUs were not found through direct sequencing of semen samples, and that the structure and composition of cultured communities varied dependent on the culturing medium used and the presence of oxygen. These findings further emphasise the value of combining *in vitro* culturing with next-generation molecular techniques in order to fully explore a transient, complex and phenotypically diverse community.

In investigating potential drivers of within and between individual variation in seminal microbiota, we found that the sole predictor of variation in community composition between 117 samples taken from 51 males of our junglefowl population over two breeding seasons was the

allelic richness and amino acid divergence at the BLB locus of the MHC. This is consistent with an immuno-genetic determinant of the microbiota, as this locus segregates for alleles involved in recognition of extra-cellular, non-host (i.e. non-self) cells such as bacteria. A genetic determinant of the microbiota could allow inheritance of, and selection for and against, given microbiota compositions and structures, and also provide a mechanism of mate selection and kin recognition through a microbiota-mediated cue of genetic information about a conspecifics and potential mates (see below).

We looked more specifically at mating as a driver of variation between seminal microbiotas using semen samples taken from males following inclusion in one of six mixed-sex mating trial groups, where males and females were allowed to interact freely and the majority of mating interactions during the trial were recorded. We found that males belonging to different mating trial groups significantly differed in community composition and structure, and that the extent of variation of microbial communities within mating trial groups was correlated with the group's sexual network metrics. This is suggestive of a mating effect on the seminal microbiota, though there was no effect of the mating metrics of individual males during a mating trial upon community structure and composition. This does not rule out an effect of mating. For example, a recent study of a social spider, *Stegodyphus dumicola*, concluded that the behaviour of infected individuals has less effect in predicting transmission dynamics of bacteria across a population than the social structure of the population. Group size and structure, associated with density, modularity and average path length of the group social network, predicted transmission dynamics of an introduced bacterial infection in the group, whereas behavioural traits of the index case did not (Keiser *et al*., 2017). However, our results do warrant strong consideration of the possibility that differences between males in different mating trial groups are due to non-mating aspects of the group correlated with sexual network metrics such as general activity of the group, average male condition or average hormone levels in males.

We also investigated the effect of insemination on the female red junglefowl reproductive microbiota, and found that the cloacal microbiotas of virgin females were less species-rich than non-virgin females and that insemination did seem to have an effect on community composition. This effect seems to be due to both, the physical act of insemination itself (with PBS or a semen sample) and the effect of the introduction of an ejaculate, and is greatest when virgin females are first inseminated, compared with subsequent inseminations of the same female. Further, community composition changed following the cessation of an artificial insemination series, with female samples becoming more similar to their virgin sample and decreasing in the proportion of their OTUs shared with the last inseminating male semen sample.

Implications of Findings

The demonstration of live bacteria in male semen provides evidence of a reproductive microbiota with the potential to affect the fitness of a male and his mates, through transmission to a female, or through affecting the chances of fertilisation of a mate by the male's sperm. The further demonstration that this microbiota could be at least partially genetically determined, and evidence suggesting that reproductive microbiotas of both males and females can be disrupted by mating, is further evidence that the reproductive microbiota may be an important consideration in studies of the evolutionary ecology of mating behaviours and systems.

First, the notion that the bacterial community of a host can be disrupted by mating identifies a potential new factor modulating the costs and benefits of mating. If the reproductive microbiota of an individual is at a stable, functional state but mating can change the structure or composition of this community and interrupt function as suggested by studies of the temporal dynamics of the vaginal microbiota (e.g. Gajer *et al*., 2012 ;Vodstrcil *et al*., 2017) then mating is costly to an individual. Equally, if such disruption results in a more favourable community or structure, perhaps through the transmission of beneficial microbiota members as suggested by Lombardo *et al*. (1999), then mating is beneficial. Thus, in work seeking to understand the evolution of certain mating behaviours or systems, more clarity may be gained from considering

the role of reproductive microbiotas in mediating the advantages and disadvantages of different host decisions and actions. Whilst in chapter 5 we have provided some evidence for an effect of insemination on female reproductive microbiotas, and that this effect may be context dependent (we showed a more disruptive effect of insemination in virgin females, compared with subsequent inseminations), further work exploring the exact nature of disruptions in different contexts would allow more definite predictions as to the role the reproductive microbiota may play in driving the evolution of mating systems.

As well as affecting the costs and benefits of a mating directly, it is also possible that the microbiota could provide cues as to whether the mating is likely to be costly or beneficial to an individual. For instance, as it has been demonstrated that microbiota can provide olfactory cues as to an individual's social group in hyenas, *Crocuta crocuta* (Theis *et al*., 2013) and as to relatedness of individuals in meerkats, *Suricata suricatta* (Leclaire *et al*., 2017) through volatile fatty acid production, and allow mate choice in the fruit fly, *Drosophila melanogaster* (Sharon *et al*., 2010). Bill coloration is associated with antibacterial activity in mallards, *Anas platyrhynchos* (Rowe et al., 2011) and therefore potentially infection status of a male, and pathogenic infection of the male red junglefowl affects secondary sexual characteristics which determine female mating preferences (Zuk *et al*., 1990). If the reproductive microbiota could similarly provide an olfactory cue, or affect sexual signals, as to the quality of a potential mate through conveying information as to health, fertility, genetic make-up or infection status it would therefore inform the signal receiver as to the relative costs and benefits of a mating with a given individual. The evidence for a genetic basis of the microbiota described in chapter 4 of this study potentially provides such a microbiota-mediated mechanism to provide either a pre- or post-copulatory cue to females as to genetic quality or relatedness of a potential mate, as well as the likelihood of carrying infection.

The reproductive microbiota could further drive sexual selection through determining whether or not a male's sperm successfully fertilises a female's ova. The reproductive microbiota

could mediate success of fertilisation in a number of ways; the bacteria in an ejaculate or in the cloaca of a female could affect the ability of the spermatozoa to reach or fertilise the ova as suggested by a detrimental effect of bacteria on sperm performance *in vitro* (e.g. Haines *et al*., 2013; Donoghue *et al*., 2004) or bacteria contained within a male's ejaculate could elicit a female immune response with spermatozoa being caught in the cross fire. If there is an immuno-genetic determinant of the reproductive microbiota of both males and females, then this could determine the fertilising compatibility of males and females. In this scenario, males with an MHC haplotype similar to a female are less likely to carry bacteria which will elicit an immune response in the female following insemination, and thus may be more likely to fertilise her ova. This hypothesis predicts that MHC-similar mates would be favoured through post-copulatory sexual selection, contradicting the current predictions that MHC-based mate choice should select for MHC-dissimilar mates due to increased MHC diversity of the offspring being protective against pathogenic infection (e.g. Worley *et al*., 2010). There is evidence of mate choice for MHC dissimilar mates in the red junglefowl (Gillingham *et al*., 2009, Løvlie *et al*., 2013) but there is also evidence of MHC-assortative parent pairs as observed in wolves, *Canis lupis* (Galaverni *et al*., 2016) and badgers, *Meles meles* (Sin *et al*., 2015). These latter cases, where parent pairs are more MHC similar to each other than expected at random, could be explained by this mechanism of microbiota driven post-copulatory sexual selection favouring fertilisation, and variation between species as to which is the most powerful selective force – microbiota compatibility or a heterozygote advantage - may explain variation in whether offspring are the result of selection for MHC assortative or disassortative parent pairs.

Limitations and constraints of the present study

My work indicates that the next-generation sequencing methodological protocol can be further refined and developed, particularly to tailor different aims of microbiota studies. Whilst this methodology represents a parsimonious approach to provide a systematic and unbiased description of patterns of variation in bacterial communities within and between individuals, this coverage is unlikely to be a complete representation. Rarefaction curves in chapter 2 reveal that

OTUs continue to be accumulated past the OTU level that we rarefied to, albeit at a low and decreasing rate, and OTUs not detected by direct sequencing were detected in cultured semen samples so we know there are a great number of rare OTUs that we may have missed. Therefore, if the aim of a study was to more fully characterise and explore the members of a reproductive microbiota, the inclusion of all the sequence reads generated for a sample may be advisable, rather than rarefying to an even sequencing depth to allow comparison between samples. Additionally, one could spike communities with known concentrations of bacterial strains, or perform qPCR on selected community members in order to more accurately measure the actual amounts of each bacterial taxa in semen. Further, the creation and sequencing of simulated communities made up of known relative concentrations of a variety of bacterial taxa would give a measure of how closely the sequenced community represents the original bacterial community of a sample, and whether the relative abundances of certain taxa are likely to be under or overestimated by our protocol.

As the focus of this work was the reproductive microbiota at the site of mating, we did not explore the physical origins of these bacteria from within the hosts, nor the physical extent of any mating-induced disruption. Dissections of the male and female reproductive and digestive tracts would therefore be of interest in answering questions relating to this, for instance if one wanted to understand the potential impacts of mating-induced disruption on host health and function, or if one wanted to look for the potential of sexual transmission of gut bacteria. Through sampling the microbiota of six sites along the human female reproductive tract, Chen *et al*., (2017) extend knowledge of the female reproductive microbiota beyond that gained from looking just at distal (vaginal) samples, revealing a continuum of distinct microbiotas along the tract, and bacterial taxa and functions which correlate with menstrual cycle and reproductive disorders. One could also characterise the functional genome of these communities i.e. the suite of genes and subsequent functions provided by the combined genomes of all microbiota members, as has been done, for example, for the gut microbiota of humans (Arumugam *et al*., 2011) and for the faecal and seminal microbiotas of mice, *Mus musculus* (Javurek *et al*., 2016), and Gajer *et al*. (2012)

described the metabolic profile of the human vaginal microbiota. Such work combined with work such as ours considering drivers of variation in microbiota composition may shed light on the likely impact of such variation in terms of host function; do they gain or lose genes likely to play a function in host reproductive health as a result of mating, or are those bacteria transmitted likely to be pathogenic or symbiotic given their genome components?

