Seasonal trends in the midgut microbiota of honeybees

Jane Ludvigsen

Norwegian University of Life Sciences
Department of Chemistry, Biotechnology and Food Science
Master Thesis 60 credits 2013

Aknowledgements

This master thesis was performed at the Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Science with Professor Knut Rudi and Ph.Ds; Ekaterina Avershina and Anbjørg Rangberg as supervisors.

I would first like to thank Professor Knut Rudi for your enthusiasm and positive attitude towards all aspects of this thesis` work, including me. Thank you Ekaterina; your technical insights amazes me of which I could not have managed without. Thank you Anbjørg for being helpful and supporting at the start of this work, when I needed it the most, and thanks Claus for helping me get rid of my fear of bees and for always being positive and at hand when needed. To all of the employees in the Environmental microbiology and Microbial diversity groups, you made me look forward to every day at work through your positive remarks and by making me feel welcome- thank you.

Thanks to the staff at Sykehuset Østfold -HF, bacteriology apartment for technical supplies.

I would also like to thank Professor Gro Amdam for letting us use your bees and for helping in the set up of this work, and also Nick, the Arizona lab technician, who sample the Arizona bees. Not to forget, all of the employees in the Integrative neuroscience and sosiogenomics group at the Norwegian University of Life Science for all help regarding bee work.

A special thanks goes to my boyfriend Christian, who`s patience and support kept me focused and helped me sustain even through stressful times. To my Brother and Sister and their families- a big thank you for believing in me and for your endless support.

In loving memory of my Mother and Father

You inspired me to reach beyond myself and to follow my heart

Ås, May 2013

_____

Jane Ludvigsen

Abstract

Honeybees are prominent crop pollinators and thus important for an effective food production. Without any apparent reason, massive bee colony losses appeared around the world, which started about a decade ago, and thereby the focus was set on bee health. The gut microbiota composition is crucial for health and immune system development both in mammals and insects, and factors shaping the gut microbiota has been intensely studied in humans and animals, three of which are frequently addressed are diet, hence exposure to the surrounding environment, host symbiosis and host genetics. The bee gut microbiota is bee specific and dominated by eight bacteria phylotyphes, which are found in almost all bees, and thus they are believed to be bee symbionts. This specific and simple microbiota in combination with the fact that bees are numerous, easy to monitor and the gut is easily accessible, the bee gut has been found to be a well suited model for gut microbiota studies. In this thesis the stability of the midgut microbiota of honeybees was addressed through sampling of bees throughout a season from May until October, where the impact of diet was the main focus. The data was analyzed using Sanger sequencing in combination with MCR-ASL computer analysis, which identified the dominating bacteria in the midgut and their relative ratio, hence bacteria composition. QPCR analysis was used to calculate the relative bacteria load in the data.

The overall data showed that the midgut microbiota was not stable, and the dominating bacteria displayed apparent trends throughout the season. Analysis of distinct data sets including bees exposed to different diets, and with varying nutritional preferences, reviled bacteria in the midgut, which had nutritional preferences and exhibited possible niche adaptations. In addition results specified that the impact on the midgut microbiota by different diets was greater than the impact of same diet for a longer period of time. Biochemical profiling of the bee symbiont *G.apicola*, showed that this bacteria can utilize both glucose and fructose, something which has not been proven before using phenotypical tests. Competitive interactions between the two bee symbionts; *G.apicola* and *S.alvi*, was also detected.

Sammendrag

Honningbier er viktige plante bestøvere og dermed også viktige for en produktiv matproduksjon. For ca et tiår siden begynte mange bie kolonier å dø uten at en bestemt grunn kunne påvises, og dermed ble det satt fokus på god bie helse. Bakteriesammensetnngen i magen, er ytterst viktige for god helse og stimulering av immunsystemet, og faktorer, som påvirker bakteriefloraen i magen har blitt intensivt analysert både i mennesker og dyr. Tree faktorer som ofte vender tilbake som viktige er diet, derav påvirkning av miljøet rundt, symbiose mellom vert og bakterier og genetikk hos vert. Bakteriefloraen i bie mage er spesifikk, og de bakteriene, som dominer tilhører åtte phylotyper. Disse har blitt funnet hos nesten alle bier og regnes dermed til å være bie symbionter. Denne spesifikke og enkle bakteriefloraen i kombinasjon med at det finnes mange bier, at de er enkle å overvåke og at mage-tarm systemet er lett tilgjengelig, gjør at bier er blitt sett på som vel egnede modeller for mageflora studier. I denne oppgaven ble stabiliteten til midtmage floraen hos honningbier adressert ved å samle bier gjennom en sesong fra Mai til Oktober, hvor det ble særlig lagt vekt på effekten av diet. Ved å bruke sanger sekvensering i kombinasjon med MCR-ALS data analyse ble de dominerende bakteriene i midtmage hos bier identifisert og den relative bakteriesammensetningen ble bestemt. QPCR analyse ble brukt til å beregne relativ bakterie mengde.

Resultatene viste at bakteriefloraen i midtmage ikke var stabil og at de dominerende bakteriene hadde tydelige varierende trender igjennom hele sesongen. Analysering av separate datasett, som inneholdt bier med varierende diet og preferanse for ulike næringsstoffer, viste at det fantes bakterier i midtmagen med nærings preferanser og at de muligens hadde spesialisert seg til visse oppgaver eller miljøer. I tillegg så viste resultater at effekten på bakteriefloraen av ulike dietter var større enn lik diet over lengre tid. Biokjemiske tester av bie symbionten G.apicola, viste at denne bakterien kan nyttiggjøre seg av både glukose og fruktose, noe som ikke før har blitt bevist ved bruk av fenotypiske tester. Konkurrerende samspill mellom de to bie symbiontene *G-apicola* og *S.alvi* ble også funnet.

Abbrevations

API	--	Analtyical profile index
BFG	--	Beta-,Firm -5 and Gamma-5 phylotypes
BLAST	--	Basic Local Alignment Tool
BP	--	Base pairs
CCD	--	Colony colaps disorder
CFU	--	Colony forming units
CO_2	--	Carbon dioxide
Ct	--	Treshhold cycle
dd	--	Dideoxy
DNA	--	Deoxyribonucleic acid
ds	--	Double stranded
EtOH	--	Ethanol
FISH	--	Fluorescence in situ hybridisation
gDNA	--	Genomic DNA
GI-tract	--	Gastro intestinal tract
HMR	--	High resolution melt
HP	--	High pollen
LP	--	Low pollen
MCR-ALS	--	Multivariate Curve Resolution with Alternating Least Squares
NCBI	--	National Center for Biotechnology Information
Min	--	Minutes
NTP	--	Nucleotide triphosphate
OTU	--	Operational taxonomy unit
PBS	--	Phosphate buffered saline
PCR	--	Polymerase chain reaction
qPCR	--	Quantitative polymerase chain reaction
RDP	--	Ribosomal Database Project
rRNA	--	Ribosomal ribonucleic acid
RT	--	Room temperature
SD	--	Standard deviation
Sec	--	Seconds
SEM	--	Standard error of the mean
ss	--	Single stranded
TSA	--	Tryphtose soy agar
TSB	--	Tryphtose soy broth
UMB	--	Norwegian University of Life Science
WT	--	Wild type

Table of content

1.0 Introduction

Bees are an important crop pollinators and thus important for a productive food production (Vanengelsdorp & Meixner 2010). When the sudden collapse of several bee colonies (Colony Collapse Disorder, CCD) in the USA, Europe and Asia began without any apparent reason around ten years ago, the focus was set on bee health (Evans & Schwarz 2011). It is important to preserve the bee colonies healthy and viable, and the occurrence of CCD renders the bee colonies to collapse due to the massive loss of adult foraging bees. Normally some bee colonies also die due to seasonal variation during the winter-spring transition. The reason for this, as for the cause of CCD, are up to today not been determined, but studies done by Dainat et al. (2012) and Amdam et al. (2004a) show the parasite *V.destructor* to be important in weakening the bee community health and physiology respectively, and also viruses and general stress induce CDD, which again seems to change the gut microbiota composition (Cornman et al. 2012). Many different factors seem to play a role, as for instance; use of pesticides and antibiotics, nutrition abundance and variance, bee pathogens; including parasites, protozoans, viruses, pathogenic fungi and bacteria, and also climate change, and in addition, the overwintering bee`s nutrition storage capacity plays a crucial part for survival and health of single bee colonies (Evans & Schwarz 2011). Health can be linked to gut microbiota in many mammals and insects and therefore a lot of focus has been set on this topic in the last years, including both symbionts and pathogens. Since there seems to be a complex set of reason for bee colony deaths, a reasonable approach to gut microbiota would be to address it as a complex community, and to investigate the bacteria in their normal habitat, which was the main objective in this thesis.

1.1 The honeybee – *Apis mellifera*

1.1.1 Important nutrients and their seasonal variation

Apis mellifera is the European honey bee, which has been managed for centuries by people across the world, due to their capability to produce honey and wax. Throughout the 19th century honey was the only natural sweetener available to people in Europe (Vanengelsdorp & Meixner 2010). Naturally, the bees would use the produced honey for food storage through the winter when foraging and food availability is low, but managed honeybee colonies do not

have this opportunity, because as honey is produced by the bees during summer, it is taken away in autumn and used for food production by people. Instead they are fed sugar during the winter and the sugar mix is composed to equal the natural honey reserve. In Scandinavia and equally tempered zones, no foraging occurs when temperature drops, because the bees cannot survive outside the hive when it is to cold. Then they huddle inside the hive to keep warm and their metabolism slows down. The honey exhibit naturally occurring antiseptic properties and is important for the bee, in addition to nectar and pollen, for building food storages in the form of fat bodies. For the bees nectar is a source of sugar, and pollen is the main source of protein, vitamins, minerals and lipids, which in particular essential for the larvae and young bees for proper development. Different pollen contains different minerals and vitamins and a variety of pollen types are necessary to cover the bees` nutritional needs. The natural season variability in food intake follows the seasonal climate change and the bees use the plants available at that time of season for food and honey production (Davis 2004).

The *Apis mellifera* is highly adaptable and can easily be transported around to serve as crop pollinators in large scale food production industry and agriculture (Vanengelsdorp & Meixner 2010). This happens around the world today and a huge number of bee colonies are set out into fields with different food production, as for instance apples or oil seeds and then moved around to the next field, which gives the bees limiting possibilities to include a diverse nutritional diet.

1.1.2 Lifecycle

The queen bee lays eggs throughout the whole season, but with a peak in mid-late May to build up the colony for the summer, but also for swarming to new places, if necessary, if the hive is too small (Davis 2004). This is why beekeepers build multiple boxes on top of each other to expand the hive, so that the queen will not swarm. From the egg is laid it only takes 21 days until a newly hatched bee emerges from its nest, and a seasonal cycle is seen, where number of newly hatched bees is varying in accordance to climatic changes. All workers are female and the male bee is called drones, which main purpose is to mate with the queen. The life of a worker honey bee is complex from the time the bee is born as an egg, to become a larvae, and then to develop into a worker bee, which life spans a range of stages until it dies as a adult foraging bee (fig 1.1). There is a hierarchy of work tasks that has to be done inside and outside the colony to keep the colony strong and functional. The stages in a honey bees` life are environmental dependent, which will to some extent affect ageing and then again affect

which stage the worker bee will develop into (Davis 2004). A generalized approach is that they first become nurse bees, which feed the larvae, and then they develop into housekeeping bees, which clean the hive and take care of the queen and food supplies. The next stage involves maintaining the bee hive and building combs, and maintaining food reserves. Guarding bees make sure no intruders come through the opening of the hive and, the last stage is the foraging bee, which collects pollen and nectar, which then is shared with the rest of the colony.

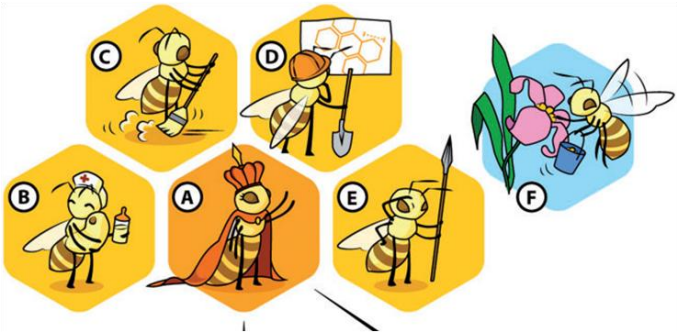

Figure 1.1: The different stages in a bees life from picture B to F. A= Queen, B= Nurse bee, C= Cleaning inside the colony bee, D= maintaining food preservatives, E= Guard the opening bee, F= Foraging bee. Picture is modified and reprinted from (Rangberg et al. 2012)

1.1.2 The digestive system

The transition from larvae to bee is extraordinary and evolves breakdown of the larvae internal secretion system and buildup of the new bee digestive system. Some similarities between the human gastro intestinal tract (GI- tract) and the bee GI- tract can be seen (fig 1.2a&b), because they can both be separated into three distinct parts; Forgut, Midgut, and Hindgut.

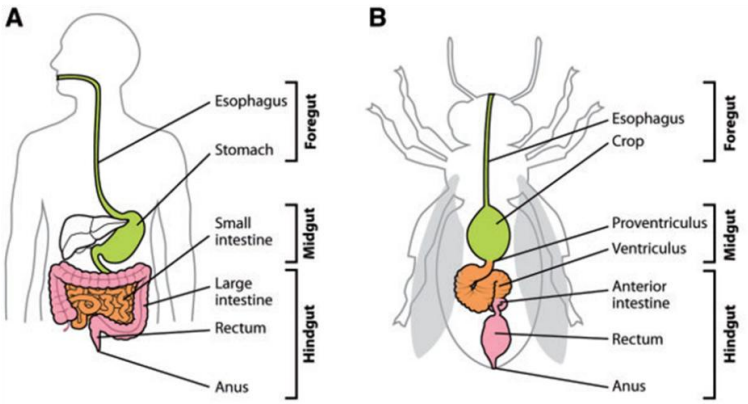

Figure 1.2: The digestive systems of humans and honeybees. The human intestinal tract (A), and the honey bee intestinal tract (B). The coloring is comparable in respect to function of the different parts of the digestive system. Picture is modified and reprinted from (Rangberg et al. 2012).

The GI- tract of the bee is a tube, which spans from the mouth to the anus (fig 1.2b). The crop functions as the storage of nectar and water, and the midgut (ventriculus) is where the digestion of nutrients takes place (Crailsheim 1988; Crailsheim 1990; Davis 2004). The midgut is embedded with a peritrophic membrane, probably for protection of the epithelia wall, and exchange of soluble substances through this membrane to the haemolymph (circulatory system) is one of the important functions of the midgut. The hindgut includes the anterior intestine (ileum) and the rectum (Davis 2004).

1.2 Gut microbiota- function and importance of health

1.2.1 Transfer of bacteria within the bee colony

When the worker bee emerges, and the digestive system is rebuilt, it is sterile and has to be colonized by the shared bacteria flora within the colony. Because the bees are social insects, like humans are social, they shear a common bacteria flora, which is transmitted from adult to newly hatched bees through feeding and secretion inside the colony hive (Koch et al. 2013). The commensal gut bacteria community, either humans or bees, will develop with age and expands to include more numerous bacteria groups until it reaches a stable composition (Martinson et al. 2012; Palmer et al. 2007). The development of gut microbiota composition is influenced by the surroundings, and is especially susceptible to change during the development.

1.2.2 Gut microbiota and influence on health

The commensal flora includes bacteria, which are potential symbionts. These symbionts have most likely coevolved with their host (Ley et al. 2006a; Moran 2006), and in bees different bacteria have been shown to protect against pathogens (Evans & Armstrong 2006; Koch & Schmid-Hempel 2011b; Vasquez et al. 2012; Wu et al. 2013). In addition, these symbionts will have specified tasks in the gut (Engel et al. 2012), and this niche specificity is important for a well functioning gut. The sharing of bacteria flora does not limit itself to the commensal flora but also pathogens will be transmitted, and this could have seriously consequences if the newly hatched bees are infected, due to their poorly developed immune system (Wilson-Rich et al. 2008). This can coincide with the fact that the health of many animals, insects and also humans depends on a well functioning gut microbiota, and a not fully developed microbiota

or a change in the gut microbiota composition can have effect on the nutritional uptake thus physiology (De Filippo et al. 2010; Hildebrandt et al. 2009; Ley et al. 2006b) and the immune system development (Vebø et al. 2011; Young et al. 2004). The innate immune system in the first line of defense in invertebrates and in bees it shows altered expression and activation when pathogens are encountered (Chaimanee et al. 2012; Evans 2004). It is possible that this immune system is shaped by the gut symbionts (Nyholm & Graf 2012) and therefore a proper development of and correct composition of the gut microbiota is of outmost importance. A review by Hamdi et al. (2011) also argue for the importance of gut symbionts in bee health and that dysbiosis in the gut microbiota could lead to disease.

Many studies have been done on nutrition impacts on humans and animals, and it has been demonstrated that different nutrition intake has an effect on the metabolism and gut microbiota composition (De Filippo et al. 2010; Fabbrini et al. 2013; Hildebrandt et al. 2009; Ley et al. 2006b), and a study done by Alaux et al. (2011) also showed that different diets in honey bees influenced the honey bee metabolism. Many of these studies show that the gut microbiota composition is important for the function of the host and thus the honeybee, because the bacteria digest nutrients and is thereby involved in the host`s nutrition uptake, and also that the bee surroundings play a dominating part of shaping the gut microbiota. In addition, some studies have addressed the importance of genetics in gut microbiota, and found that the overall composition is thereby controlled (Tims et al. 2013).

1.2.3 Bee gut bacteria

Studies on the 16S rRNA gene report that the bee gut harbor a distinct microbiota (Jeyaprakash et al. 2003) of eight abundant phylotypes, which makes up ~95% of the total bacteria (Martinson et al. 2011; Moran et al. 2012). These studies represent both clone libraries of preamplified gut bacteria 16S and 454 pyrosequencing, and despite different methodology throughout the analysis, they found the same distinct gut bacteria in the *Apis mellifera*, which were different from other insects. The Moran lab and her associate researchers were the first to classify the eight phylotypes into readily comparable taxonomic group, and this taxonomy is presently standing. Table 1.1 lists the eight phylotypes and show the bacteria taxonomy in which the phylotypes clustered accordingly to the study done by Martinson et al. (2011) in which the taxonomy was set. The gamma -1 phyloype clustered between the *Enterobacteriaceae*- and *Pasteurellaceae bacterium*, which makes it difficult to assign the accurate taxonomy classification. Two of the phylotypes have also been assigned

taxonomy on the strain level gamma -1 phylotype = *Gilliamella apicola*, and the beta phylotype = *Snodgrassella alvi*.

Table 1.1: Taxonomy classification of the eight abundant phylotypes identified from honeybee gut. The mid column shows where the phylotypes cluster (Martinson et al. 2011).

Moran`s phylogroups	Taxonomy
Alpha -1	Rhizobiales
Alpha -2	Acetobacteriaceae and Gluconobacter
Beta	Neisseriaceae
Bifido	Bifidobacterium
Firm -4 and firm 5	Lactobacillus
Gamma -1 and Gamma -2	Gammaproteobacteria

In addition to these studies on whole bee guts, one study using terminal restriction fragment lenght analysis (T-RFLP) on the honey bee midut bacteria composition, is known (Disayathanoowat et al. 2012). In this study they found that the identifyed bacteria all belonged in four groups, which they named the *Beta*- and *Gammaproteobacteria*, *Firmicutes* and *Actinomycetes*. They tested bees of different ages and bees, in different geographic locations, and they found that the *Firmicutes* was present only amongst worker bees and that the *Gammaproteobacteria* was present in all ages. They also found some divergence in bacteria composition between places, which has also been seen in a study by (Moran et al. 2012). This study also shows that the colony bacteria composition varies amongst bees at one location and that multiple strains of each of the phylotypes; *Gilliamella apicola* and *Snodgrassella alvi*, can be present in a single bee.

A different approach to study the bee gut microbiota was conducted by (Martinson et al. 2012), in where they investigated the abundance of three of the most abundant phylotypes; the beta, gamma -1, and firm -5 (BFG), both amongst different ages and gut parts; Crop, Midgut, Ileum, and Rectum, using qPCR, 454 pyrosequencing, and FISH analysis, which results are important to think about when analyzing the bee gut. The midgut contained about 1-4% of total amount of BFG and the Gamma -1 was the most abundant bacteria amongst the BFG, but also the beta phylotype was relative abundant compared to the firm-5. In the midgut the amount of BFG varied amongst different ages, where a low amount was seen at day1 and a dramatic increase occures after day1 and the adult foraging bees (30 days) harbored most bacteria. This difference will give bee to bee variation if bees are sampled from a colony with

bees of different ages. Fluorescence in situ hybridisation (FISH) analysis showed that the midgut is closely colonized with the beta phylotype along its entire gut wall and the bacteria load increases towards the ileum where the gamma -1 phylotype dominates in the interior of the gut, giving indication of a kind of niche specific property by the beta phylotype, by attaching to the gut wall. The same study (Martinson et al. 2012) also found four dominating bacteria groups or operational taxonomy units (OTUs) when investigating the midgut, which corresponded to the beta, gamma, -1, gamma -2 and firm -5 phylotypes.

A recent study, in which bee gut bacteria were cultivated, isolated 35 bacteria and all of them could be classified using the same phylotypes as described by sequencing methods (Yoshiyama & Kimura 2009). In addition, a phenotypic profiling of two bacteria isolates; *G.apicola* and *S.alvi*, have been done by Kwong and Moran (2012), which gives important information about metabolic traits, growth characteristics and which also addressed the difficult phylogenetic relationship between the *G.apicola* and *Enterobacteriaceae/ Pasteurellaceae* families, and this cluster of bacteria has been proposed to give rise to a new order; Orbales and family; Orbaceae. Lately functional analysis have been done on the most prominent bee specific bacteria, and the niche specificity found in other studies has been verified (Engel et al. 2012; Engel & Moran 2013).

1.3 Culture independent analysis

During the last decades new methods of bacteria classification have taken over for or supplement the more traditional culture dependent phenotypic approaches and the DNA-DNA hybridization method. These new methods use the genome sequence, from bacteria, for identification purposes and no acquirement of culturing is needed.

To be able to use genomic material for analysis, one must first lyse the cells and then extract the DNA from the cells. Figure 1.3 gives a schematic outline of a general culture independent analysis approach. The choice of cell lysis- method depends on the material which shall be lysed (Burgmann et al. 2001). One method is mechanical lysis, using bead beating with glass beads, and this is an efficient way of lysing all cells, Gram negatives and Gram positives, yielding negligible bias, especially in easy to lyse materials such as stool samples (Salonen et al. 2010; Skanseng et al. 2006). The DNA extraction method is used to separate the DNA from cell wall structures and other compounds, which could potentially inhibit further

analysis. DNA extraction can be done in several ways either by chemical separation of the DNA from the rest of the cell (Moran et al. 2012), or by the use of readymade extraction columns (Koch & Schmid-Hempel 2011a), but also by a much used method, which is based on the use of magnetic silica particles in combination with chaotropic salts (Boom et al. 1990). Because of the small amount of DNA present in a cell, an amplification step is often needed for further analysis. This amplification step is often done by Polymerase Chain Reaction (PCR), and this method is described more in detail below. Depending on the application of the analysis this PCR reaction is either directly connected to the sequencing reaction, or an additional modified sequencing PCR is performed for labeling the DNA. The different methods of sequencing are explained more in detail later in this chapter.

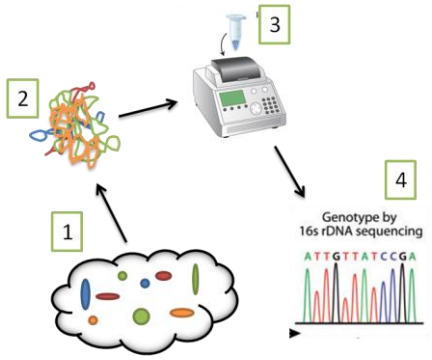

Figure 1.3: Flow chart of a general work flow for culture independent analysis. Environmental sample with different bacteria (1), Extracted DNA (2), PCR of the desired amplicon (3), Sequencing of the PCR product (4). Part of chart taken from New England BioLabs®inc`s webpage: https://www.neb.com/applications/dna-amplification-and-pcr and (Rangberg et al. 2012)

1.3.2 16S rRNA gene

The bacterial DNA contains different householding genes that can be used for identification of that bacteria. One such gene is the 16S rRNA, which holds house-holding functions as part of the small ribosomal subunit (Woese 1987). Due to the constitutive nature of this gene in every bacteria, some regions are well preserved , but it also contains variable regions (fig.1.4), which gives a span of levels for identification purposes (Lu et al. 2000; Van de Peer et al. 1996).

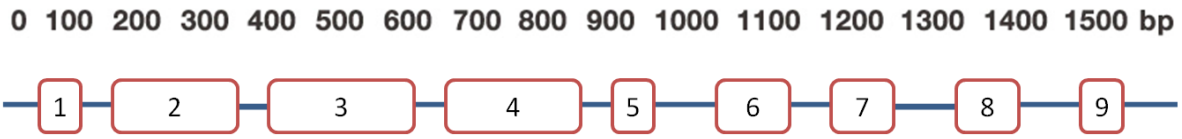

Figure 1.4: Illustration of the 16S rRNA gene showing its conserved regions in blue and its variable regions in red in accordance with its basepar (bp) order. The figure is modified from (http://www.alimetrics.net/en/index.php/dna-sequence-analysis)

This gene is up to date the most used gene for comparing evolutionary relationships between bacteria, and is used in studies on bacteria composition in combination with very different methodology (Jeyaprakash et al. 2003; Palmer et al. 2007; Rudi et al. 2007). Sequences derived from this gene is deposited in different databases like National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) and Ribosomal Database Project (RDP) (Cole et al. 2009), which can be used as a reference for sequence comparison and thus bacteria identification.