The investigation of drivers of variation in the seminal microbiota of males (chapter 4) offers promising evidence of an effect of the genes of the Major Histocompatibility Complex (MHC) on reproductive microbiota composition. However, the scope of this study was constrained by limited variation in MHC genotypes (and consequently of MHC metrics) in our available red junglefowl males. There were also several other potential sources of variation possibly obscuring an immunogenetic effect; age, cohort, seasonal effect and recent inclusion in a mating trial group were included in analysis but not controlled, and so whilst they were shown to not contribute significantly to variation, variation in composition as a result of these they may still add sufficient noise to the data to prevent detection of an MHC genotype effect. A more focused and systematic sampling of males to give a wider range of MHC genotypes, and controlling more for these other potential sources of variation would provide a more powerful analysis to expand upon these promising results and further explore any immunogenetic effect. Wider genetic and relatedness information about the males would also allow clarification as to whether any association between MHC metrics and microbiota composition was a result of immuno-genetic mediation, or a result of a wider whole genome effect or of other factors correlated with the relatedness of males i.e. behavioural traits have been shown to be heritable in the red junglefowl (Agnvall *et al*., 2012) but may also influence the microbiota through increased exposure to novel sources of infection of more exploratory birds in the form of social interactions or new environments or food sources. Such a study could also be combined with characterisation of different sample types i.e. faecal samples, with gut and reproductive tract dissections, and with investigations of evidence of immune cell activity to begin an exploration of where and how any immuno-genetic mediation of the reproductive microbiota might occur.

In the second part of chapter 4, the investigation of an effect of mating on the seminal microbiota, only samples collected following the completion of a mating trial were collected; analysis of male samples prior to inclusion in the mating trial group would greatly enhance this analysis. Upon initiation of the mating trial it is likely that variation in the seminal microbiota of males already exists, and this may prevent the detection of mating effects in samples collected at the end of the trial. For instance, increased similarity of male seminal microbiotas relative to their initial similarity due to mating with the same females may not be detected because whilst they are more similar to each other than previous to the trial, they are no more similar to each other than they are to males who haven't mated with the same females but were more similar to each other prior to the trial, perhaps because they have been kept in the same pen, or have a similar past mating history. Existing variation could also lead to differences in mating behaviour during the trial i.e. if microbiota variation is linked to a trait that influences mating behaviour or access to females. For example sexually selected sexual ornaments which influence female mate choice have been linked to antibacterial activity in mallards (Rowe et al., 2011) and parasite infection in the red junglefowl (Zuk *et al*., 1990), and so it may be possible that variation in microbiota is associated with variation in sexual ornaments and this influences variation in mating success.

Further, samples from 'control' males could differentiate between an effect of mating on the seminal microbiota and other potentially confounding sources of variation, for instance if males in the same mating trial group are more similar to each other due to the shared environment as birds eat the same food and come into contact with each other and each other's faeces. Social group membership predicts the similarity of gut microbiota composition in red-bellied lemurs, *Eulemur rubriventer* (Raulo *et al*., 2017) and savannah baboons, *Papio cynocephalus* (Tung *et al*., 2015). Sampling of males kept in 'mating' trial groups with no females could allow further distinguishing of a co-habitation or seasonal effect from the effect of female exposure and mating during a mating trial group. Further a mixed male-female trial group, where insemination is prevented through the use of harness worn by males as was used to prevent

inseminations by male kittiwakes in (White *et al*., 2010) could separate an insemination effect from an effect of exposure to females regardless of mating.

With regards as to the investigation of an effect of insemination on females (chapter 5), the evidence indicating an insemination effect upon the cloacal microbiota of females is somewhat limited by the patchy nature of the dataset and a scarcity of control females (i.e. females inseminated with PBS). A more concerted sampling regime with consistent sampling methods could allow for a more complete dataset with more power to explore our hypotheses. Importantly, the availability of more control females would mean that samples from these females could be included in statistical analyses explicitly testing the difference between insemination and insemination with semen. Different types of control females could also be used, including females, which are not inseminated at all, to differentiate between changes due to insemination and those due to seasonal variation and random temporal fluctuations, and females who are mated naturally by males to distinguish between insemination and the entire act of mating, and also give a more natural picture of the full effect of mating on reproductive microbiotas.

CONCLUSION

In this work the description of the red junglefowl seminal microbiota and patterns of intra- and inter-individual variation have revealed an ecologically complex, highly variable and temporally dynamic seminal microbiota, a large proportion of which is live bacteria. This has implications both for the methodological design and interpretation of microbiota studies, as well as the study of the evolutionary ecology of mating systems.

The demonstration of variation of microbiota composition and structure with culturing condition, and between different semen sample types, highlights how a combination of complementary methods and approaches is necessary to fully understand the composition, ecology and potential functions of microbiotas. Appreciation of such variation is also necesaryry to properly inform the design, and interpret the results, of work studying such communities.

A small amount of variation in seminal microbial composition is explained by haplotype at a single MHC locus, potentially suggesting an immune-genetic determinant of ejaculate microbiota which could provide a genetic basis of selection for and against given microbiota compositions, and also provide microbiota-mediated mechanisms of mate choice and kin recognition. Evidence of live bacteria in male semen suggests important consequences of the seminal microbiota for sexual selection through mediating the outcome of sperm competition, and through affecting the costs and benefits of mating if transmitted the female. An effect of insemination on the female's reproductive microbiota provides further potential for reproductive microbiotas to mediate the costs and benefits of mating, and drive sexual selection.

This thesis therefore emphasises how the understanding of the composition and function of the seminal microbiota, the patterns and drivers of compositional variation, and the consequences of such variation for mating outcomes, is necessary to contextualise the design, interpretation and conclusions of studies of the evolutionary ecology of host mating systems.

REFERENCES

Agnvall, B., Jöngren, M., Strandberg, E., Jensen, P., 2012. Heritability and Genetic Correlations of Fear-Related Behaviour in Red Junglefowl–Possible Implications for Early Domestication. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0035162

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12, R18. https://doi.org/10.1186/gb-2011-12-2-r18

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andere, C., Monteavaro, C., Palacio, M., Catena, M., Rodríguez, E., Collins, A., 2011. Apis mellifera semen: bacterial contamination and susceptibility to antibiotics. Apidologie 42, 551–559. https://doi.org/10.1007/s13592-011-0051-7

Antwis, R.E., Griffiths, S.M., Harrison, X.A., Aranega-Bou, P., Arce, A., Bettridge, A.S., Brailsford, F.L., de Menezes, A., Devaynes, A., Forbes, K.M., Fry, E.L., Goodhead, I., Haskell, E., Heys, C., James, C., Johnston, S.R., Lewis, G.R., Lewis, Z., Macey, M.C., McCarthy, A., McDonald, J.E., Mejia-Florez, N.L., O'Brien, D., Orland, C., Pautasso, M., Reid, W.D.K., Robinson, H.A., Wilson, K., Sutherland, W.J., 2017. Fifty important research questions in microbial ecology. FEMS Microbiol. Ecol. 93. https://doi.org/10.1093/femsec/fix044

Archie, E.A., Theis, K.R., 2011. Animal behaviour meets microbial ecology. Anim. Behav. 82, 425–436. https://doi.org/10.1016/j.anbehav.2011.05.029

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Doré, J., MetaHIT Consortium (additional Members), Weissenbach, J., Ehrlich, S.D., Bork, P., 2011. Enterotypes of the human gut microbiome. Nature 473, 174–180. https://doi.org/10.1038/nature09944

Atlas, R., M., Bartha, R., 1998. Microbial Ecology: Fundamentals and Applications, Foutrh. ed.

Bailey, M.T., Coe, C.L., 1999. Maternal separation disrupts the integrity of the intestinal microflora in infant rhesus monkeys. Dev. Psychobiol. 35, 146–155.

Banks, J.C., Cary, S.C., Hogg, I.D., 2009. The phylogeography of Adelie penguin faecal flora. Environ. Microbiol. 11, 577–588. https://doi.org/10.1111/j.1462-2920.2008.01816.x

Birkhead, T.R., Martinez, J.G., Burke, T., Froman, D.P., 1999. Sperm mobility determines the outcome of sperm competition in the domestic fowl. Proc. R. Soc. Lond. B Biol. Sci. 266, 1759–1764. https://doi.org/10.1098/rspb.1999.0843

Bolnick, D.I., Snowberg, L.K., Caporaso, J.G., Lauber, C., Knight, R., Stutz, W.E., 2014. Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. Mol. Ecol. 23, 4831–4845. https://doi.org/10.1111/mec.12846

Borgdorff, H., Tsivtsivadze, E., Verhelst, R., Marzorati, M., Jurriaans, S., Ndayisaba, G.F., Schuren, F.H., van de Wijgert, J.H., 2014. Lactobacillus-dominated cervicovaginal microbiota associated with reduced HIV/STI prevalence and genital HIV viral load in African women. ISME J. 8, 1781–1793. https://doi.org/10.1038/ismej.2014.26

Borovkova, N., Korrovits, P., Ausmees, K., Türk, S., Jõers, K., Punab, M., Mändar, R., 2011. Influence of sexual intercourse on genital tract microbiota in infertile couples. Anaerobe, Cruising in the Amazing World of Microbial Ecosystems 17, 414–418. https://doi.org/10.1016/j.anaerobe.2011.04.015

Bresciani, C., Cabassi, C.S., Morini, G., Taddei, S., Bettini, R., Bigliardi, E., Ianni, F.D., Sabbioni, A., Parmigiani, E., 2014. Boar Semen Bacterial Contamination in Italy and Antibiotic Efficacy in a Modified Extender. Ital. J. Anim. Sci. 13, 3082. https://doi.org/10.4081/ijas.2014.3082

Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D., Lawley, T.D., 2016. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533, 543–546. https://doi.org/10.1038/nature17645

Bystrykh, L.V., 2012. Generalized DNA Barcode Design Based on Hamming Codes. PLoS ONE 7(5): e36852. https://doi.org/10.371/journal.pone.0036852

Buhr, R.J., Cox, N.A., Stern, N.J., Musgrove, M.T., Wilson, J.L., Hiett, K.L., 2002. Recovery of Campylobacter from Segments of the Reproductive Tract of Broiler Breeder Hens. Avian Dis. 46, 919–924.