1.3.3 Qualitative and quantitative- PCR

Qualitative PCR

Polymerase chain reaction (PCR) is a method which effective amplify genetic material. From a small amount of template one can get a high amount of the desired amplicon. A part of the method was first described by Kleppe et al. (1971), and then to be fully developed by Mullis (1990). The reaction proceeds in a three step manner, controlled by temperature, which include denaturation, annealing, and elongation. The denaturation step is when the dsDNA is separated due to high temperature, then, when lowering the temperature, two specific primers will anneal to their target sites on each side of the desired amplicon. The annealing temperature is essential for correct binding of the primers and ensures specific amplification. Finally the polymerase elongates the annealed primers, so that the amplicon is copied in its full length. When this reaction is cycled the amplicon will be exponentially amplified until the reaction reaches a hold. This amplification makes it possible to continue with further analysis on the desired PCR product. Qualitative PCR is normally just called PCR or diagnostic PCR because it only detects and amplifies the target in a sample, without giving an amount of the target present in the sample analyzed, and an additional step is necessary for detection and amplicon quantification.

Quantitative PCR

Quantitative PCR (qPCR) method is used to determine the amount of bacteria present in a sample, and it gives an estimate of the load of amplicon in the sample analyzed, either by endpoint measurement or by real time measurement (Kubista et al. 2006; Zhang & Fang 2006). Monitoring the PCR cycling in real time, by using a fluorescent compound, that will give out a signal during each cycle, which is proportional to the amount of bacteria in that sample (Holland et al. 1991), is today the most used application of this analysis.

By using a specific fluorescent labeled probe, that targets inside the amplicon, the signal will only reflect the amount of amplicon in the PCR. There are different probe designs available, but a much used one is the TaqMan probe (Haugland et al. 1999). The Taqman probe is designed to emit no light in its original form, due to a quencher molecule, which absorbs the light when in the presence of the fluorochrome, but when the polymerase slides along and elongates the amplicon, the quencher will separate from the fluorochrome on the TaqMan probe and light is emitted (Kubista et al. 2006).

If an intercalating fluorescent dye is used, the signal will reflect all dsDNA in the sample, which will also include primer-dimers, and therefore it is necessary to perform a meltingcurve (high resolution melt analysis, HRM) analysis, to be sure that most of the signal reflect the correct amplicon (Ririe et al. 1997). The meltingcurve plotted in the 2nd derivative against temperature, will present a peak in the fluorescence plot where the fluorescence drops due to DNA denaturation by temperature increase. Primer-dimers and targeted amplicon is differentiated by the peak at different temperatures, due to the length difference and bonding properties of the two sequences (Kubista et al. 2006). There are different types of dyes that can be used, and two of the most used dyes today are SYBRGreen (Morrison et al. 1998) or EvaGreen (Ihrig et al. 2006; Mao et al. 2007).

The cycle time (Ct) will be calculated, by the machine, by setting a threshold for the detection of the fluorescence signal, and this Ct value reflects the amount of DNA in the sample (Kubista et al. 2006).

1.3.4 Sequencing

The sequencing field has advanced from single bacteria sequencing (Sanger sequencing) by the use of staining for visualization and separation by poly acrylamide gel electrophorese (Sanger et al. 1977) and later by the use of fluorochromes (Smith et al. 1986) to an explosive increase of different methods over the last ten years, which include sequencing genes from whole bacteria communities (mixed sequencing) (Trosvik et al. 2007) and deep sequencing, which goes by the term Next-generation sequencing (NGS).

Sanger sequencing

The Sanger sequencing method uses labeled dideoxy nucleotides (ddNTP) for labeling and sequencing ssDNA constructed from pure isolates. The principle of the method is based on capillary electrophoreses in which labeled ssDNA is separated through a capillary which

enable nucleotides to be detected one by one. The detection is done by labeled ddNTP`s that binds to DNA under synthesis in a labeling PCR reaction, and because ddNTP lacks the OH-group responsible for crosslinking NTP`s the synthesis of DNA stops. This gives ssDNA with different length labeled with a ddNTP at the end. The four ddNTP`s are labeled with four diffenent fluorochromes that emits light when detected in the capillary. This light is captured and accumulated at each position as the sequence passes through the capillary, giving an electropherogram that equals the sequence of the nucleotides. One advantage of the Sanger sequencing is that it displays long reads.

Mixed sequencing

Mixed sequencing uses the same method as described on Sanger sequencing, but it is applied to environmental samples with more than one bacteria present, hence mixed sequencing, and was first applied by Trosvik et al. (2007). The resulting electropherogram reflects the dominant bacteria in the sample, and a whole dataset will then reflect the bacterial community in the sample. This dataset can be analyzed further using a multivariate curve resolution with alternating least squares analysis (MCR-ALS), to yield more detailed information about the dominant bacteria. Figure 5 gives a simplified description of the method.

The MCR-ALS method was developed by (Zimonja et al. 2008) and is used in several studies of environmental samples (Avershina et al. 2013; Sekelja et al. 2012). The MCR-ALS program compares all the electropherograms obtained from the samples in the dataset and finds the most dominating components (bacteria) for them all. This is done by first aligning the elecropherograms and then scanning them for nucleotides one by one, and finding the ones that are most frequent at each position (fig 1.5a). Each dominant component is assigned its unique sequence, which can be used to identify the bacteria using Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/) or RDP (http://rdp.cme.msu.edu/) databases. In addition a relative ratio of the dominant components in each sample is calculated (fig 1.5b).

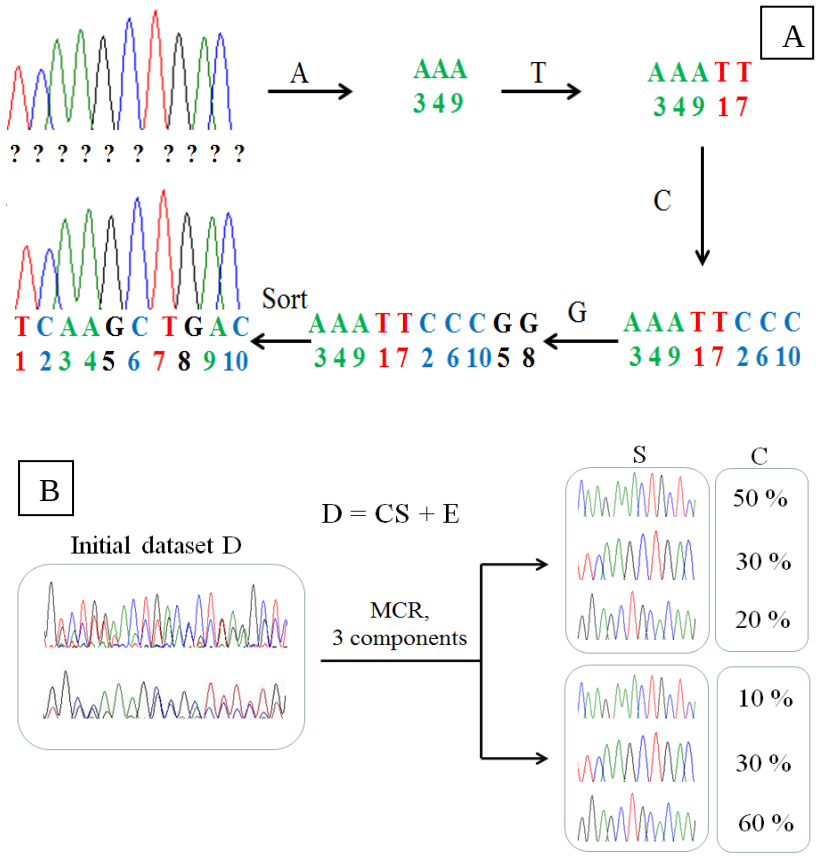

Figure 1.5: Simplified description of the mixed sequencing combined with the MCR-ALS analysis. (A): The aligned mixed sequences are scanned for nucleotides one by one at each position in all of the samples in the dataset. The program looks for the components that are the most dominant in all the samples in the dataset, and the nucleotides dominating at each position is sorted and displayed as the component sequence. (B): The relative ratio of the different dominating components in the dataset is calculated for each sample. Here shown with 3 components in two samples. The figure is based upon and reprinted from (Avershina et al. 2013).

Next generation sequencing

This term includes all sequencing methods that either analyses the whole genome of a single bacteria or the whole community in a sample. The whole community approach includes sequencing only one gene, like the 16S rRNA, or the whole community genome in a sample. This has given us an enormous advantage in finding out more about which bacteria live together and how they interact in different environments such as the bee gut (Engel et al. 2012; Engel & Moran 2013). There are many methods available today and different applications are needed depending on what to investigate (Glenn 2011), such as 454 pyrosequencing (http://454.com) first applied by Roche, Illumina sequencing (http://www.illumina.com), and IonTorrent (http://www.iontorrent.com). The methods create thousands of short sequences which are computer intensive to assemble and some present biases are method dependent (Quail et al. 2012; Schwartz et al. 2011).

Illumina sequencing

Before the sequencing reaction can take place a PCR with different forward and reverse primers has to be performed on the samples. The primers can be tagged, and this creates amplicons, which can be traced back to the original sample when samples are pooled for sequencing. This method saves money and a lot of samples can be analyzed together. Illumina sequencing is preformed on glass slides where amplicons are amplified, after paired end ligation, on the slide by bridging to the next neighbor oligo attached on the slide. Spots with similar amplicons will result, which are sequenced using reversible fluorescence end labeled dideoxy nucleotide detection (ddNTP), where all four ddNTPs are added and a signal is detected from the different bacteria spots, then the label is washed away and a new set of ddNTPs are added and the nucleotides are read one by one (http://www.illumina.com).

1.4 Culture dependent analysis

Culture dependent analyses are considered the traditional approach for identifying bacteria and makes the fundament on which today`s bacteria are classified. Culturing of bacteria is important when one wants to investigate morphology, functional properties and interactions between bacteria. To find out if and when different genes are expressed and in what niche environment a bacteria may grow, it is often necessary to study the phenotypic traits of a bacteria under growth conditions (Carini et al. 2013). The morphology and physiology of a bacteria can give much information about its properties and relatives, and working with pure cultures gives an advantage when exploiting biomarkes such as lipopolysaccarides, and 16S rRNA, for identification purposes (Kwong & Moran 2012). One can expect bacteria with similar physiology to have the same function, but that they have evolved from the same starting point, is on the other side not necessary the case (Fraser et al. 2009). Because many bacteria have evolved in different directions through evolution, due to dispersal in various environments and selection pressures, the genome can have changed in such a way that locking only at the gene composition, without taking into account the physical properties of these bacteria, may give a misleading guidance about their properties and relatedness, and vice versa (Fraser et al. 2009; Willey et al. 2009). To combine both culture independent and culture dependent analysis when working with bacteria is preferable, to cover all aspects of the bacteria metabolism for the bacteria of interest (Shade et al. 2012).

1.4.1 Growing bacteria

Bacteria can live in the almost every environment and this creates bacteria within different environmental niches, which demand and adjust to a variety of nutritional needs and conditions. When growing bacteria it is crucial to supply them the nutrition and growth conditions they need to be able to grow. Different conditions like atmosphere, temperature, nutrients, pH and light must be controlled to optimize bacteria growth (fig 1.6), and working with bacteria isolates it is important to have pure cultures, which can be achieved by streaking one colony onto a new agarplate numerous times. Bacteria growth rates can vary enormously, from only about 20 min to several days. If grown in a liquid media a most bacteria will follow a growth curve including an initial state, where the bacteria adjusts to the new environment, then an exponential state will emerge, with doubling times proportional to the growth rate, and at last a steady state in with bacteria begin to die, due to lack of nutrients or accumulation of toxic compound (Jasnos et al. 2005). When bacteria are grown together and if they compete for the same nutrient one of them will outgrow the other, or bacteria can also be mutualistic and dependent on each other, either by metabolizing end products from the other bacteria or that they perform different tasks which stabilize the environment they live in (Chen & Weimer 2001)

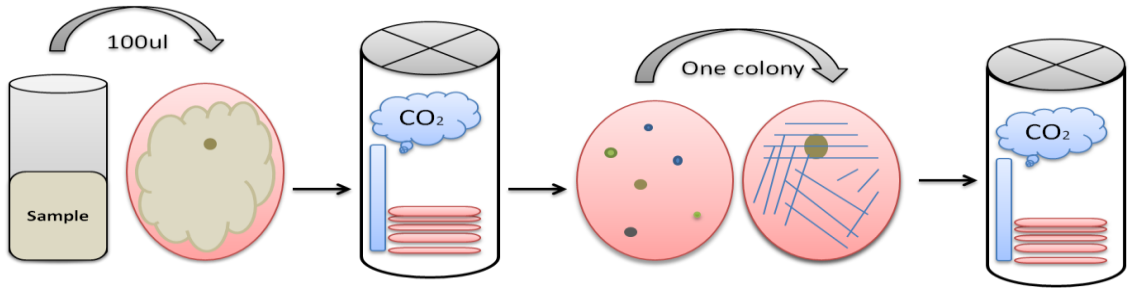

Figure 1.6: Flow chart showing how to grow bacteria in CO_2- enriched atmosphere and isolation of pure culture. 100µl of the sample is transferred to an agar plate containing growth medium for the desired bacteria. The sample is spread all over the plate, and grown in a chamber; here shown with agar plates (in red) and a sachet producing CO_2- enriched atmosphere (in blue). The chamber is closed with a lid on top (grey). This chamber can be used with the preferred temperature. After incubation for the required time, that the bacteria need for growth, colonies can be counted or spread one by one for pure culture isolation. The spread plate is the incubated again under the same conditions.

1.5 The aim of the thesis

The honey bee *Apis mellifera* is a social bee and their social interactions amongst one another inside a colony can readily model a human family or other social animals when it comes to sharing and transfer of bacteria (Martinson et al. 2012). The bee is numerous, easy to monitor,

and its gut microbiota composition is relative simple compared to other animals, and these aspects make the honey bee a suitable model for gut microbiota studies.

Studies done on bee gut microbiota have been focusing on what kind of bacteria that are present in bees around the world and what function they perform in the bee gut (Engel et al. 2012; Jeyaprakash et al. 2003; Martinson et al. 2011). One study has also addressed the differences and similarities in gut microbiota composition between places and between bees from different colonies (Moran et al. 2012).

In reference to previously studies done on gut microbiota of honeybees, the aim of this thesis was: **To determine the stability of the midgut microbiota composition in honey bees throughout a season.**

The rationale is that we still lack knowledge about the seasonal stability of the honeybee microbiota. This knowledge will be important for understanding the effects, of other factors such as diet, geography and genetics, on gut microbiota composition.

In this thesis we focus on the mid gut where the digestion of nutrition is taking place, and we include the following subgoals:

- Determine the midgut composition throughout a season for stationary bee colonies.
- Determine the impact of honey bee diet on midgut microbiota.
- Determine the impact of potential competition between midgut bacteria.

1.5.1 Choice of methodology

We chose an approach that combines mixed sequencing with MCR- ALS analysis in order to determine the stability of the dominant bacteria in the midgut microbiota composition. Illumina deep sequencing was used for verification of the MCR-ALS, while QPCR was applied for ratio calculation between amount of bacteria DNA and bee DNA, which addressed the relative bacteria load present in the midgut.

Culturing of bacteria from the mid gut was used for comparative analysis between the isolates and the different components obtained from MCR- ALS data, but also to identify potential interactions between isolated bacteria through a competition experiment including two of the isolates. Additional biochemical tests were used to identify physiological properties of the two most interesting bacteria isolates.

2.0 Materials and methods

The description of buffers, working solutions, broths and agars used in this chapter is listed in Appendix 2. Primer and probe sequences (Invitrogen™ Life Technologies™, USA) and thermo-cycling protocols for PCRs and amplicon optimization are listed in Appendix 3, and all commercial agents are listed in Appendix 1

2.1 Experimental setup

For the culture independent part of this study, two datasets named; Seasonal stability and Diet, are included, to address the subgoals in this thesis. The Diet dataset includes two sets with bees (UMB and Arizona), which angles different aspects of nutritional intake, and figure 2.1 gives a simplified outline of the experimental setup. For the culture dependent analyses, the same bee colonies, as for the seasonal stability set were used, and bacteria were cultured and isolated from the midguts for further analysis.

In addition, bees were separately sampled to address potential technical variation in the methodology, referred to as the stability test and the extraction control.

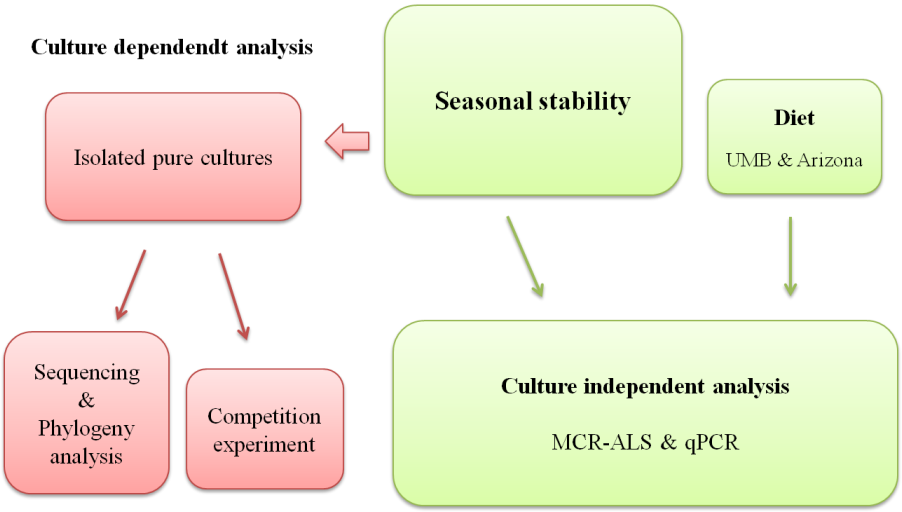

Figure 2.1: Simplified outline of the experimental setup in this thesis. The datasets; Seasonal stability and Diet, make up the culture independent part and is marked in green. The culture dependent part originates from the bee colonies included in the Seasonal stability set and the workflow is marked in red.

All methods of this section is presented Appendix 4, which gives a schematic outline of the workflow from bee sampling to finished sequence analysis and statistics, and this work flow ensured the same technical conditions for all sample sets.

2.2 Bee sampling

2.2.1 Seasonal stability

The bees for the Seasonal stability set were collected from three different hives, named 2, 4 and 6, placed at one location at Norwegian University of Life science (UMB) Ås, Norway. The bees from theses hives represent three free foraging colonies, thus represent wild type (WT) bees in traditional beekeeping in Norway. The Seasonal stability set bees were collected from the bottom box in each hive, to ensure that a representative from all bee stages would be present in the dataset, and the bees were picked with tweezers from three of the ten removable frames; no 3, 6 and 9, where the 9th frame is the one nearest the opening of the box (Fig.2.2a). All of the bees from the three colonies were sampled together in a wooden box in where they mixed by flying around (fig.2.2b)

For the September and October measurements each hive were sampled separately in three different boxes, where each box represent only one colony. This was done to address the colony variation if present.

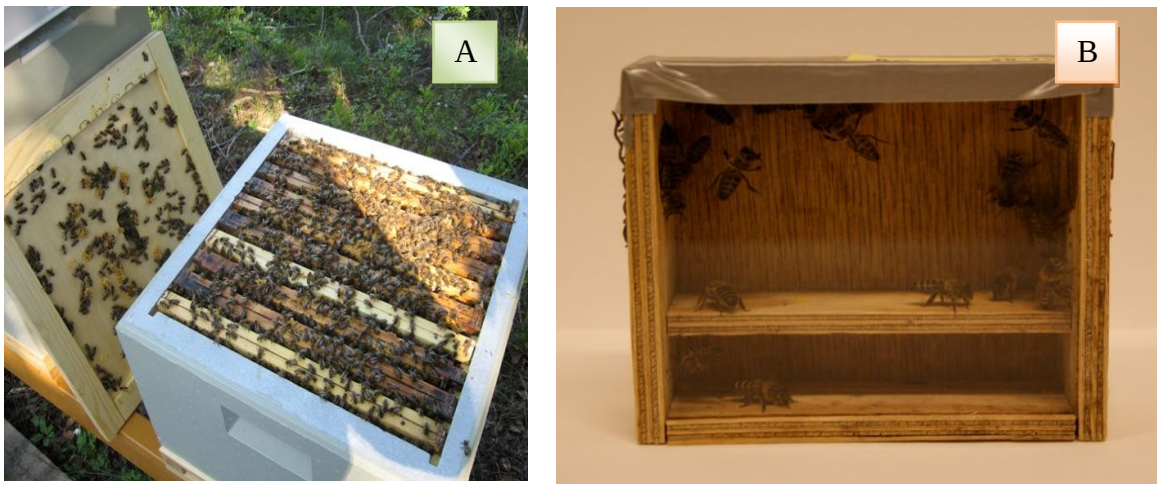

Figure 2.2: Bottom box shoving 10 removable frames (A), and a wooden box for bee sampling in which the bees could wander and fly around (B). A is reprinted from http://www.royken.kommune.no/Tjenester/Oppvekst/Ungdomsskoler/Slemmestad-ungdomsskole/Fokusomrader/Alternativ-skole/ (accessed 24.3-13). Photo B was taken by Anbjørg Rangberg

30 bees were included from each month, from May until October 2012, to address seasonal stability in the midgut composition as shown in table 2.1. Number of bees collected from each hive in September and October are also listed in this table. We collected some more bees than

we needed to include, because we expected some disruption of the mid gut during the gut removal. The total number of bees included in the Seasonal stability dataset was 180.

Table 2.1: Number of bees included from each month and from each colony in the Seasonal stability datasets used in this study. It also lists the time of sampling and number of bees sampled for the extraction control and the culture dependent study, and the kind of nutrition the different bee colonies utilized at the time sampled. Total numbers of bees included in the set is listed. Timepoint for formic acid is marked with *.

Nutrition	Sample Timepoint	Number of bees
Honey	**May**	30
	June	30
	July	30
	August	30
Sugar 1 week	***September** col 2	10
	***September** col 4	10
	***September** col 6	10
Sugar 1 mnd	**October** col 2	10
	October col 4	10
	October col 6	10
Total		**180**

Nutritional variation

The bees were not exposed to the same type of nutrition throughout the Seasonal stability experiment sampling, because when the honey was collected in September, the bees were fed sugar as a substitute for honey (Table 1). They were continuously fed sugar and in October they had utilized this nutrition for about one month. The sugar composition is listed in the appendix 1.

Formic acid treatment

In September the hives were treated with formic acid to kill *Varroa* parasites present on the bees. This was done by the UMB beekeeper for three days in a row with a cotton cloth soaked with formic acid inside the hive. This procedure creates a damp pressure inside the hive which makes the parasites fall off. When the amount of formic acid is measured and controlled at the appropriate temperature, just the right pressure will form so that this will not harm the bees http://www.norges-birokterlag.no/sykdominfo.cfm?pArticleId=13978 (accessed 6.3-13).

2.2.2 Diet dataset

The two set of bees for the Diet dataset were collected by others, the UMB beekeeper and a technician in Arizona respectively, but by the same procedure as previously described.

UMB set

The bee colonies used for the UMB set were not the same as the one for the Seasonal stability dataset. Number of bees sampled, when the bees were sampled, and the nutritional state of each colony for the UMB set, is shown in table 2.2. Colony A was a WT colony living on honey reserves. Colony B and C are the same colonies living on fed sugar sampled at different time points. Total number of bees included in the UMB set was 80.

Table 2.2: Number of bees included from each colony in the Diet set. Also listing when they were sampled and nutrition utilization at the time sampled. Total numbers of bees included in the set is listed.

Time sampled	Nutrition utilization	Name of colony	Number of bees
November 2011	WT	**A**	24
November 2011	Sugar for 1 month	**B** col 1	8
		B col 2	8
		B col 3	8
February 2012	Sugar for 4 months	**C** col 2	15
		C col 3	15
Total			**78**

Arizona set

The Arizona set includes bees collected in Arizona, which were genetically bred to have different nutritional preferences. Table 2.3 lists the number of bees included from each of the colonies of genetically different bred bees. 30 bees were included from each of the different genotypes, hence wild type bees (WT), low pollen bees (LP) and high pollen bees (HP). Low pollen means that the bees favor nectar, and high pollen means that they favor pollen as their main nutrition, but do feed on both during foraging. The bees were all sampled at one time in February 2011, and the total number of bees included in the Arizona set was 90. In Arizona during February the bees are foraging.

Table 2.3: Number of bees included from each colony in the Arizona set, and their genetically nutritional preferences are listed. Total numbers of bees included in the set is shown.

Genetical nutritional pref.	Colony name	Number of bees
WT	WT col O-8	10
	WT col P-11	10
	WT col O-7	10
LP	LP col B62	10
	LP col G55	10
	LP col G76	10
HP	HP col Y88	10
	HP col W56	10
	HP col R48	10
Total		**90**

2.2.3 Bacteria isolation

For the culture dependent analysis 30 bees were included, and they were sampled in July 2012 from the same three colonies and frames as described earlier (fig 2.1).

2.3 Gut removal

The bees were anesthetized on ice before the guts were dissected (Amdam et al. 2004b) (Fig 2.3a). To be able to isolate the midgut we used a method of gut dissecting which allowed us to remove the whole digestive system from the bee in one operation (Fig. 2.3b). This efficient technique was thought us by a senior technician, and it insured a sterile gut dissection.

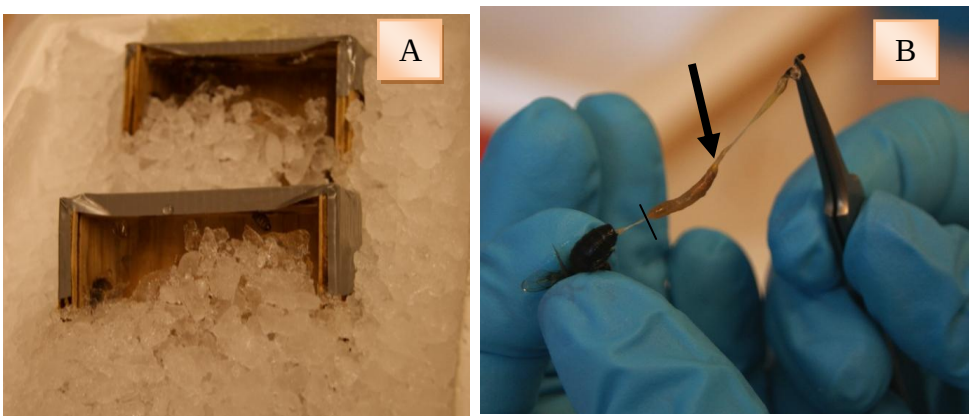

Figure 2.3: Midgut removal of honey bees. (A) Bees on ice prior to gut removal. (B) Gut removal by sterile technique. The honey bee was hold by the head as the whole gut was pulled out by the sting using sterile forceps. The midgut was collected by cutting in the transition between the midgut and the ileum (arrow). The crop got separated from the midtgut during pulling (at the line). Photos taken by Anbjørg Rangberg

Midgut isolation procedure:

The bees were put on ice directly after sampling and kept for about 1-1 1/2h. The bee was washed in 50% ethanol (EtOH) before dissecting. Then the whole gut was dissected out using a sterile dissecting forceps, and the sting was pulled out as the bee was held by the head (Fig 9b). The midgut got separated from the honey crop by itself during this procedure, and the midgut was collected into 2 ml micro tubes (Sarsted, Germany), by cutting with a sterile dissecting scissor, in the transition between the midgut and the ileum (fig 9b).