Burrows, W.H., Quinn, J.P., 1937. The Collection of Spermatozoa from the Domestic Fowl and Turkey. Poult. Sci. 16, 19–24. https://doi.org/10.3382/ps.0160019

Camarda, A., Newell, D.G., Nasti, R., Di Modugno, G., 2000. Genotyping Campylobacter jejuni Strains Isolated from the Gut and Oviduct of Laying Hens. Avian Dis. 44, 907–912. https://doi.org/10.2307/1593065

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Carabotti, M., Scirocco, A., Maselli, M.A., Severi, C., 2015. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. Ann. Gastroenterol. Q. Publ. Hell. Soc. Gastroenterol. 28, 203–209.

Castro, J., Henriques, A., Machado, A., Henriques, M., Jefferson, K.K., Cerca, N., 2013. Reciprocal interference between Lactobacillus spp. and Gardnerella vaginalis on initial adherence to epithelial cells. Int. J. Med. Sci. 10, 1193–1198. https://doi.org/10.7150/ijms.6304

Cauwerts, K., Pasmans, F., Devriese, L.A., Martel, A., Haesebrouck, F., Decostere, A., 2006. Cloacal Lactobacillus isolates from broilers show high prevalence of resistance towards macrolide and lincosamide antibiotics. Avian Pathol. 35, 160–164. https://doi.org/10.1080/03079450600598137

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E.A., Grange, T., Geigl, E.-M., 2010. An Efficient Multistrategy DNA Decontamination Procedure of PCR Reagents for Hypersensitive PCR Applications. PLoS ONE 5, e13042. https://doi.org/10.1371/journal.pone.0013042

Chen, C., Song, X., Wei, W., Zhong, H., Dai, J., Lan, Z., Li, F., Yu, X., Feng, Q., Wang, Z., Xie, H., Chen, X., Zeng, C., Wen, B., Zeng, L., Du, H., Tang, H., Xu, C., Xia, Y., Xia, H., Yang, H., Wang, J., Wang, J., Madsen, L., Brix, S., Kristiansen, K., Xu, X., Li, J., Wu, R., Jia, H., 2017. The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. Nat. Commun. 8, 875. https://doi.org/10.1038/s41467-017-00901-0

Chen, Y., Blaser, M.J., 2007. Inverse associations of Helicobacter pylori with asthma and allergy. Arch. Intern. Med. 167, 821–827. https://doi.org/10.1001/archinte.167.8.821

Cho, I., Blaser, M.J., 2012. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13, 260–270. https://doi.org/10.1038/nrg3182

Chung, H., Pamp, S.J., Hill, J.A., Surana, N.K., Edelman, S.M., Troy, E.B., Reading, N.C., Villablanca, E.J., Wang, S., Mora, J.R., Umesaki, Y., Mathis, D., Benoist, C., Relman, D.A., Kasper, D.L., 2012. Gut immune maturation depends on colonization with a host-specific microbiota. Cell 149, 1578–1593. https://doi.org/10.1016/j.cell.2012.04.037

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., Harris, H.M.B., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., Fitzgerald, G.F., Deane, J., O'Connor, M., Harnedy, N., O'Connor, K., O'Mahony, D., Sinderen, D. van, Wallace, M., Brennan, L., Stanton, C., Marchesi, J.R., Fitzgerald, A.P., Shanahan, F., Hill, C., Ross, R.P., O'Toole, P.W., 2012. Gut microbiota composition correlates with diet and health in the elderly. Nature 488, 178. https://doi.org/10.1038/nature11319

Collet, J., Richardson, D.S., Worley, K., Pizzari, T., 2012. Sexual selection and the differential effect of polyandry. Proc. Natl. Acad. Sci. U. S. A. 109, 8641–8645. https://doi.org/10.1073/pnas.1200219109

Collias, N.E., Collias, E.C., 1996. Social organization of a red junglefowl,Gallus gallus, population related to evolution theory. Anim. Behav. 51, 1337–1354. https://doi.org/10.1006/anbe.1996.0137

Consortium, T.H.M.P., Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., Giglio, M.G., Hallsworth-Pepin, K., Lobos, E.A., Madupu, R., Magrini, V., Martin, J.C., Mitreva, M., Muzny, D.M., Sodergren, E.J., Versalovic, J., Wollam, A.M., Worley, K.C., Wortman, J.R., Young, S.K., Zeng, Q., Aagaard, K.M., Abolude, O.O., Allen-Vercoe, E., Alm, E.J., Alvarado, L., Andersen, G.L., Anderson, S., Appelbaum, E., Arachchi, H.M., Armitage, G., Arze, C.A., Ayvaz, T., Baker, C.C., Begg, L., Belachew, T., Bhonagiri, V., Bihan, M., Blaser, M.J., Bloom, T., Bonazzi, V., Brooks, J.P., Buck, G.A., Buhay, C.J., Busam, D.A., Campbell, J.L., Canon, S.R., Cantarel, B.L., Chain, P.S.G., Chen, I.-M.A., Chen, L., Chhibba, S., Chu, K., Ciulla, D.M., Clemente, J.C., Clifton, S.W., Conlan, S., Crabtree, J., Cutting, M.A., Davidovics, N.J., Davis, C.C., DeSantis, T.Z., Deal, C., Delehaunty, K.D., Dewhirst, F.E., Deych, E., Ding, Y., Dooling, D.J., Dugan, S.P., Dunne, W.M., Durkin, A.S., Edgar, R.C., Erlich, R.L., Farmer, C.N., Farrell, R.M., Faust, K., Feldgarden, M., Felix, V.M., Fisher, S., Fodor, A.A., Forney, L.J., Foster, L., Francesco, V.D., Friedman, J., Friedrich, D.C., Fronick, C.C., Fulton, L.L., Gao, H., Garcia, N., Giannoukos, G., Giblin, C., Giovanni, M.Y., Goldberg, J.M., Goll, J., Gonzalez, A., Griggs, A., Gujja, S., Haake, S.K., Haas, B.J., Hamilton, H.A., Harris, E.L., Hepburn, T.A., Herter, B., Hoffmann, D.E., Holder, M.E., Howarth, C., Huang, K.H., Huse, S.M., Izard, J., Jansson, J.K., Jiang, H., Jordan, C., Joshi, V., Katancik, J.A., Keitel, W.A., Kelley, S.T., Kells, C., King, N.B., Knights, D., Kong, H.H., Koren,

O., Koren, S., Kota, K.C., Kovar, C.L., Kyrpides, N.C., Rosa, P.S.L., Lee, S.L., Lemon, K.P., Lennon, N., Lewis, C.M., Lewis, L., Ley, R.E., Li, K., Liolios, K., Liu, B., Liu, Y., Lo, C.-C., Lozupone, C.A., Lunsford, R.D., Madden, T., Mahurkar, A.A., Mannon, P.J., Mardis, E.R., Markowitz, V.M., Mavromatis, K., McCorrison, J.M., McDonald, D., McEwen, J., McGuire, A.L., McInnes, P., Mehta, T., Mihindukulasuriya, K.A., Miller, J.R., Minx, P.J., Newsham, I., Nusbaum, C., O'Laughlin, M., Orvis, J., Pagani, I., Palaniappan, K., Patel, S.M., Pearson, M., Peterson, J., Podar, M., Pohl, C., Pollard, K.S., Pop, M., Priest, M.E., Proctor, L.M., Qin, X., Raes, J., Ravel, J., Reid, J.G., Rho, M., Rhodes, R., Riehle, K.P., Rivera, M.C., Rodriguez-Mueller, B., Rogers, Y.-H., Ross, M.C., Russ, C., Sanka, R.K., Sankar, P., Sathirapongsasuti, J.F., Schloss, J.A., Schloss, P.D., Schmidt, T.M., Scholz, M., Schriml, L., Schubert, A.M., Segata, N., Segre, J.A., Shannon, W.D., Sharp, R.R., Sharpton, T.J., Shenoy, N., Sheth, N.U., Simone, G.A., Singh, I., Smillie, C.S., Sobel, J.D., Sommer, D.D., Spicer, P., Sutton, G.G., Sykes, S.M., Tabbaa, D.G., Thiagarajan, M., Tomlinson, C.M., Torralba, M., Treangen, T.J., Truty, R.M., Vishnivetskaya, T.A., Walker, J., Wang, L., Wang, Z., Ward, D.V., Warren, W., Watson, M.A., Wellington, C., Wetterstrand, K.A., White, J.R., Wilczek-Boney, K., Wu, Y., Wylie, K.M., Wylie, T., Yandava, C., Ye, L., Ye, Y., Yooseph, S., Youmans, B.P., Zhang, L., Zhou, Y., Zhu, Y., Zoloth, L., Zucker, J.D., Birren, B.W., Gibbs, R.A., Highlander, S.K., Methé, B.A., Nelson, K.E., Petrosino, J.F., Weinstock, G.M., Wilson, R.K., White, O., 2012. Structure, function and diversity of the healthy human microbiome. Nature 486, 207. https://doi.org/10.1038/nature11234

Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J.M., Relman, D.A., 2012. The application of ecological theory toward an understanding of the human microbiome. Science 336, 1255–1262. https://doi.org/10.1126/science.1224203

Cox, N.A., Stern, N.J., Hiett, K.L., Berrang, M.E., 2002. Identification of a New Source of Campylobacter Contamination in Poultry: Transmission from Breeder Hens to Broiler Chickens. Avian Dis. 46, 535–541.

Cramer, E.R.A., Laskemoen, T., Eroukhmanoff, F., Haas, F., Hermansen, J.S., Lifjeld, J.T., Rowe, M., Sætre, G.-P., Johnsen, A., 2014. Testing a post-copulatory pre-zygotic reproductive barrier in a passerine species pair. Behav. Ecol. Sociobiol. 68, 1133–1144. https://doi.org/10.1007/s00265-014-1724-9

Crhanova, M., Hradecka, H., Faldynova, M., Matulova, M., Havlickova, H., Sisak, F., Rychlik, I., 2011. Immune response of chicken gut to natural colonization by gut microflora and to salmonella enterica serovar enteritidis infection. Infect. Immun. 79, 2755–2763. https://doi.org/10.1128/IAI.01375-10

Dawkins, R., 1989. The Extended Phenotype. Oxford: Oxford University Press. p. xiii. ISBN 0-19-288051-9.

De Palma, G., Blennerhassett, P., Lu, J., Deng, Y., Park, A.J., Green, W., Denou, E., Silva, M.A., Santacruz, A., Sanz, Y., Surette, M.G., Verdu, E.F., Collins, S.M., Bercik, P., 2015. Microbiota and host determinants of behavioural phenotype in maternally separated mice. Nat. Commun. 6, 7735. https://doi.org/10.1038/ncomms8735

Deriu, E., Boxx, G.M., He, X., Pan, C., Benavidez, S.D., Cen, L., Rozengurt, N., Shi, W., Cheng, G., 2016. Influenza Virus Affects Intestinal Microbiota and Secondary Salmonella Infection in the Gut through Type I Interferons. PLOS Pathog. 12, e1005572. https://doi.org/10.1371/journal.ppat.1005572

Dethlefsen, L., Huse, S., Sogin, M.L., Relman, D.A., 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16s rRNA sequencing. PLoS Biol 6, e280. https://doi.org/10.1371/journal.pbio.0060280

Dethlefsen, L., Relman, D.A., 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc. Natl. Acad. Sci. 108, 4554–4561. https://doi.org/10.1073/pnas.1000087107

Dikow, R.B., 2011. Genome-level homology and phylogeny of Shewanella (Gammaproteobacteria: lteromonadales: Shewanellaceae). BMC Genomics 12, 237. https://doi.org/10.1186/1471-2164-12-237

Dillon, R.J., Vennard, C.T., Charnley, A.K., 2000. Pheromones: Exploitation of gut bacteria in the locust. Nature 403, 851. https://doi.org/10.1038/35002669

Donoghue, A.M., Blore, P.J., Cole, K., Loskutoff, N.M., Donoghue, D.J., 2004. Detection of Campylobacter or Salmonella in Turkey Semen and the Ability of Poultry Semen Extenders to Reduce Their Concentrations. Poult. Sci. 83, 1728–1733. https://doi.org/10.1093/ps/83.10.1728

Duncan, S.H., Louis, P., Flint, H.J., 2004. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Appl. Environ. Microbiol. 70, 5810–5817. https://doi.org/10.1128/AEM.70.10.5810-5817.2004

Dunne, E.F., Chapin, J.B., Rietmeijer, C.A., Kent, C.K., Ellen, J.M., Gaydos, C.A., Willard, N.J., Kohn, R., Lloyd, L., Thomas, S., Birkjukow, N., Chung, S., Klausner, J., Schillinger, J.A., Markowitz, L.E., 2008. Rate and Predictors of Repeat Chlamydia trachomatis Infection Among Men. Sex. Transm. Dis. 35, S40. https://doi.org/10.1097/OLQ.0b013e31817247b2

Eberhard, W.G., 1996. Female Control: Sexual Selection by Cryptic Female Choice, Princeton University Press, Princeton

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., Relman, D.A., 2005. Diversity of the Human Intestinal Microbial Flora. Science 308, 1635–1638. https://doi.org/10.1126/science.1110591

Eckert, L.O., Moore, D.E., Patton, D.L., Agnew, K.J., Eschenbach, D.A., 2003. Relationship of Vaginal Bacteria and Inflammation With Conception and Early Pregnancy Loss Following In-Vitro Fertilization. Infect. Dis. Obstet. Gynecol. 11, 11–17. https://doi.org/10.1155/S1064744903000024

Edens, F.W., Parkhurst, C.R., Casas, I.A., Dobrogosz, W.J., 1997. Principles of ex ovo competitive exclusion and in ovo administration of Lactobacillus reuteri. Poult. Sci. 76, 179–196. https://doi.org/10.1093/ps/76.1.179

Eggert-Kruse, W., Rohr, G., Probst, S., Rusu, R., Hund, M., Demirakca, T., Aufenanger, J., Runnebaum, B., Petzoldt, D., 1998. Antisperm antibodies and microorganisms in genital secretions—a clinically significant relationship? Andrologia 30, 61–71. https://doi.org/10.1111/j.1439-0272.1998.tb02828.x

Eschenbach, D.A., Patton, D.L., Hooton, T.M., Meier, A.S., Stapleton, A., Aura, J., Agnew, K., 2001. Effects of vaginal intercourse with and without a condom on vaginal flora and vaginal epithelium. J. Infect. Dis. 183, 913–918. https://doi.org/10.1086/319251

Etches, R.J., 1996. Reproduction in Poultry. CAB International.

Faith, D.P., Minchin, P.R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69, 57–68. https://doi.org/10.1007/BF00038687

Filippo, C.D., Cavalieri, D., Paola, M.D., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., Lionetti, P., 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. 107, 14691–14696. https://doi.org/10.1073/pnas.1005963107

Firman, R.C., Gasparini, C., Manier, M.K., Pizzari, T., 2017. Postmating Female Control: 20 Years of Cryptic Female Choice. Trends Ecol. Evol. 32, 368–382. https://doi.org/10.1016/j.tree.2017.02.010

Forman, D., Newell, D.G., Fullerton, F., Yarnell, J.W., Stacey, A.R., Wald, N., Sitas, F., 1991. Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation. BMJ 302, 1302–1305.

Foster, J.A., Neufeld, K.-A.M., 2013. Gut–brain axis: how the microbiome influences anxiety and depression. Trends Neurosci. 36, 305–312. https://doi.org/10.1016/j.tins.2013.01.005

Franzenburg, S., Walter, J., Künzel, S., Wang, J., Baines, J.F., Bosch, T.C.G., Fraune, S., 2013. Distinct antimicrobial peptide expression determines host species-specific bacterial associations. Proc. Natl. Acad. Sci. U. S. A. 110, E3730–E3738. https://doi.org/10.1073/pnas.1304960110

Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S., Kondo, N., 1994. One subspecies of the red junglefowl (Gallus gallus gallus) suffices as the matriarchic ancestor of all domestic breeds. Proc. Natl. Acad. Sci. U. S. A. 91, 12505–12509.

Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., Kondo, N., Ohno, S., 1996. Monophyletic origin and unique dispersal patterns of domestic fowls. Proc. Natl. Acad. Sci. U. S. A. 93, 6792–6795.

Gajer, P., Brotman, R.M., Bai, G., Sakamoto, J., Schütte, U.M.E., Zhong, X., Koenig, S.S.K., Fu, L., Ma, Z. (Sam), Zhou, X., Abdo, Z., Forney, L.J., Ravel, J., 2012. Temporal dynamics of the human vaginal microbiota. Sci. Transl. Med. 4, 132ra52-132ra52. https://doi.org/10.1126/scitranslmed.3003605

Galaverni, M., Caniglia, R., Milanesi, P., Lapalombella, S., Fabbri, E., Randi, E., 2016. Choosy Wolves? Heterozygote Advantage But No Evidence of MHC-Based Disassortative Mating. J. Hered. 107, 134–142. https://doi.org/10.1093/jhered/esv090

Gallo, M.F., Warner, L., King, C.C., Sobel, J.D., Klein, R.S., Cu-Uvin, S., Rompalo, A.M., Jamieson, D.J., 2011. Association between Semen Exposure and Incident Bacterial Vaginosis. Infect. Dis. Obstet. Gynecol. 2011, e842652. https://doi.org/10.1155/2011/842652

Gillingham, M.A.F., Richardson, D.S., Løvlie, H., Moynihan, A., Worley, K., Pizzari, T., 2009. Cryptic preference for MHC-dissimilar females in male red junglefowl, Gallus gallus. Proc. R. Soc. B Biol. Sci. 276, 1083–1092. https://doi.org/10.1098/rspb.2008.1549

Graves, B.M., Duvall, D., 1995. Effects of sexually transmitted diseases on heritable variation in sexually selected systems. Anim. Behav. 50, 1129–1131. https://doi.org/10.1016/0003-3472(95)80112-X

Spear, G.T., Kersh, E., Guenthner, P., Vishwanathan, S.A., Gilbert, D., Zariffard, M.R., Mirmonsef, P., Landay, A., Zheng, L., Gillevet, P., 2012. Longitudinal assessment of pigtailed macaque lower genital tract microbiota by pyrosequencing reveals dissimilarity to the genital

microbiota of healthy humans. AIDS Res. Hum. Retroviruses 28, 1244–1249. https://doi.org/10.1089/aid.2011.0382