2.4 Genotypic analyzes

The culture independent part of this thesis focuses on the 16s rRNA gene presented in every bacteria, and follows a generalized approach for culture independent studies previously described in the introduction chapter (fig 1.3). A part of the 16S rRNA gene (amplicon) was used for the identification of bacteria in the midgut of bees, but also for relative quantification of the bacteria present, in all three sample sets.

The methodology used in this part of the work is well established for varies stool samples, but it was for the first time tried out on bee midgut during this thesis. The methodology optimizations include; Stability testing of storage buffer, primer testing and optimization of primer conditions for PCR and qPCR, and optimization of DNA dilutions for both qPCR and mixed sequencing.

2.4.1 Sample preparation, cell lysis and DNA extraction

Sample preparation

DNA will easily degrade if it is kept out in the open at room temperature, and therefore it is important to use preservation solution for storing the DNA before it is to be analyzed. We conducted a stability test on the S.T.A.R buffer (Roche, Germany), which was used for transport, processing and storage of the bee gut samples. We wanted to test if the gut composition was stable or changing accordingly to if the midgut stored in the buffer was frozen at once after collecting or kept in the buffer for 3 days at room temperature (RT). To test for this was essential before we analyzed the Arizona set, which arrived after being shipped at RT for three days. We chose this buffer because it is particular made for stool sample transport and DNA recovery (Espy et al. 2006).

Preparation protocol:

The micro tubes (Sarsted, Germany) were filled with approximately 0,2g < 106 um glass beads - acid washed (Sigma-Aldrich, Germany), and 500ul S.T.A.R buffer (Roche, Germany) before the mid gut was added. Some of the samples were processed the same day as gut dissection was preformed, but some samples was frozen at -20 °C for later analyzing.

Cell lysis

Mechanical lysis (bead beating) was performed on midguts and bacteria isolates, in S.T.A.R buffer (Roche, Germany), on the MagNA Lyzer (Roche, Germany). Frozen midguts were thawed on ice and lysed in the micro tubes they were sampled in, using the bead beating protocol; 6500rpm for 20 sec; 1 min cooling at 4°C; 6500rpm for 20 sec. For the bacteria isolates, each bacterium was swabbed from the blood plate using a sterile loop, and dissolved in the buffer for lysis at 6000rpm for 20sec.

Stability test: From 10 samples of lysed midguts, 150ul of the lysate were distributed into 20 new micro tubes (Sarsted, Germany), after centrifugation at 1300rmp for 5min. Representing the same sample; 10 tubes were stored in RT for 3 days before freezed at -20°C, and 10 tubes were frozen directly at -20°C.

DNA extraction

Due to the large number of samples, we used an automated magnetic particle approach, on all of our samples sets and bacteria isolates, for DNA extraction, which has previously been described by Skanseng et al. (2006). This approach uses silica particles which bind the DNA in the combination with high salt concentrations. Through a series of wash buffer steps, using lower salt concentrations, the DNA is purified and eventually released from the particles and into the elution buffer.

The Seasonal stability dataset were extracted in three turns; May & June, July & August, and September & October. The Diet set, Arizona set, and bacteria isolates were extracted in three separate turns. To address potential extraction procedure bias, an extraction control was included in each plate run, which was made to equal one midgut and a mean measurement of the three colonies, sampled in July 2012. In addition a none-template control (DNA elution buffer) was includes in each run to check for cross contamination. For each month some of the samples were run in parallel.

DNA extraction protocol:

The lyzed bacteria were spun down on a microcentrifuge at 13000rpm for 5 minutes. The sample plate was **manually** prepared with 10µl of magnetic silica particles (Chemicell (SiMAG GmbH, Gremany) in a 1:4 dilution of the genomic DNA (gDNA) and binding buffer. The three wash plates were **automatically** prepared on the epMotion 5070 pipetting robot (Eppendorf, Germany) with 1ml of each of the wash buffers 1, 2 and 3, and the elution plate with 100µl of 10nM TE buffer in 96 well KingFisher® plates. The extraction protocol was automatically performed on the KingFisher® Flex robot (ThermoScientific, USA); Incubation and mixing of sample plate for 5 min at RT, wash buffer 1 incubation and mixing for 40 sec at RT, wash buffer 2 incubation for 40 sec at RT, wash buffer 3 incubation and slow mixing for 15 sec at RT, elution buffer incubation for 1 min and 15 sec at 65 °C- repeated 8x, remove magnetic beads for 30 sec at RT. Quant-iT PicoGreen® dsDNA assay (Life Technologies™, USA) was used to confirm successful extraction.

The bee brain DNA, for vitellogenin qPCR positive control, was extracted using the DNeasy Blood and Tissue kit (Qiagen, USA), after a senior technician had sterile dissected out one brain.

Long term storage of DNA

The eluted DNA and the crushed midguts/bacteria isolates were frozen down at -20 °C in elution buffer and S.T.A.R buffer (Roche) respectively until the experiment was finished, and then at -40 °C for long term storage.

2.4.2 DNA quantification

For quantification of the DNA, the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life technologies™, USA) was used. This method measure the total double stranded DNA (dsDNA) amount in a solution, by the use of a fluorescent substance that intercalates itself in the dsDNA. Not bound to the dsDNA the PicoGreen substance emits little or no light, but bound to the dsDNA the light emission escalates a 1000 times and can be measured in a spectrofluorometer (Singer et al. 1997). The concentration of DNA in the samples can be calculated from the fluorescence measured, by using a standard curve with known concentrations.

For the bee brain DNA extraction quantification ND-1000 spectrophotometer (Nanodrop) (Thermo Scientific, USA) was used to estimate a successful extraction.

DNA quantification protocol:

5ul of the samples were added to 1X Quant-iT PicoGreen® (Life Technologies™, USA) working solution to a total volume of 100 ul in a black, 96 well nunc™ (ThermoFisher, USA) microtiter plate. The plate was read after 5 min of incubation using the KC4 software, in a FLX 800cse spectrofluorometer, with excitation at 485nm and emission at 528nm and with a sensitivity of 50.

Standard curve was made using dilutions of Lambda DNA (λ- DNA; standard 100 ng/ul) supplied together with Quant-iT PicoGreen® dsDNA kit (Life Technologies™, USA) in MilliQ purified water (Millipore Corporation, USA). Final concentrations are given in Appendix 2. To normalize the data with regards to background fluorescence, a blank sample (DNA elution buffer) was also included in the standard curve.

2.4.3 PCR

Universal 16S rRNA primers (Nadkarni et al. 2002) were used to amplify about 450 bases of each gene of the total bacterial 16S rRNA genes in all the sample sets, and the universal bacteria CoverAll™ primers (developed by Genetic Analysis; http://www.genet-analysis.com/, accessed 16.04-13) were used for the 16S rRNA gene amplification (about 1200bp) of the bacteria isolates.

PCR protocol:

Both PCR reactions were run in a 96 well PCR plate (VWR, USA) using a 2720 ThermalCycler (Applied Biosystems, USA), and the HOT FIREpol® DNA polymerase (Solis BioDyne, Estonia). Final concentration of the constituents in the working solution for one reaction was; HOT FIREpol® DNA polymerase 0,05U/µl, HOT FIREpol ® buffer B2 1X, $MgCl_2$ 2,5mM, dNTP 0,2mM, forward/reverse primer 0,2µM. Nuclease-free water (Ambion, USA) was added together with 1 µl of DNA at a total volume of 25 µl. The following thermo cycling protocol was used for the Universal 16S rRNA primers; Activation at 95 °C for 15min., Then 30 cycles of: Denaturation at 95 °C for 30sec, Annealing at 60 °C for 30sec, Elongation at 72 °C for 60sec, and a Final elongation step at 72 °C for 7 min.

For the CoverAll™ primers; difference from the above thermocycling protocol was that the annealing temperature was set to 55 °C, and the elongation time was 1min and 20sec, and no final elongation step was needed. The PCR product was stored at 4 °C for shorter time or

frozen at -20°C. A positive control = *Lactobacillus kunkeii*, and a negative control = Nuclease-free water (Ambion, USA), was added to each run.

Quant-iT PicoGreen® dsDNA assay (Life Technologies™, USA) was used to confirm successful PCR amplification. Correct size of the amplicon was confirmed using 1 % agarose gel electrophoresis stained, with ethidium bromide (Electran®) (VWR International, England), at 70V for 50 min with 100 bp DNA ladder (Solis BioDyne, Estonia) as a size marker. The agarose gel method will separate negatively charged DNA fragments according to size, using electric current, where DNA migrates from negative to positive pole through the agarose matrix. The Molecular Imager® Gel Doc™ XR Imaging System with Quantity One 1-D analysis software, v.4.6.7 (Bio-Rad laboratories, USA) was used for visualization of agarose gels.

2.4.4 Quantitative PCR (qPCR)

To be able to address the ratio between amount of bacteria DNA and bee DNA we used bee specific vitellogenin gene primers, which amplifies 150bp of this gene (Amdam et al. 2004b) and the same universal 16S rRNA primers as in the qualitative analysis. Vitellogenin is a yolk protein and is part of the reproductive organs of female bees and is therefore a good marker for the presence of bee DNA in a sample (Amdam et al. 2004b). For the vitellogenin qPCR the intercalating fluorescence substance EvaGreen was used, and the Universal 16S rRNA qPCR was performed using a TaqMan probe, which binds to the ssDNA inside the amplicon (Nadkarni et al. 2002) HRM protocol is listed in appendix 2. In addition bacteria specific primers (*Gilliamella* and *Snodgrassella*), targeting the 16S rRNA gene (Martinson et al. 2012), were used to quantify relative bacteria amount of the bacteria isolates in the competition experiment in combination with the same Universal 16S rRNA primers described above.

Vitellogenin qPCR protocol:

The PCR reaction was run in a 96 well LightCycler plate (Roche, Germany) on the LightCycler 480 (Roche, Germany), with 5X HOT FIREPol® EvaGreen qPCR Mix Plus (Solis Bio Dyne, Estonia) in final concentration of 1X. 1 µl of diluted gDNA (1:2) was added to the working solution for a final volume of 20 µl. Determination of the dilution factor for qPCR was based upon a dilution series experiment conducted prior to analysis. The following thermo cycling protocol was used; Activation at 95 °C for 15min, Then 40 cycles of:

Denaturation at 95 °C for 30sec, Annealing at 55 °C for 30sec, Elongation at 72 °C for 30sec. The fluorescence was measured each cycle at 55 °C. A HRM analysis was preformed after each run, and A pos control = Bee brain DNA and neg control = Nuclease-free water (Ambion, USA) were added to each run.

The same protocol as for the vitellogenin gene, was used for the bacteria specific qPCR, in the competition experiment, and a HRM was run to document specific amplification.

Universal 16S rRNA qPCR protocol:

The PCR reaction was run in a 96 well LightCycler plate (Roche, Germany) on the LightCycler 480 (Roche), with 5X HOT FIREPol® Probe qPCR Mix Plus (Solis BioDyne) in a 1X concentration, with 1 µl of diluted gDNA (1:2) at a final volume of 20 µl. The following thermo cycling protocol was used; Activation at 95 °C for 15min., Then 40 cycles of: Denaturation at 95 °C for 30sec, Annealing and Elongation at 60 °C for 60sec.The fluorescence was measured each cycle at 60 °C. A pos control = *L.kunkeii* and a neg control = Nuclease-free water (Ambion, USA) were added to each run.

2.4.5 Sequencing

All samples were sequenced using the mixed sequencing method described in the introduction. Four steps were performed during the sequencing procedure; PCR purification, Sanger sequencing ddNTP labeling, ddNTP removal, and mixed sequencing. In addition Paired end Illumina sequencing (250bp) was performed on the extraction control for validation of the mixed sequencing, and for retrieving a more detailed bacteria composition.

PCR purification

The PCR product was treated with exonuclease (Exo1), to get rid of excess universal 16S rRNA primers before the sequencing reaction. The Exo1 reagent, supplied in the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, USA), was used in a 1X concentration working solution, where 1µl diluted PCR product (1:2) was added to a final volume of 5µl. The reaction was run in a 96 well PCR plate (VWR, USA) in a 2720 ThermalCycler (Applied Biosystems, USA), using the following thermo cycling protocol; Exonuclease activity at 37 ºC for 60 min, Inactivation of exonuclease at 85 ºC, for 15 min: The thermo cycle was only run once, and Exo1 treated PCR product was stored at 10 ºC/ 4 ºC.