Haines, M.D., Parker, H.M., McDaniel, C.D., Kiess, A.S., 2013. Impact of 6 different intestinal bacteria on broiler breeder sperm motility in vitro. Poult. Sci. 92, 2174–2181. https://doi.org/10.3382/ps.2013-03109

Hashemi, F.B., Ghassemi, M., Roebuck, K.A., Spear, G.T., 1999. Activation of human immunodeficiency virus type 1 expression by Gardnerella vaginalis. J. Infect. Dis. 179, 924–930. https://doi.org/10.1086/314674

Hashway, S.A., Bergin, I.L., Bassis, C.M., Uchihashi, M., Schmidt, K.C., Young, V.B., Aronoff, D.M., Patton, D.L., Bell, J.D., 2014. Impact of a Hormone-Releasing Intrauterine System on the Vaginal Microbiome: A Prospective Baboon Model. J. Med. Primatol. 43, 89–99. https://doi.org/10.1111/jmp.12090

Hiett, K.L., Cox, N.A., Buhr, R.J., Stern, N.J., 2002. Genotype Analyses of *Campylobacter* Isolated from Distinct Segments of the Reproductive Tracts of Broiler Breeder Hens. Curr. Microbiol. 45, 0400–0404. https://doi.org/10.1007/s00284-002-3771-0

Hong, P.-Y., Lee, B.W., Aw, M., Shek, L.P.C., Yap, G.C., Chua, K.Y., Liu, W.-T., 2010. Comparative Analysis of Fecal Microbiota in Infants with and without Eczema. PLOS ONE 5, e9964. https://doi.org/10.1371/journal.pone.0009964

Hou, D., Zhou, X., Zhong, X., Settles, M.L., Herring, J., Wang, L., Abdo, Z., Forney, L.J., Xu, C., 2013. Microbiota of the seminal fluid from healthy and infertile men. Fertil. Steril. 100, 1261–1269.e3. https://doi.org/10.1016/j.fertnstert.2013.07.1991

Hupton, G., Portocarrero, S., Newman, M., Westneat, D.F., 2003. Bacteria in the reproductive tracts of red-winged blackbirds. The Condor 105, 453–464. https://doi.org/10.1650/7246

Ivanov, I.B., Kuzmin, M.D., Gritsenko, V.A., 2009. Microflora of the seminal fluid of healthy men and men suffering from chronic prostatitis syndrome. Int. J. Androl. 32, 462–467. https://doi.org/10.1111/j.1365-2605.2008.00878.x

Jakobsson, H.E., Abrahamsson, T.R., Jenmalm, M.C., Harris, K., Quince, C., Jernberg, C., Björkstén, B., Engstrand, L., Andersson, A.F., 2014. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by Caesarean section. Gut 63, 559–566. https://doi.org/10.1136/gutjnl-2012-303249

Jakobsson, H.E., Jernberg, C., Andersson, A.F., Sjölund-Karlsson, M., Jansson, J.K., Engstrand, L., 2010. Short-Term Antibiotic Treatment Has Differing Long-Term Impacts on the Human Throat and Gut Microbiome. PLOS ONE 5, e9836. https://doi.org/10.1371/journal.pone.0009836

Javurek, A.B., Spollen, W.G., Ali, A.M.M., Johnson, S.A., Lubahn, D.B., Bivens, N.J., Bromert, K.H., Ellersieck, M.R., Givan, S.A., Rosenfeld, C.S., 2016. Discovery of a Novel Seminal Fluid Microbiome and Influence of Estrogen Receptor Alpha Genetic Status. Sci. Rep. 6. https://doi.org/10.1038/srep23027

Jensen, R.G., Smith, K.L., Edmondson, J.E., Merilan, C.P., 1956. THE CHARACTERISTICS OF SOME RUMEN LACTOBACILLI1. J. Bacteriol. 72, 253–258.

Johnston, S.D., Lever, J., McLeod, R., Oishi, M., Qualischefski, E., Omanga, C., Leitner, M., Price, R., Barker, L., Kamau, K., Gaughan, J., D'Occhio, M., 2014. Semen collection and seminal

characteristics of the Australian saltwater crocodile (*Crocodylus porosus*). Aquaculture 422, 25–35. https://doi.org/10.1016/j.aquaculture.2013.11.002

Juricova, H., Videnska, P., Lukac, M., Faldynova, M., Babak, V., Havlickova, H., Sisak, F., Rychlik, I., 2013. Influence of Salmonella enterica serovar enteritidis infection on the development of the cecum microbiota in newly hatched chicks. Appl. Environ. Microbiol. 79, 745–747. https://doi.org/10.1128/AEM.02628-12

Kaufman, J., Milne, S., Göbel, T.W.F., Walker, B.A., Jacob, J.P., Auffray, C., Zoorob, R., Beck, S., 1999. The chicken B locus is a minimal essential major histocompatibility complex. Nature 401, 923–925. https://doi.org/10.1038/44856

Keiser, C.N., Pinter-Wollman, N., Ziemba, M.J., Kothamasu, K.S., Pruitt, J.N., 2017. The index case is not enough: Variation among individuals, groups and social networks modify bacterial transmission dynamics. J. Anim. Ecol. n/a-n/a. https://doi.org/10.1111/1365-2656.12729

Kermes, K., Punab, M., Lõivukene, K., Mändar, R., 2003. Anaerobic seminal fluid micro-flora in chronic prostatitis/chronic pelvic pain syndrome patients. Anaerobe 9, 117–123. https://doi.org/10.1016/S1075-9964(03)00085-4

Kokko, H., Ranta, E., Ruxton, G., Lundberg, P., 2002. Sexually transmitted disease and the evolution of mating systems. Evolution 56, 1091–1100. https://doi.org/10.1111/j.0014-3820.2002.tb01423.x

Koren, O., Goodrich, J.K., Cullender, T.C., Spor, A., Laitinen, K., Kling Bäckhed, H., Gonzalez, A., Werner, J.J., Angenent, L.T., Knight, R., Bäckhed, F., Isolauri, E., Salminen, S., Ley, R.E., 2012. Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell 150, 470–480. https://doi.org/10.1016/j.cell.2012.07.008

Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., Huttenhower, C., Ley, R.E., 2013. A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. PLOS Comput Biol 9, e1002863. https://doi.org/10.1371/journal.pcbi.1002863

Korhonen, C.J., Srinivasan, S., Huang, D., Ko, D.L., Sanders, E.J., Peshu, N.M., Krieger, J.N., Muller, C.H., Coombs, R.W., Fredricks, D.N., Graham, S.M., 2017. Semen Bacterial Concentrations and HIV-1 RNA Shedding Among HIV-1–Seropositive Kenyan Men. J. Acquir. Immune Defic. Syndr. 1999 74, 250–257. https://doi.org/10.1097/QAI.0000000000001244

Kreisinger, J., Čížková, D., Kropáčková, L., Albrecht, T., 2015. Cloacal Microbiome Structure in a Long-Distance Migratory Bird Assessed Using Deep 16sRNA Pyrosequencing. PLoS ONE 10, e0137401. https://doi.org/10.1371/journal.pone.0137401

Kulkarni, S., Heeb, P., 2007. Social and sexual behaviours aid transmission of bacteria in birds. Behav. Processes 74, 88–92. https://doi.org/10.1016/j.beproc.2006.10.005

Lake, P.E., 1957. The male reproductive tract of the fowl. J. Anat. 91, 116–129.

Leclaire, S., Jacob, S., Greene, L.K., Dubay, G.R., Drea, C.M., 2017. Social odours covary with bacterial community in the anal secretions of wild meerkats. Sci. Rep. 7, 3240. https://doi.org/10.1038/s41598-017-03356-x

Leclaire, S., Nielsen, J.F., Drea, C.M., 2014. Bacterial communities in meerkat anal scent secretions vary with host sex, age, and group membership. Behav. Ecol. aru074. https://doi.org/10.1093/beheco/aru074

Leclaire, S., Nielsen, J.F., Thavarajah, N.K., Manser, M., Clutton-Brock, T.H., 2013. Odour-based kin discrimination in the cooperatively breeding meerkat. Biol. Lett. 9, 20121054. https://doi.org/10.1098/rsbl.2012.1054

Ley, R.E., Turnbaugh, P.J., Klein, S., Gordon, J.I., 2006. Microbial ecology: Human gut microbes associated with obesity. Nature 444, 1022–1023. https://doi.org/10.1038/4441022a

Li, X.J., Piao, X.S., Kim, S.W., Liu, P., Wang, L., Shen, Y.B., Jung, S.C., Lee, H.S., 2007. Effects of Chito-Oligosaccharide Supplementation on Performance, Nutrient Digestibility, and Serum Composition in Broiler Chickens. Poult. Sci. 86, 1107–1114. https://doi.org/10.1093/ps/86.6.1107

Liu, C.M., Osborne, B.J.W., Hungate, B.A., Shahabi, K., Huibner, S., Lester, R., Dwan, M.G., Kovacs, C., Contente-Cuomo, T.L., Benko, E., Aziz, M., Price, L.B., Kaul, R., 2014. The Semen Microbiome and Its Relationship with Local Immunology and Viral Load in HIV Infection. PLoS Pathog 10, e1004262. https://doi.org/10.1371/journal.ppat.1004262

Lizé, A., McKay, R., Lewis, Z., 2014. Kin recognition in *Drosophila*: the importance of ecology and gut microbiota. ISME J. 8, 469. https://doi.org/10.1038/ismej.2013.157

Lombardo, M.P., Thorpe, P.A., 2000. Microbes in tree swallow semen. J. Wildl. Dis. 36, 460–468. https://doi.org/10.7589/0090-3558-36.3.460

Lombardo, M.P., Thorpe, P.A., Cichewicz, R., Henshaw, M., Millard, C., Steen, C., Zeller, T.K., 1996. Communities of Cloacal Bacteria in Tree Swallow Families. The Condor 98, 167–172. https://doi.org/10.2307/1369521

Lombardo, M.P., Thorpe, P.A., Power, H.W., 1999. The beneficial sexually transmitted microbe hypothesis of avian copulation. Behav. Ecol. 10, 333–337. https://doi.org/10.1093/beheco/10.3.333

Lorenzen, E., Kudirkiene, E., Gutman, N., Grossi, A.B., Agerholm, J.S., Erneholm, K., Skytte, C., Dalgaard, M.D., Bojesen, A.M., 2015. The vaginal microbiome is stable in prepubertal and sexually mature Ellegaard Göttingen Minipigs throughout an estrous cycle. Vet. Res. 46. https://doi.org/10.1186/s13567-015-0274-0

Løvlie, H., Gillingham, M.A.F., Worley, K., Pizzari, T., Richardson, D.S., 2013. Cryptic female choice favours sperm from major histocompatibility complex-dissimilar males. Proc R Soc B 280, 20131296. https://doi.org/10.1098/rspb.2013.1296

Lozupone, C.A., Li, M., Campbell, T.B., Flores, S.C., Linderman, D., Gebert, M.J., Knight, R., Fontenot, A.P., Palmer, B.E., 2013. Alterations in the Gut Microbiota Associated with HIV-1 Infection. Cell Host Microbe 14, 329–339. https://doi.org/10.1016/j.chom.2013.08.006

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., Knight, R., 2012. Diversity, stability and resilience of the human gut microbiota. Nature 489, 220–230. https://doi.org/10.1038/nature11550

Lung, O., Kuo, L., Wolfner, M.F., 2001. Drosophila males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. J. Insect Physiol. 47, 617–622. https://doi.org/10.1016/S0022-1910(00)00151-7

Lusk, R.W., 2014. Diverse and Widespread Contamination Evident in the Unmapped Depths of High Throughput Sequencing Data. PLoS ONE 9, e110808. https://doi.org/10.1371/journal.pone.0110808

MacManes, M.D., 2011. Promiscuity in mice is associated with increased vaginal bacterial diversity. Naturwissenschaften 98, 951–960. https://doi.org/10.1007/s00114-011-0848-2

Mändar, R., Punab, M., Borovkova, N., Lapp, E., Kiiker, R., Korrovits, P., Metspalu, A., Krjutškov, K., Nõlvak, H., Preem, J.-K., Oopkaup, K., Salumets, A., Truu, J., 2015. Complementary seminovaginal microbiome in couples. Res. Microbiol. 166, 440–447. https://doi.org/10.1016/j.resmic.2015.03.009

Maroto Martín, L.O., Muñoz, E.C., De Cupere, F., Van Driessche, E., Echemendia-Blanco, D., Rodríguez, J.M.M., Beeckmans, S., 2010. Bacterial contamination of boar semen affects the litter size. Anim. Reprod. Sci. 120, 95–104. https://doi.org/10.1016/j.anireprosci.2010.03.008

Martin, M.O., Gilman, F.R., Weiss, S.L., 2010. Sex-specific asymmetry within the cloacal microbiota of the striped plateau lizard, Sceloporus virgatus. Symbiosis 51, 97–105. https://doi.org/10.1007/s13199-010-0078-y

McDonald, G.C., Pizzari, T., 2015. Mating behaviour: Sexual networks and sexual selection, in: Animal Social Networks. Oxford University Press, pp. 24–37.

McDonald, G.C., Spurgin, L.G., Fairfield, E.A., Richardson, D.S., Pizzari, T., 2017. Pre- and postcopulatory sexual selection favor aggressive, young males in polyandrous groups of red junglefowl. Evolution 71, 1653–1669. https://doi.org/10.1111/evo.13242

Milinski, M., Griffiths, S., Wegner, K.M., Reusch, T.B.H., Haas-Assenbaum, A., Boehm, T., 2005. Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proc. Natl. Acad. Sci. U. S. A. 102, 4414–4418. https://doi.org/10.1073/pnas.0408264102

Miller, E.A., Livermore, J.A., Alberts, S.C., Tung, J., Archie, E.A., 2017. Ovarian cycling and reproductive state shape the vaginal microbiota in wild baboons. Microbiome 5, 8. https://doi.org/10.1186/s40168-017-0228-z

Mimee, M., Citorik, R.J., Lu, T.K., 2016. Microbiome therapeutics — Advances and challenges. Adv. Drug Deliv. Rev., Synthetic Biology: Innovative approaches for pharmaceutics and drug delivery 105, Part A, 44–54. https://doi.org/10.1016/j.addr.2016.04.032

Mitchell, C.M., Fredricks, D.N., Winer, R.L., Koutsky, L., 2012. Effect of Sexual Debut on Vaginal Microbiota in a Cohort of Young Women. Obstet. Gynecol. 120, 1306–1313.

Moeller, A.H., Li, Y., Mpoudi Ngole, E., Ahuka-Mundeke, S., Lonsdorf, E.V., Pusey, A.E., Peeters, M., Hahn, B.H., Ochman, H., 2014. Rapid changes in the gut microbiome during human evolution. Proc. Natl. Acad. Sci. U. S. A. 111, 16431–16435. https://doi.org/10.1073/pnas.1419136111

Mohan, J., Saini, M., Joshi, P., 1995. Isolation of a spermatozoa motility inhibiting factor from chicken seminal plasma with antibacterial property. Biochim. Biophys. Acta BBA - Gen. Subj. 1245, 407–413. https://doi.org/10.1016/0304-4165(95)00119-0

Moos, W.H., Faller, D.V., Harpp, D.N., Kanara, I., Pernokas, J., Powers, W.R., Steliou, K., 2016. Microbiota and Neurological Disorders: A Gut Feeling. BioResearch Open Access 5, 137–145. https://doi.org/10.1089/biores.2016.0010

Morris, E.K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T.S., Meiners, T., Müller, C., Obermaier, E., Prati, D., Socher, S.A., Sonnemann, I., Wäschke, N., Wubet, T., Wurst, S., Rillig, M.C., 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. Ecol. Evol. 4, 3514–3524. https://doi.org/10.1002/ece3.1155

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., Gordon, J.I., 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 332, 970–974. https://doi.org/10.1126/science.1198719

Nelson, D.B., Hanlon, A.L., Wu, G., Liu, C., Fredricks, D.N., 2015. First Trimester Levels of BV-Associated Bacteria and Risk of Miscarriage Among Women Early in Pregnancy. Matern. Child Health J. 19, 2682–2687. https://doi.org/10.1007/s10995-015-1790-2

Neuendorf, E., Gajer, P., Bowlin, A.K., Marques, P.X., Ma, B., Yang, H., Fu, L., Humphrys, M.S., Forney, L.J., Myers, G.S.A., Bavoil, P.M., Rank, R.G., Ravel, J., 2015. Chlamydia caviae infection alters abundance but not composition of the guinea pig vaginal microbiota. Pathog. Dis. 73. https://doi.org/10.1093/femspd/ftv019

Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] - Available from: https://www.ncbi.nlm.nih.gov/nucleotide/

Nishiyama, H., 1955. Studies on the accessory reproductive organs in the cock. J. Foe. Agr. Kyushu Univ., 10, 277-305.

Nishiyama, H., Fujishima, T., 1961. On the ejection of the accessory reproductive fluid of the cock during natural copulation. Mem. Fac. Agr. Kagoshima Univ., 4, 27-4

Omland, K.E., 1996. Female mallard mating preferences for multiple male ornaments. Behav. Ecol. Sociobiol. 39, 361–366. https://doi.org/10.1007/s002650050301

Otti, O., McTighe, A.P., Reinhardt, K., 2013. In vitro antimicrobial sperm protection by an ejaculate-like substance. Funct. Ecol. 27, 219–226. https://doi.org/10.1111/1365-2435.12025

Palma, G. de, Nova, E., Pozo Rubio, T., Sanz, Y., 2010. Interplay between human leukocyte antigen genes and the microbial colonization process of the newborn intestine. [WWW Document]. URL http://digital.csic.es/handle/10261/94237 (accessed 2.26.15).

Parker, G.A., 1970. Sperm Competition and Its Evolutionary Consequences in the Insects. Biol. Rev. 45, 525–567. https://doi.org/10.1111/j.1469-185X.1970.tb01176.x

Pellatt, E.J., Birkhead, T.R., 1994. Ejaculate size in Zebra Finches Taeniopygia guttata and a method for obtaining ejaculates from passerine birds. Ibis 136, 97–101. https://doi.org/10.1111/j.1474-919X.1994.tb08136.x

Pizzari, T., 2007. Post-insemination sexual selection in birds. Soc. Reprod. Fertil. Suppl. 65, 137–154.

Pizzari, T., Birkhead, T.R., 2000. Female feral fowl eject sperm of subdominant males. Nature 405, 787. https://doi.org/10.1038/35015558

Pizzari, T., Lø, H., Cornwallis, C.K., 2004. Sex–specific, counteracting responses to inbreeding in a bird. Proc. R. Soc. Lond. B Biol. Sci. 271, 2115–2121. https://doi.org/10.1098/rspb.2004.2843

Poiani, A., 2010. Do Cloacal Pathogenic Microbes Behave as Sexually Transmitted Parasites in Birds? Open Ornithol. J. 3.