Sanger sequencing ddNTP labeling

For the midgut DNA, the labeling reaction was performed using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, USA), and run in a 96 well sequencing/pcr plate (VWR) using a 2720 ThermalCycler (Applied Biosystems, USA). A 1X Working solution was made and 1µl of exo1 treated PCR product was added to a total volume of 10µl. The following thermo cycling protocol was used; Activation at 95 ºC for 1 min, then 25 cycles of; Denaturation at 96 ºC for 15 sec, Annealing/elongation at 60 ºC for 4 min. For the bacteria isolates, the labeling reaction was run with CoverAll™ primers instead of the sequencing primer included in the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, USA). Both forward and reverse primers were run separately on all isolates, creating complimentary 3`-end labeled sequences. The labeled PCR product was stored at 10 ºC/ 4 ºC.

ddNTP Removal

The Agencourt® CleanSEQ® Dye-terminator Removal (Beckman Coulter, USA) was used to purify the end labeled sequences. Magnetic particles and a 96S Super Magnet plate (Alpaqua®, USA) was used for separating labeled ssDNA from the excess ddNTP left from the labeling reaction. This is a necessary step to ensure specific fluorescence signal in the mixed sequencing. The protocol for 10 ul BigDye® end- labeled product was manually performed using the same 96 well pcr plate (VWR), with the labeled PCR products, used in the BigDye® labeling reaction. 0,1mM EDTA (Sigma-Aldrich, Sweden) was used for elution, and 30µl of the eluate was transferred to an AB-1100 Thermo-Fast 96 PCR detection plate (Thermo scientific, USA). For more detailed description of the protocol see: http://www.agencourt.com/technical

Mixed sequencing

Mixed sequencing was performed on 30µl of the CleanSEC® elute. The sequencing reaction was performed by a technician at Hedemark University College on a 3130xl Genetic analyzer (ABI, USA).

Illumina sequencing

The extraction control sample was used for this analysis, and it was run in parallel with tagged universal 16S rRNA PRK primers (Yu et al. 2005) for initial PCR and then pooled with other

samples, after quantification and normalization, and before sequencing. The PRK primers had on forehand been tested to yield product from the bee midtgut, during the amplicon optimization. The protocol was done by a laboratory technician and the samples were sent to the University of Oslo for MiSec® sequencing (Illumina, USA).

2.4.7 Data processing

In addition to the qPCR data, sequence data from the MCR-ALS analysis, bacteria isolates and Illumina sequencing were processed. A regression analysis (correlation analysis, r^2) was done on the on the stability test parallels (done by a senior professor).

MCR-ALS analysis

MCR-ALS analysis was executed for us, yielding specific component sequences and relative ratio of the components in each sample, which we processed further by taxonomy classification of the components/bacteria, making phylogenetic trees and calculating the bacteria composition in each sample.

Taxonomy assignment

The component sequences were uploaded as fasta format and analysed in the RDP database using the Seqmacht module with default settings, which displays a hierarchy of results down to genus level. By using the; view selectable matches option, the best hits are displayed accordingly to calculated; S_ab score and unique common oligomers. S_ab score means a calculated seqmach score using the ratio between shared 7-base oligomers in the query sequence and a given RDP sequence then divided by the lowest number of 7-base oligos in either of the two sequences, and unique common oligomers means the number of uniquely occurring oligomers within a given sequence. Reference at: http://rdp.cme.msu.edu/seqmatch/seqmatch_help.jsp#resultFormat

We selected to assign the components names on the genus level where a relative high S_ab score (close to 1,0) and unique common oligomers score was calculated, and if this was not the case the component was assigned at the family level in the taxonomy hierarchy.

Phylogentic trees

To address the evolutionary distance among bacteria, sequences of the 16S rRNA- gene can be used to make phylogenetic trees (Woese 1987). Making a phylogram, using the neighbour

joining method will give an unrooted tree based on the evolutionary distance between bacteria sequences, by including internal nodules and branches with different length, which reflect each evolutionary divergence of the bacteria (Saitou & Nei 1987).

We made a phylogenetic tree out of the components and bacteria isolates for comparative analysis. The tree was made by first aligning the component sequences in BioEdit Sequence Alignment Editor (Ibis Biosciences, USA)(Hall 1999), with the use of the clustalW multiple alignment tool. The alignment was cut to the same length in the edit mode and imported into Phylogeny.fr (http://www.phylogeny.fr/version2_cgi/index.cgi; free web program distributed by Methodes et Algorithmes pour la Bio-informatique LIRMM) as fasta format (Dereeper et al. 2008). The BioNJ module was used to make a neighbor joining tree with the following settings; bootstrap: 100, substitution model: kimura 2 parameters, and transition/transversion ratio: 2 (Gascuel 1997).

Bacteria composition calculations

Excel was used for calculation of DNA concentration by PicoGreen® measurements, using the slope and intercept of the standard curve and the equation for a straight line, giving the concentration by $\mathbf{X = (fluorescent\ signal - intercept)/slope}$.

A mean of the sample parallels, in addition to a mean of all the samples in each month, was calculated, to address the relative abundance to each component in the seasonal stability set. For September and October the mean was also calculated for each colony to address colony bacteria composition divergence. In the Diet set, the mean was taken for colony A, B and C separately, and for the Genetics set a mean was calculated for WT, LP, and HP separately.

Relative ratio calculation

The Ct values were retrieved from the LighCycler 480 (Roche, Germany), and raw data was processed in excel and imported into LinReg PCR program for PCR efficiency calculation. For the qPCR data the parallel mean Ct value and the mean Ct value, for whole sets, were calculated, as were r^2 of the plate parallels. If the detection limit is the same for both runs, one can calculate the log relative bacteria/bee DNA ratio from the Ct value and slope (Kubista et al. 2006; Ståhlberg et al. 2005), using the formula described below, where bacteria values are written in red and bee values in green.

Log relative ratio Bact/Bee = Ct/Slope – Ct/Slope Slope = -1/log (efficiency)

= Ct/ (-1/ log efficiency) - Ct/ (-1/ log efficiency) ⟶

= - (Ct x log efficiency) – (- Ct x log efficiency) ⟶

Log relative ratio Bact/Bee = (log (efficiency) x Ct) - (log (efficiency) x Ct)

Bacteria isolates

The Sanger sequences were processed in the CLC Main work bench 6 program. Consensus sequences were assembled from the two complimentary sequences and identified using BLAST and RDP database. The closest matches in RDP, as described previously, and by using BLAST-hits of more than 95% match, sequence taxonomy was assigned. A neighbor joining phylogenetic tree, of all isolates, was made using the consensus, in the CLC Main work bench 6 software (CLC bio, USA), with bootstrap value of 100.

Ct-values and SEM were calculated for the competition experiment.

Illumina sequencing

The retrieved data was analyzed through a main pipeline for next generation sequence analysis, QIIME (Quantitative Insights Into Microbial Ecology) (Caporaso et al. 2010) by a computer scientist to yield result, which we imported into Excel for interpretation.

Statistical tests

To address the technical and biological variation the standard deviation (SD) and the standard error of the mean (SEM) was calculated for each component, both for the different time points and for colonies, in all datasets, by the equation;

$$\mathbf{SEM = SD/\sqrt{Number\ of\ samples}}$$

The SEM was calculated in this thesis to show how well the calculated mean values in our datasets corresponds to the true population means in the bee colonies, which was what we wanted to investigate (Barde & Barde 2012; Webster & Merry 1997).

Two sided, heteroscedastic student T-test was performed on log relative abundance and bacteria composition with significant difference set at p = 0.05.

2.5 Bacteria cultivation and pure culture isolation

30 midguts were isolated using the same technique as described under gut removal, and three micro tubes (Sarstedt, Gremany) were prepared with 500µl 1X phosphate buffered saline (PBS) and 15 % glycerol to include 10 midguts each. The tubes were kept on ice for 10-15min before they were frozen at -80°C.

2.5.1 Making plates

For Colony Forming Units (CFU) counts and bacteria isolation, tryphtose soy agar plates (TSA) (Merck KGaA, Germany) with 5% horse blood (ThermoFisher Scientific, USA) were used (Kwong & Moran 2012). The agar was autoclaved at 121°C for 15min and cooled to 40°C before adding the blood. For the competition experiment, tryptose soy broth (TSB) (Merck KGaA, Germany), without horse blood, was used (Kwong & Moran 2012).

2.5.2 Growing bacteria

The frozen guts in one micro tube were homogenized using tissue homogenizer, and a 10X dilution series was made with sterile PBS accordingly to figure 2.4. 100µl of the dilution was spread, in parallels, on TSA plates with 5% horse blood (blood agar plates), using a sterile digrasky spaltel.

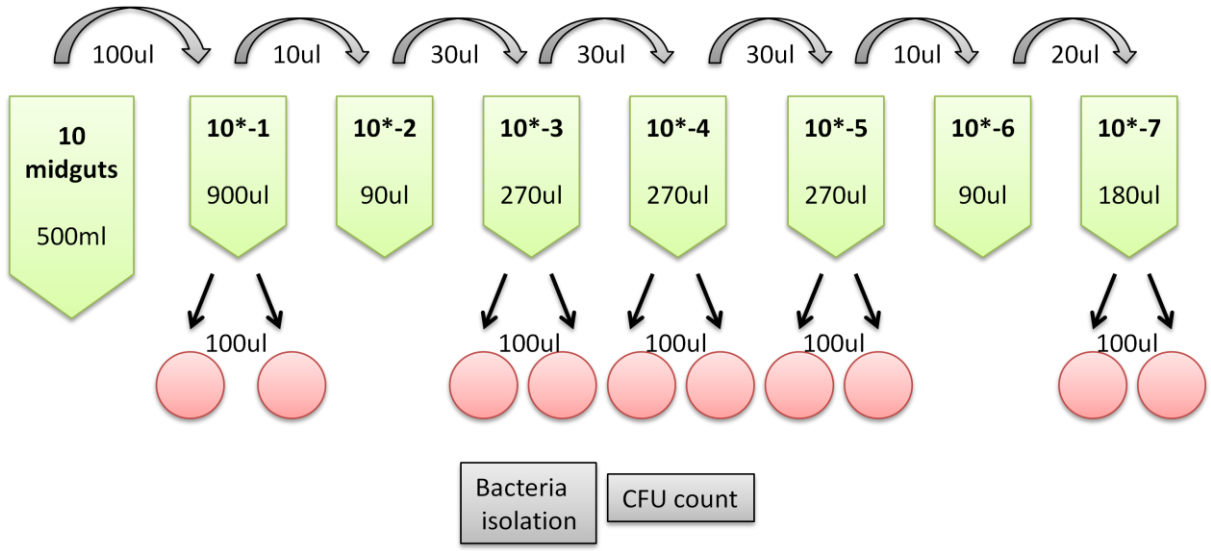

Figure 2.4: Dilution series for midgut CFU count and bacteria isolation. The different dilutions and how they were made are shown, indicated in green with corresponding grey arrows. 100µl from five of the dilutions were spread on parallel blood plates indicated by black arrows. The two black boxes underneath the blood plates show which plates which were used for picking bacteria for bacteria isolation and CFU count.

The agar plates were incubated for 2 days at 37 °C in CO_2- enriched atmosphere in a growth chamber (fig 1.6). The CO_2- enriched atmosphere was created using the GasPack™ EZ CO_2 container system (Becton Dickinson (BD), Ireland). 10% of each blood agar batches were additionally incubated for sterile controls, and as a control for correct atmosphere, a strain of *Neisseria gonhorrea* was streaked on a Thayer Martin modified agar plate and added in the growth chamber.

CFU

Total CFU/ midgut, was calculated from the parallel mean CFU plate counts, using the formula: **((Mean CFU*10 $^{(\text{dilution X})}$) * 5)/10**.

Isolation of pure cultures

Colonies were picked somewhat randomly but trying to discriminate by different morphology. The colonies were streaked on new blood agar plates to yield isolated colonies then incubated as before, repeated 2x to ensure pure cultures. Morphology and color of colony was documented.

Long term storage of bacteria

The isolates were frozen in Hart Infusion Broth (DIFCO Laboratories, USA) containing 11% glycerol (Merck KGaA, Germany) at -80 °C.

2.5.3 Physiology and biochemical tests

Gram staining

The bacteria cell wall is different amongst bacteria, and this feature can be utilized to stain bacteria, and then be able to visualize them using a microscope. A well known staining method is the Gram stain which stains Gram positive bacteria blue using crystalviolet and Gram negative bacteria are stained red through safranin, and is a much used method for differentiating bacteria in clinical diagnostics (Takenaka et al. 2012). Gram positive bacteria have a thick peptidoglycan layer, which retians the krystalviolet inside the cell wall when flushed with alcohol for decolorizaton. The Gram negative bacteria do not have this thick layer and is decolorized in this process, but colored again with the red safranin.

Two interesting bacteria isolated from the midgut were fixed on glass slides, and Gram staining was performed using the following protocol: crystalviolet for 1min, iodide for 1min,

rinse with distilled EtOH for 20-30sec, safranin for 1min. The slides were looked at using a light microscope with 100X - oil lens.

Biochemical traits

Biochemical profiles of the two most interesting bacteria was profiled using different Analytical profile index systems (API, Bio merieux® sa, France); API NH, API 20E, and API 20NE, and the tests were performed accordingly to the manufactures recommendations.