Poiani, A., Gwozdz, J., 2002. Cloacal microorganisms and mating systems of four Australian bird species. Emu 102, 291–296. https://doi.org/10.1071/MU01055

Poiani, A., Wilks, C., 2000a. Sexually Transmitted Diseases: A Possible Cost of Promiscuity in Birds? The Auk 117, 1061–1065. https://doi.org/10.1642/0004-8038(2000)117[1061:STDAPC]2.0.CO;2

Poiani, A., Wilks, C., 2000b. Sexually Transmitted Diseases: A Possible Cost of Promiscuity in Birds? The Auk 117, 1061–1065. https://doi.org/10.1642/0004-8038(2000)117[1061:STDAPC]2.0.CO;2

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raulo, A., Ruokolainen, L., Lane, A., Amato, K., Knight, R., Leigh, S., Stumpf, R., White, B., Nelson, K.E., Baden, A.L., Tecot, S.R., 2017. Social behaviour and gut microbiota in red-bellied lemurs (Eulemur rubriventer): In search of the role of immunity in the evolution of sociality. J. Anim. Ecol. n/a-n/a. https://doi.org/10.1111/1365-2656.12781

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S.K., McCulle, S.L., Karlebach, S., Gorle, R., Russell, J., Tacket, C.O., Brotman, R.M., Davis, C.C., Ault, K., Peralta, L., Forney, L.J., 2011. Vaginal microbiome of reproductive-age women. Proc. Natl. Acad. Sci. 108, 4680–4687. https://doi.org/10.1073/pnas.1002611107

Rehewy, M.S., Hafez, E.S., Thomas, A., Brown, W.J., 1979. Aerobic and anaerobic bacterial flora in semen from fertile and infertile groups of men. Arch. Androl. 2, 263–268.

Reiber, M.A., Conner, D.E., 1995. Effect of Mating Activity on the Ability of Salmonella enteritidis to Persist in the Ovary and Oviduct of Chickens. Avian Dis. 39, 323–327. https://doi.org/10.2307/1591874

Reiber, M.A., Mcinroy, J.A., Conner, D.E., 1995. Enumeration and Identification of Bacteria in Chicken Semen. Poult. Sci. 74, 795–799. https://doi.org/10.3382/ps.0740795

Ren, T., Grieneisen, L.E., Alberts, S.C., Archie, E.A., Wu, M., 2015. Development, diet and dynamism: longitudinal and cross-sectional predictors of gut microbial communities in wild baboons. Environ. Microbiol. n/a-n/a. https://doi.org/10.1111/1462-2920.12852

Reusch, T.B.H., Häberli, M.A., Aeschlimann, P.B., Milinski, M., 2001. Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. Nature 414, 300. https://doi.org/10.1038/35104547

Ringel-Kulka, T., Cheng, J., Ringel, Y., Salojärvi, J., Carroll, I., Palva, A., Vos, W.M. de, Satokari, R., 2013. Intestinal Microbiota in Healthy U.S. Young Children and Adults—A High

Throughput Microarray Analysis. PLOS ONE 8, e64315. https://doi.org/10.1371/journal.pone.0064315

Rodin, D.M., Larone, D., Goldstein, M., 2003. Relationship between semen cultures, leukospermia, and semen analysis in men undergoing fertility evaluation. Fertil. Steril. 79, 1555–1558. https://doi.org/10.1016/S0015-0282(03)00340-6

Rose WA, McGowin CL, Spagnuolo RA, Eaves-Pyles TD, Popov VL, Pyles RB. (2012). Commensal bacteria modulate innate immune responses of vaginal epithelial cell multilayer cultures. PLoS One 7: e32728.

Rosenberg, E., Zilber-Rosenberg, I., 2016. Microbes Drive Evolution of Animals and Plants: the Hologenome Concept. mBio 7, e01395-15. https://doi.org/10.1128/mBio.01395-15

Round, J.L., Mazmanian, S.K., 2010. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. Proc. Natl. Acad. Sci. U. S. A. 107, 12204–12209. https://doi.org/10.1073/pnas.0909122107

Rowe, M., Czirják, G.Á., Lifjeld, J.T., Giraudeau, M., 2013. Lysozyme-associated bactericidal activity in the ejaculate of a wild passerine. Biol. J. Linn. Soc. 109, 92–100. https://doi.org/10.1111/bij.12044

Rowe, M., Czirják, G.Á., McGraw, K.J., Giraudeau, M., 2011. Sexual ornamentation reflects antibacterial activity of ejaculates in mallards. Biol. Lett. 7, 740–742. https://doi.org/10.1098/rsbl.2011.0276

Rowe, M., Pruett-Jones, S., 2013. Extra-pair paternity, sperm competition and their evolutionary consequences in the Maluridae. Emu - Austral Ornithol. 113, 218–231. https://doi.org/10.1071/MU12084

Ruiz-de-Castañeda, R., Vela, A.I., Lobato, E., Briones, V., Moreno, J., 2011. Prevalence of potentially pathogenic culturable bacteria on eggshells and in cloacae of female Pied Flycatchers in a temperate habitat in central Spain. J. Field Ornithol. 82, 215–224. https://doi.org/10.1111/j.1557-9263.2011.00324.x

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 12, 87. https://doi.org/10.1186/s12915-014-0087-z

Shalika, S., Dugan, K., Smith, R.D., Padilla, S.L., 1996. The effect of positive semen bacterial and Ureaplasma cultures on in-vitro fertilization success. Hum. Reprod. 11, 2789–2792. https://doi.org/10.1093/oxfordjournals.humrep.a019211

Shankar, V., Agans, R., Holmes, B., Raymer, M., Paliy, O., 2013. Do gut microbial communities differ in pediatric IBS and health? Gut Microbes 4, 347–352. https://doi.org/10.4161/gmic.24827

Sharon, G., Segal, D., Ringo, J.M., Hefetz, A., Zilber-Rosenberg, I., Rosenberg, E., 2010. Commensal bacteria play a role in mating preference of Drosophila melanogaster. Proc. Natl. Acad. Sci. 107, 20051–20056. https://doi.org/10.1073/pnas.1009906107

Sheldon, B.C., 1993. Sexually transmitted disease in birds: Occurrence and evolutionary significance. Philos. Trans. R. Soc. Lond. B Biol. Sci. 339, 491–497. https://doi.org/10.1098/rstb.1993.0044

Silk, M.J., Croft, D.P., Delahay, R.J., Hodgson, D.J., Boots, M., Weber, N., McDonald, R.A., 2017. Using Social Network Measures in Wildlife Disease Ecology, Epidemiology, and Management. Bioscience 67, 245–257. https://doi.org/10.1093/biosci/biw175

Sin, Y.W., Annavi, G., Newman, C., Buesching, C., Burke, T., Macdonald, D.W., Dugdale, H.L., 2015. MHC class II-assortative mate choice in European badgers (Meles meles). Mol. Ecol. 24, 3138–3150. https://doi.org/10.1111/mec.13217

Smith, A. U., 1949. The control of bacterial growth in fowl semen. J. Agric. Sci. 39:194-200. St. Louis, M. E., D. L. Morse, M. E. Potter

Sommer, F., Bäckhed, F., 2013. The gut microbiota — masters of host development and physiology. Nat. Rev. Microbiol. 11, 227–238. https://doi.org/10.1038/nrmicro2974

Spear, G.T., Gilbert, D., Sikaroodi, M., Doyle, L., Green, L., Gillevet, P.M., Landay, A.L., Veazey, R.S., 2010. Identification of Rhesus Macaque Genital Microbiota by 16S Pyrosequencing Shows Similarities to Human Bacterial Vaginosis: Implications for Use as an Animal Model for HIV Vaginal Infection. AIDS Res. Hum. Retroviruses 26, 193–200. https://doi.org/10.1089/aid.2009.0166

Stewart, R., Rambo, T.B., 2000. Cloacal microbes in house sparrows. The Condor 102, 679–684. https://doi.org/10.1650/0010-5422(2000)102[0679:CMIHS]2.0.CO;2

Stürup, M., Baer-Imhoof, B., Nash, D.R., Boomsma, J.J., Baer, B., 2013. When every sperm counts: factors affecting male fertility in the honeybee Apis mellifera. Behav. Ecol. 24, 1192–1198. https://doi.org/10.1093/beheco/art049

Swartz, J.D., Lachman, M., Westveer, K., O'Neill, T., Geary, T., Kott, R.W., Berardinelli, J.G., Hatfield, P.G., Thomson, J.M., Roberts, A., Yeoman, C.J., 2014. Characterization of the Vaginal Microbiota of Ewes and Cows Reveals a Unique Microbiota with Low Levels of Lactobacilli and Near-Neutral pH. Front. Vet. Sci. 1. https://doi.org/10.3389/fvets.2014.00019

Tan, C.K.W., Doyle, P., Bagshaw, E., Richardson, D.S., Wigby, S., Pizzari, T., 2017. The contrasting role of male relatedness in different mechanisms of sexual selection in red junglefowl. Evolution 71, 403–420. https://doi.org/10.1111/evo.13145

Tannock, G.W., Savage, D.C., 1974. Influences of dietary and environmental stress on microbial populations in the murine gastrointestinal tract. Infect. Immun. 9, 591–598.

Theis, K.R., Venkataraman, A., Dycus, J.A., Koonter, K.D., Schmitt-Matzen, E.N., Wagner, A.P., Holekamp, K.E., Schmidt, T.M., 2013. Symbiotic bacteria appear to mediate hyena social odors. Proc. Natl. Acad. Sci. 110, 19832–19837. https://doi.org/10.1073/pnas.1306477110

Thrall, P.H., Antonovics, J., Bever, J.B., 1997. Sexual transmission of disease and host mating systems: Within-season reproductive success. Am. Nat. 149, 485–506.

Thrall, P.H., Antonovics, J., Dobson, A.P., 2000. Sexually transmitted diseases in polygynous mating systems: prevalence and impact on reproductive success. Proc. R. Soc. Lond. B Biol. Sci. 267, 1555–1563. https://doi.org/10.1098/rspb.2000.1178

Toivanen, P., Vaahtovuo, J., Eerola, E., 2001. Influence of major histocompatibility complex on bacterial composition of fecal flora. Infect. Immun. 69, 2372–2377. https://doi.org/10.1128/IAI.69.4.2372-2377.2001

Toth, A., Lesser, M.L., 1981. Asymptomatic Bacteriospermia in Fertile and Infertile Men. Fertil. Steril. 36, 88–91. https://doi.org/10.1016/S0015-0282(16)45625-6

Triplett, M.D., Parker, H.M., McDaniel, C.D., Kiess, A.S., 2016. Influence of 6 different intestinal bacteria on Beltsville Small White turkey semen. Poult. Sci. 95, 1918–1926. https://doi.org/10.3382/ps/pew119

Tung, J., Barreiro, L.B., Burns, M.B., Grenier, J.-C., Lynch, J., Grieneisen, L.E., Altmann, J., Alberts, S.C., Blekhman, R., Archie, E.A., 2015. Social networks predict gut microbiome composition in wild baboons. eLife 4, e05224. https://doi.org/10.7554/eLife.05224

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C., Knight, R., Gordon, J.I., 2007. The human microbiome project: exploring the microbial part of ourselves in a changing world. Nature 449, 804–810. https://doi.org/10.1038/nature06244

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., Gordon, J.I., 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444, 1027–131. https://doi.org/10.1038/nature05414

Uchihashi, M., Bergin, I.L., Bassis, C.M., Hashway, S.A., Chai, D., Bell, J.D., 2015. Influence of Age, Reproductive Cycling Status, and Menstruation on the Vaginal Microbiome in Baboons (Papio anubis). Am. J. Primatol. 77, 563–578. https://doi.org/10.1002/ajp.22378

van Dongen, W.F., White, J., Brandl, H.B., Moodley, Y., Merkling, T., Leclaire, S., Blanchard, P., Danchin, É., Hatch, S.A., Wagner, R.H., 2013. Age-related differences in the cloacal microbiota of a wild bird species. BMC Ecol. 13, 11. https://doi.org/10.1186/1472-6785-13-11

Vicari, J.J., Peek, R.M., Falk, G.W., Goldblum, J.R., Easley, K.A., Schnell, J., Perez-Perez, G.I., Halter, S.A., Rice, T.W., Blaser, M.J., Richter, J.E., 1998. The seroprevalence of cagA-positive Helicobacter pylori strains in the spectrum of gastroesophageal reflux disease. Gastroenterology 115, 50–57. https://doi.org/10.1016/S0016-5085(98)70364-6

Videnska, P., Faldynova, M., Juricova, H., Babak, V., Sisak, F., Havlickova, H., Rychlik, I., 2013. Chicken faecal microbiota and disturbances induced by single or repeated therapy with tetracycline and streptomycin. BMC Vet. Res. 9, 30. https://doi.org/10.1186/1746-6148-9-30

Videnska, P., Sedlar, K., Lukac, M., Faldynova, M., Gerzova, L., Cejkova, D., Sisak, F., Rychlik, I., 2014. Succession and Replacement of Bacterial Populations in the Caecum of Egg Laying Hens over Their Whole Life. PLoS ONE 9, e115142. https://doi.org/10.1371/journal.pone.0115142

Vilvanathan, S., Kandasamy, B., Jayachandran, A.L., Sathiyanarayanan, S., Tanjore Singaravelu, V., Krishnamurthy, V., Elangovan, V., 2016. Bacteriospermia and Its Impact on Basic Semen Parameters among Infertile Men. Interdiscip. Perspect. Infect. Dis. 2016, e2614692. https://doi.org/10.1155/2016/2614692

Virecoulon, F., Wallet, F., Fruchart-Flamenbaum, A., Rigot, J.-M., Peers, M.-C., Mitchell, V., Courcol, R.J., 2005. Bacterial flora of the low male genital tract in patients consulting for infertility. Andrologia 37, 160–165. https://doi.org/10.1111/j.1439-0272.2005.00673.x

Vizzier-Thaxton, Y., Cox, N.A., Richardson, L.J., Buhr, R.J., McDaniel, C.D., Cosby, D.E., Wilson, J.L., Bourassa, D.V., Ard, M.B., 2006. Apparent Attachment of *Campylobacter* and *Salmonella* to

Broiler Breeder Rooster Spermatozoa. Poult. Sci. 85, 619–624. https://doi.org/10.1093/ps/85.4.619

Vodstrcil, L.A., Twin, J., Garland, S.M., Fairley, C.K., Hocking, J.S., Law, M.G., Plummer, E.L., Fethers, K.A., Chow, E.P.F., Tabrizi, S.N., Bradshaw, C.S., 2017. The influence of sexual activity on the vaginal microbiota and *Gardnerella vaginalis* clade diversity in young women. PLOS ONE 12, e0171856. https://doi.org/10.1371/journal.pone.0171856

Vreeland, R.H., Litchfield, C.D., Martin, E.L., Elliot, E., 1980. *Halomonas elongata*, a New Genus and Species of Extremely Salt-Tolerant Bacteria. Int. J. Syst. Evol. Microbiol. 30, 485–495. https://doi.org/10.1099/00207713-30-2-485

Wang, L., Fang, M., Hu, Y., Yang, Y., Yang, M., Chen, Y., 2014. Characterization of the most abundant Lactobacillus species in chicken gastrointestinal tract and potential use as probiotics for genetic engineering. Acta Biochim. Biophys. Sin. 46, 612–619. https://doi.org/10.1093/abbs/gmu037

Wang, Y., Ametaj, B.N., Ambrose, D.J., Gänzle, M.G., 2013. Characterisation of the bacterial microbiota of the vagina of dairy cows and isolation of pediocin-producing Pediococcus acidilactici. BMC Microbiol. 13, 19. https://doi.org/10.1186/1471-2180-13-19

Weng, S.-L., Chiu, C.-M., Lin, F.-M., Huang, W.-C., Liang, C., Yang, T., Yang, T.-L., Liu, C.-Y., Wu, W.-Y., Chang, Y.-A., Chang, T.-H., Huang, H.-D., 2014a. Bacterial Communities in Semen from Men of Infertile Couples: Metagenomic Sequencing Reveals Relationships of Seminal Microbiota to Semen Quality. PLoS ONE 9, e110152. https://doi.org/10.1371/journal.pone.0110152

Weng, S.-L., Chiu, C.-M., Lin, F.-M., Huang, W.-C., Liang, C., Yang, T., Yang, T.-L., Liu, C.-Y., Wu, W.-Y., Chang, Y.-A., Chang, T.-H., Huang, H.-D., 2014b. Bacterial Communities in Semen from Men of Infertile Couples: Metagenomic Sequencing Reveals Relationships of Seminal Microbiota to Semen Quality. PLOS ONE 9, e110152. https://doi.org/10.1371/journal.pone.0110152

Westneat, D.F., Rambo, T.B., 2000. Copulation exposes female Red-winged Blackbirds to bacteria in male semen. J. Avian Biol. 31, 1–7. https://doi.org/10.1034/j.1600-048X.2000.310101.x

White, J., Mirleau, P., Danchin, E., Mulard, H., Hatch, S.A., Heeb, P., Wagner, R.H., 2010. Sexually transmitted bacteria affect female cloacal assemblages in a wild bird. Ecol. Lett. 13, 1515–1524. https://doi.org/10.1111/j.1461-0248.2010.01542.x

White, J., Richard, M., Massot, M., Meylan, S., 2011. Cloacal bacterial diversity increases with multiple mates: Evidence of sexual transmission in female common lizards. PLoS ONE 6, e22339. https://doi.org/10.1371/journal.pone.0022339

Willeén, M., Hoist, E., Myhre, E.B., Olsson, A.M., 1996. The Bacterial Flora of the Genitourinary Tract in Healthy Fertile Men. Scand. J. Urol. Nephrol. 30, 387–393. https://doi.org/10.3109/00365599609181315

Worley, K., Collet, J., Spurgin, L.G., Cornwallis, C., Pizzari, T., Richardson, D.S., 2010. MHC heterozygosity and survival in red junglefowl. Mol. Ecol. 19, 3064–3075. https://doi.org/10.1111/j.1365-294X.2010.04724.x

Worley, K., Gillingham, M., Jensen, P., Kennedy, L.J., Pizzari, T., Kaufman, J., Richardson, D.S., 2008. Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. Immunogenetics 60, 233–247. https://doi.org/10.1007/s00251-008-0288-0

Xenoulis, P.G., Gray, P.L., Brightsmith, D., Palculict, B., Hoppes, S., Steiner, J.M., Tizard, I., Suchodolski, J.S., 2010. Molecular characterization of the cloacal microbiota of wild and captive parrots. Vet. Microbiol. 146, 320–325. https://doi.org/10.1016/j.vetmic.2010.05.024

Yániz, J.L., Marco-Aguado, M.A., Mateos, J.A., Santolaria, P., 2010. Bacterial contamination of ram semen, antibiotic sensitivities, and effects on sperm quality during storage at 15°C. Anim. Reprod. Sci. 122, 142–149. https://doi.org/10.1016/j.anireprosci.2010.08.006

Yildirim, S., Yeoman, C.J., Janga, S.C., Thomas, S.M., Ho, M., Leigh, S.R., Consortium, P.M., White, B.A., Wilson, B.A., Stumpf, R.M., 2014. Primate vaginal microbiomes exhibit species specificity without universal *Lactobacillus* dominance. ISME J. 8, 2431. https://doi.org/10.1038/ismej.2014.90

Zuk, M., Thornhill, R., Ligon, J.D., Johnson, K., 1990. Parasites and mate choice in red jungle fowl. Integr. Comp. Biol. 30, 235–244. https://doi.org/10.1093/icb/30.2.235