In addition we tested the two most interesting bacteria for the enzymes catalase and oxidase using ID colorCatalase (ID-ASE) (Bio merieux® sa, France) and Bactident® Oxidase (Merk KGaA, Germany) respectively. The *Neisseria gonorrhoeae* was included a positive control

Growth conditions

Due to the formic acid treatment of the bees during our sampling period we did a growth experiment, on blood agar, where a cloth with formic acid; 1ml of 60% formic acid, was added in the growth chamber, to simulate the gas conditions experienced by the bees in the hive, to investigate whether or not the produced gas could have an inhibitory effect on bacterial growth. Two bacteria were tested and parallel-spread on blood plates accordingly to figure 2.5 below, and they were incubated both with and without formic acid under the same conditions. In the control chamber we also grew the four strains on TSA agar (Kwong & Moran 2012) and in 1ml of TSB in eppendorf tubes (Eppendorf, Germany), without stirring.

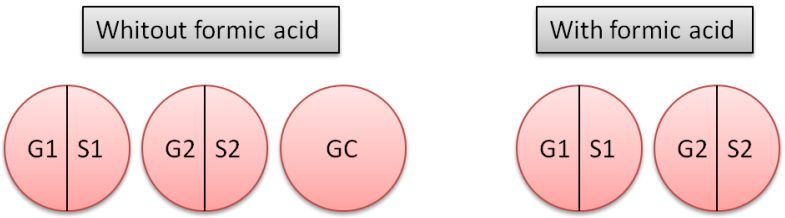

Figure 2.5: Formic acid growth experiment. Two incubation chambers were made; one with and one without formic acid. Two strains of *Gilliamella* (G1&2) and *Snodgrassella* (S1&2) were tested and spread on parallel blood agars. Both chambers incubated under same conditions and the chamber without formic acid serves as a growth control. A *N. gonnohrrea* strain (GC) was added for growth condition control.

2.5.4 Competition experiment

A competition experiment was conducted with the two strains; *Gilliamella apicola* and *Snodgrasella alvi*, isolated from the midgut, to see if they exhibited symbiotic, mutualistic or competitive growth towards eachother when grown together. The two strains were chosen after screening the isolates with strain specific primers to yield pos amplification. 1,5 ml eppendorf tubes (Eppendorf, Germany) were prepared with 0,2g < 106 um glass beads - acid washed (Sigma-Aldrich, Germany), and 100µl TSB (Merck KGaA, Germany) accordingly to figure 12. The bacteria were spread on TSA (Merck KGaA, Germany), and incubated during the whole experiment under previously described conditions. By using a sterile loop, 1µl of the bacteria was dissolved in 100µl TSB (Merck KGaA, Germany) with and without glass beads (Sigma-Aldrich, Germany), to ensure equal condition for both bacteria. Each bacteria were run in triplicates, and a negative control (only TSB) was added to the experiment. The bacteria were incubated for 1 day and then TSB (Merck KGaA, Germany) was added to a total volume of 1,5ml in each tube, for further 2 days incubation, following the work flow in figure 2.6. The bacteria were lysed and DNA was extracted for qPCR quantification of total (Universal 16S rRNA) and specific bacteria (*G. apicola* and *S. alvi*) load in each sample. The qPCR reactions were run in parallels and the products were verified on agarose gel. See section; Genotypic analyzes for protocols and primers.

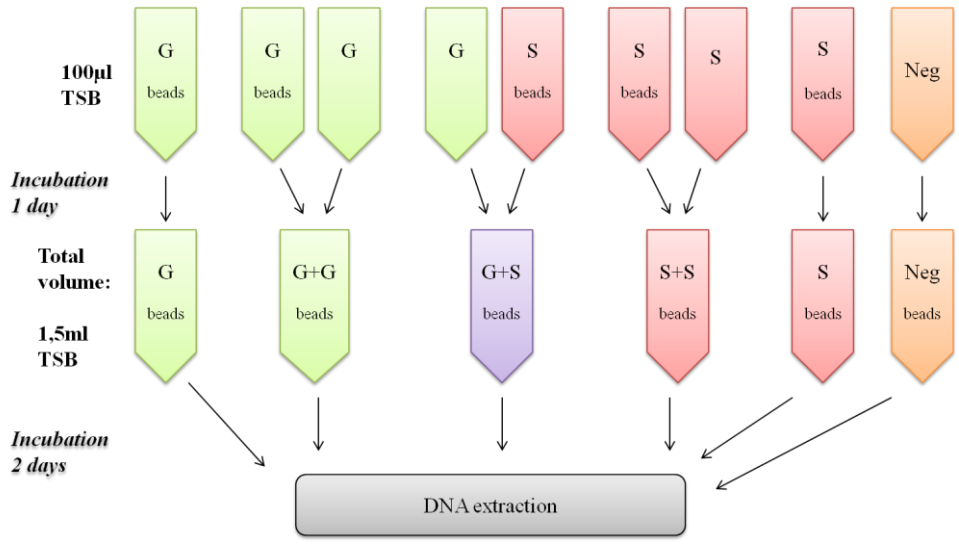

Figure 2.6: Experimental set up for one parallel in the competition experiment. The experiment was run in triplicates and qPCR quantification of total and specific bacteria load, in each sample, were determined. 1µl of bacteria were picked off the TSA agar plate and dissolved in 100 µl TSB, then incubated for one day in CO_2-enriched atmosphere at 37 °C. G = *Gilliamella*, S = *Snodgrassella*, Neg = negative contol. One single bacteria of each strain was kept separate and to strains of the same strains were joined to yield as a growth control for the joint sample with different bacteria strains. A negative control was added as sterile control. TSB was added to a total volume of 1,5ml, and incubated for 2 days, before DNA extraction was performed.

3.0 Results

3.1 Bacteria composition and relative ratio analyzes

PicoGreen® calculations of midgut gDNA concentrations spanned a range of 0,3-7,0 ng/µl. A detailed gDNA concentration range for the different sets and bacteria isolates and the quantification of PCR product using PicoGreen® measurements are listed in Appendix 5. Universal 16S rRNA gene PCR yielded satisfactory amplification of right size amplicon.

Component sequences of about 180 bp were retrieved from the mixed sequencing and MCR-ALS analysis, which were used for taxonomy assignment of the components. The MCR-ALS analysis excluded some of the samples which did not produce adequate signals and in reference to total midgut samples included in the different sets, the total coverage of sequences that fulfilled the filtering criteria of the MCR-ALS analysis, ranged from 33-100%. Mixed sequencing technical details, overview of which samples that were included in each set, on which SEM calculations are based, and a list of all the components with their assigned names are shown in Appendix 5.

Both the Universal 16S and vitellogenin gene qPCR products were the right size and the amplification was sufficient for ratio analysis. Quantification of bee brain DNA yield gave satisfying amount of DNA to be used as the positive control in the vitellogenin qPCR analysis. Plate to plate variation and the correlation regression analysis were varying, but good enough to calculate a mean colony ratio. The HRM showed one peak with no primer-dimer amplification, and the validations of the qPCR analyzes are shown in Appendix 5.

3.1.1 Seasonal stability

Four components were generated from mixed sequencing and the MCR-ALS analysis, and the assigned names were; *Pasteurella bacterium, Gilliamella apicola, Enterobactreiaceae, Snodgrassella alvi*. These four bacteria and their relative abundance make up the dominating bacteria composition in the bee midgut. The seasonal trend of the mean bacteria composition is shown in figure 3.1a, measured from May until October, and p-values calculations show coherent p-value trends between adjacent time points (table 3.3). The *G. apicola* has high dominance early on in the season but the relative amount declines from May until August ($p<0.05$). From August until October no significant difference in the relative amount of this

bacterium was calculated, but the relative amount had a total low in October. *Pasteurella bacterium* show relative amount significant difference ($p<0.05$) from July until September with a dominance peak in August, and the *Enterobacteriaceae* show a similar trend from July until September ($p<0.05$), but with the dominance peak in September, which corresponds with when the start of sugar-feeding was administrated. For the *S. alvi* the relative amount is low for the first three months and then it drops down in August to a total low, for then again to rise to higher amount in the end of the season ($p<0.05$), giving a peak in October where it dominates the composition. These data show that the midtgut bacteria composition is not constant throughout the season from May until October.

Table 3.3: Calculated p-values between different time points, for each of the four dominating components in the **Seasonal stability dataset**. Significant different values $p<0.05$ is marked in red, and coherent trends of significant differences could be detected between adjacent time points.

T-test:	*Pasteurellaceae*	*Gilliamella apicola*	*Enterobacteriaceae*	*Snodgrassella alvi*
May - June	0,381	0,046	0,205	0,938
June - July	0,125	0,009	0,959	0,169
July - Aug	0,010	0,017	0,004	0,001
Aug - Sept	0,001	0,959	0,038	0,052
Sept - Oct	0,515	0,075	0,012	0,013

The log relative bacteria/bee ratio, in the seasonal stability dataset, also show trends of declining and rising mean relative bacteria load throughout the season from May until October (fig 3.1b). The amount of bacteria was highest in May and peaked again in September, equivalent to sugar administration, and a total low was seen in October. Significant difference ($p<0.01$) between May and June measurements was calculated with variance of 0.6 log units from May until July, and similar p-value was retrieved between August and September and between September and October, displaying seasonal variation of relative amount of bacteria from May until October.

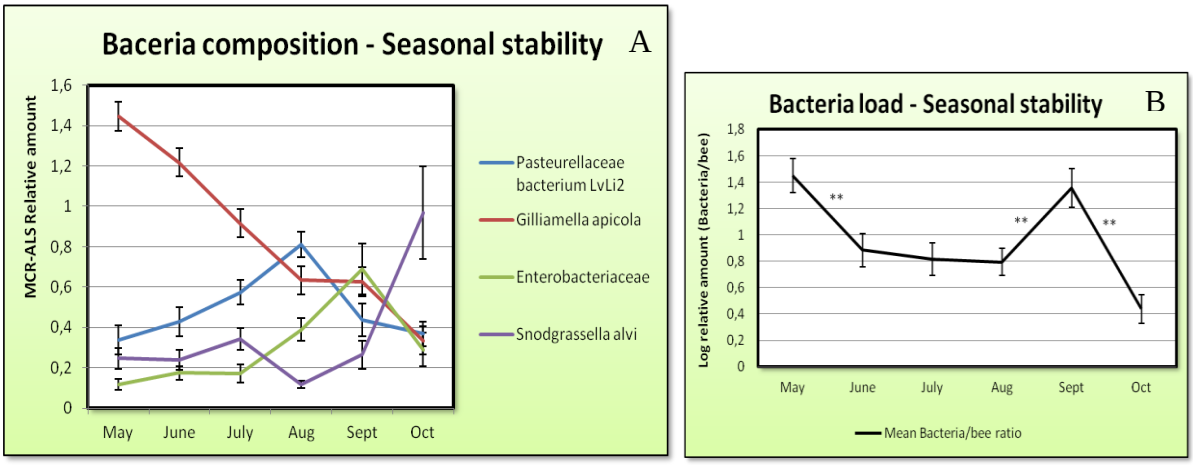

Figure 3.1: Seasonal trends of the midgut microbiota. A) Mean bacteria composition and relative ratio of bacteria, throughout a season, from May until October. The four dominating bacteria in the midgut are shown with corresponding error bars of calculated SEM for each bacterium. B) Mean log relative bacteria/ bee ratio for the season from May until October, showing error bars of calculated SEM for each month and significant difference between months is shown with **= p< 0.01

Colony variation

Colony variation was addressed through separate sampling of 10 bees from colonies 2, 4, and 6, in both September and October, and the MCR-ALS analysis included for colony 2, 4, and 6; 10, 8, and 7, and 3, 3, and 4 bees respectively for September and October. Some distinct characteristics of the bacteria composition and relative bacteria ratio, in the September and October colonies were detected from the data displayed in figure 3.2a. The *Varroa destructor* treatment showed that the bees in September were highly influenced by the *Varroa destructor* parasite with high counts of killed individuals after treatment. Colony 6 had the least parasites (around 400 individuals) and colony 4 the most, yielding over 1500 individuals.

When comparing the September colonies, colony 2 and 6 resemble in overall bacteria ratio composition and they are different from colony 4. The *Enterobacteriaceae* had a low ratio in colony 4 and a high of 0,8 and 1,0 in colony 2 and 6 respectively, which gave a significant difference between these colonies ($p<0.05$). For the October colonies a similar colony resemblance can be seen between colony 2 and 6, and between colony 4 and 6 a significant difference in the *Enterobacteriaceae* ($p<0.05$) and the *S.alvi* ($p<0.01$) was calculated, where *S.alvi* had a much higher abundance in colony 2 and 6 than in colony 4. Between the same colony measured in September and October, number 2 and 6 show the greatest change in bacteria composition, with opposite dominance of the *Enterobacteriaceae* and *S.alvi*, which contributed to the profound change, both with $p<0.01$. In colony 6, a ratio change of 1,2 units

was detected for *S.alvi*, which was then the main contributor to the *S.alvi* ratio raise from September to October in addition to colony 2.

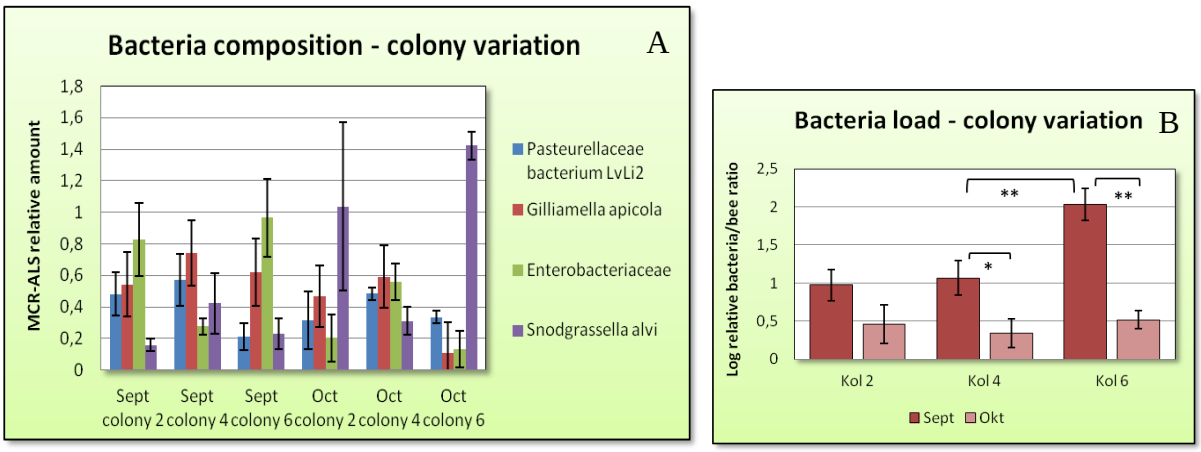

Figure 3.2: Colony comparison of September and October samples. A) Mean bacteria composition and relative ratio of bacteria for the three colonies 2, 4 and 6 included in the September and October sampling. The four dominating bacteria in the midgut are shown with corresponding error bars of calculated SEM for each bacterium. B) Mean log relative bacteria/ bee ratio for the three September and October colonies, showing error bars of calculated SEM for each month and significant difference between months is shown with *= $p<0.05$ and **= $p< 0.01$

The relative bacteria amount for the September and October colonies is shown in fig 3.2b. Internal September variation could be detected between colony 2, 4 and 6 ($p<0.01$), where colony 6 is 1.0 log unit higher than colony 2 and 4. The relative bacteria amount was in October constantly low with no internal variation between the colonies. Looking at the September to October variation, a significant difference was found both in colony 4 and 6 with $p<0.05$ and $p<0.01$ respectively, where colony 6 had the most profound change in relative bacteria amount from 2,0 log units to 0.5.

3.2.3 Diet impact on midtgut microbiota

To investigate diet impact on midgut microbiota both UMB and Arizona sets were included in this work. Both sets represent bee colonies with different nutritional intake, but the UMB set is distinct with regards to difference in diet, because the bees were forced fed sugar nutrition, whereas the Arizona set distinguish between wild type bees nutritional habits, due to genetic selection of traits in the bees when bred.

Five bacteria were retrieved from the MCR-ALS analysis for the UMB set and their composition is shown in figure 3.2a, and the calculated p-values between the colonies are listed in table 3.3. Between colonies B and C, which were sugar fed, no significant difference was detected, and the total mean composition was quite similar, but there were significant

differences in mean microbiota composition between the colonies with different diets; Colony A (honey) and colony B (sugar 1 month). *Azetobacter* had a low abundance in colony A, but significant higher in colony B and C (sugar 4 months), and *G.apicola* clearly dominated the bacteria composition in colony B and C. The most profound finding was the *Rhizobiales*, which had dominance in colony A, but was the least abundant bacteria in B and C, indicating high preference to honey diet.

Table 3.3: Calculated p-values between different colonies, for each of the five dominating components, in the **UMB set**. Significant different values of p<0.05 are marked in red.

T-test	*Azetobacter*	*Rihzobiales*	*Snodgrassella alvi*	*Gilliamella apicola*	*Lactobacilllus*
A - B	0,011	0,001	0,827	0,001	0,561
B - C	0,451	0,365	0,278	0,539	0,101

The UMB sets mean relative amount of bacteria varied between colony A and B/C with ≈ 0.6 units, but no significant difference was seen between colony B and C (fig 3.3b).

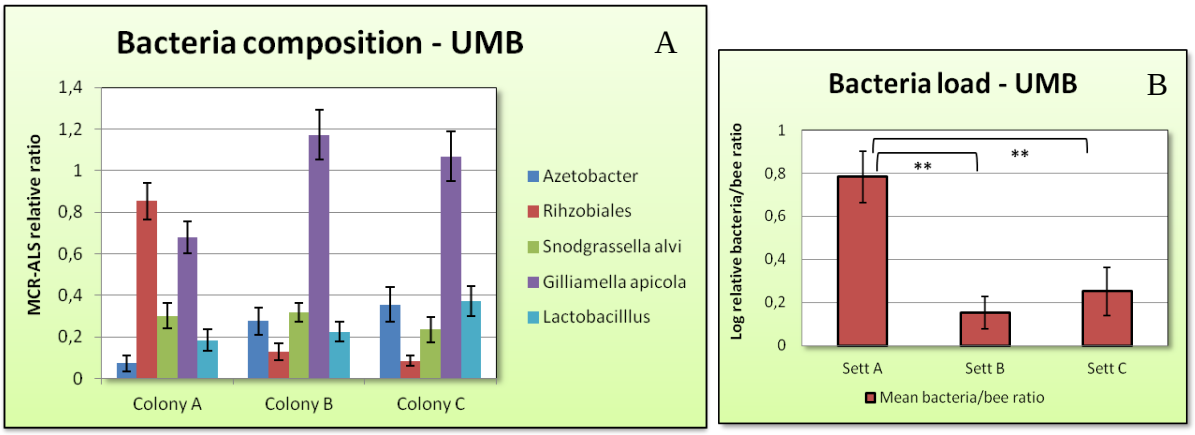

Figure 3.3: Forced fed impact on midgut microbiota. A) Mean bacteria composition and relative ratio of bacteria, between two different colonies A and B/C, where B and C are the same colony fed suger for 1 month and 4 months respectively. The five dominating bacteria in the midgut are shown with corresponding error bars of calculated SEM for each bacterium. B) Mean log relative bacteria/ bee ratio for the same colonies, showing error bars of calculated SEM for each colony and significant difference between colonies is shown with **= p< 0.01

The five components in the Arizona set stands out from the other two sets in that three different strains of *G.apicola* is present by the most domination bacteria (fig 3.4a). The different groups of bacteria, consists, after MCR-ALS analysis, of bees from 3 colonies in the WT and LP groups and from 2 colonies in the HP group. Looking at the mean group bacteria composition, little variance could be detected, where only *Lactobacillus* and *Gilliamella* 5

show significant difference between the WT and the LP groups (table 3.4). The *Gilliamella* 1 was not found significant different between HP and the two other goups, but a clear raise in relative ratio is seen in the HP group, due to one colony with high dominance by this bacterium (data not shown).

Table 3.4: Calculated p-values between different time points, for each of the five dominating components in the **Arizona set**. Significant different values p<0.05 is marked in red.

T-test	*Gilliamella 1*	*Gilliamella 2*	*Lactobacillus*	*Enterobacteriaceae*	*Gilliamella 5*
WT - LP	0,89965853	0,29611374	0,00686368	0,17152431	0,0288061
LP - HP	0,16435939	0,87569239	0,18974737	0,08777818	0,19163965
WT - HP	0,16090104	0,36142466	0,7963279	0,37486506	0,92665565

A variance between LP and HP of > 0,4 units (p< 0.01) was detected in the mean relative bacteria amount between Hp and LP (fig 3.4b), showing presence of less bacteria in the midgut of WT bees with nutritional preference for pollen grains than for nectar.

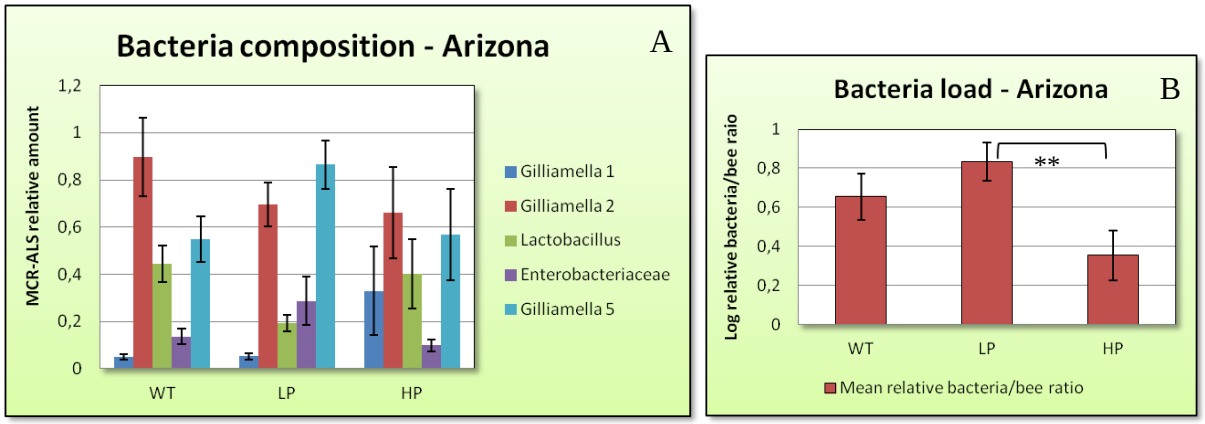

Figure 3.4: Genetic selected bred WT bees nutritional preference`s impact on midgut microbiota. A) Mean bacteria composition and relative ratio of bacteria of three different groups of bees. The five dominating bacteria in the midgut are shown with corresponding error bars of calculated SEM for each bacterium. B) Mean log relative bacteria/ bee ratio for the same groups, showing error bars of calculated SEM for each group and significant difference between groups is shown with **= p< 0.01

When relative colony composition and amount of bacteria were addressed in the Arizona set, similar results as for the September and October colonies were retrieved. Some of the three colonies within each type of bees; WT, LP, HP, differed from the next in respect to both variables, but also some of the colonies were similar and showed no significant difference (data not shown).

3.2 Isolation and characterization of midgut bacteria

3.2.1 Bacteria isolates

The bacteria from midguts grown on blood agar were picked from during day 2-5 of incubation and 29 bacteria were spread to yield pure cultures. The isolated bacteria were sequenced and the two complementary sequences had almost perfect overlapping abilities, generating a consensus sequence of about 1100-1200 bp. The assigned names, from using the generated consensus sequence and BLAST- matches, are shown in Appendix 5.

Out of the 29 isolates 9 were found genetically different (fig 3.5), and the flattening of the curve at the end of the sampling show that no new bacteria isolates were found after 20 picks, and adequately sampling of the midgut bacteria, which were capable to grow using previously described conditions, was achieved.

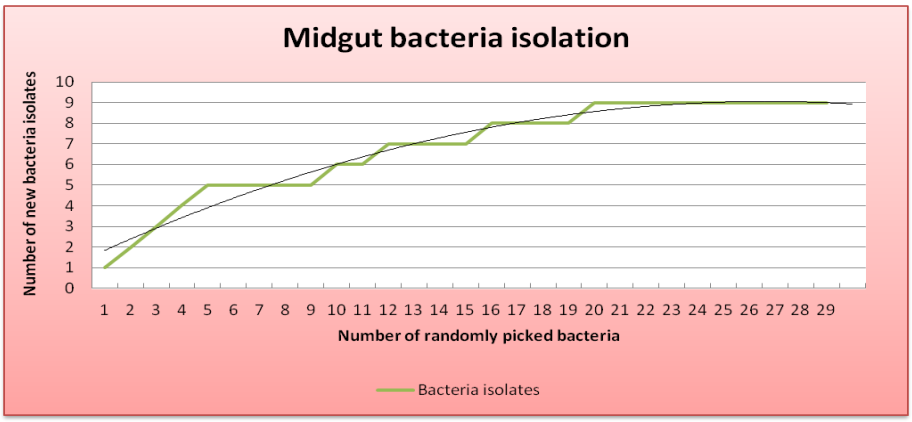

Figure 3.5: Number of genetically different bacteria found by randomly picking bacteria from blood agars.

3.2.2 Phylogenetic characterization of bacteria isolates

For phylogenetic characterization a neighbor joining phylogenetic tree including the bacteria isolates was made using the consensus sequence, which gave for the comparison, a long read alignment of the 16S rRNA gene (fig 3.6).

The isolates assigned the same name clustered together, but some slight genetic variations could be detected in the *G.apicola* cluster and the *Bifidobacterium asteroids* cluster. The *Pasteurellaceae bacterium* and the *Enterobacteriaceae bacterium* clustered at the same branch showing high genetic similarity. This was also experienced when the names were assigned and it was difficult to distinguish between these two bacteria using BLAST and RDP

databases as both these bacteria matched equally and interchangeable. In addition the *Gilliamella apicola* cluster also branched from this same node, showing low genetical divergence between these two clusters, as was also the experience using BLAST. The *S.alvi* cluster showed high similarity to the *Neiserriaceae* family through the low phylogeny distance to the growth control *Neisseria gonorrhoeae* bacterium included in the tree.

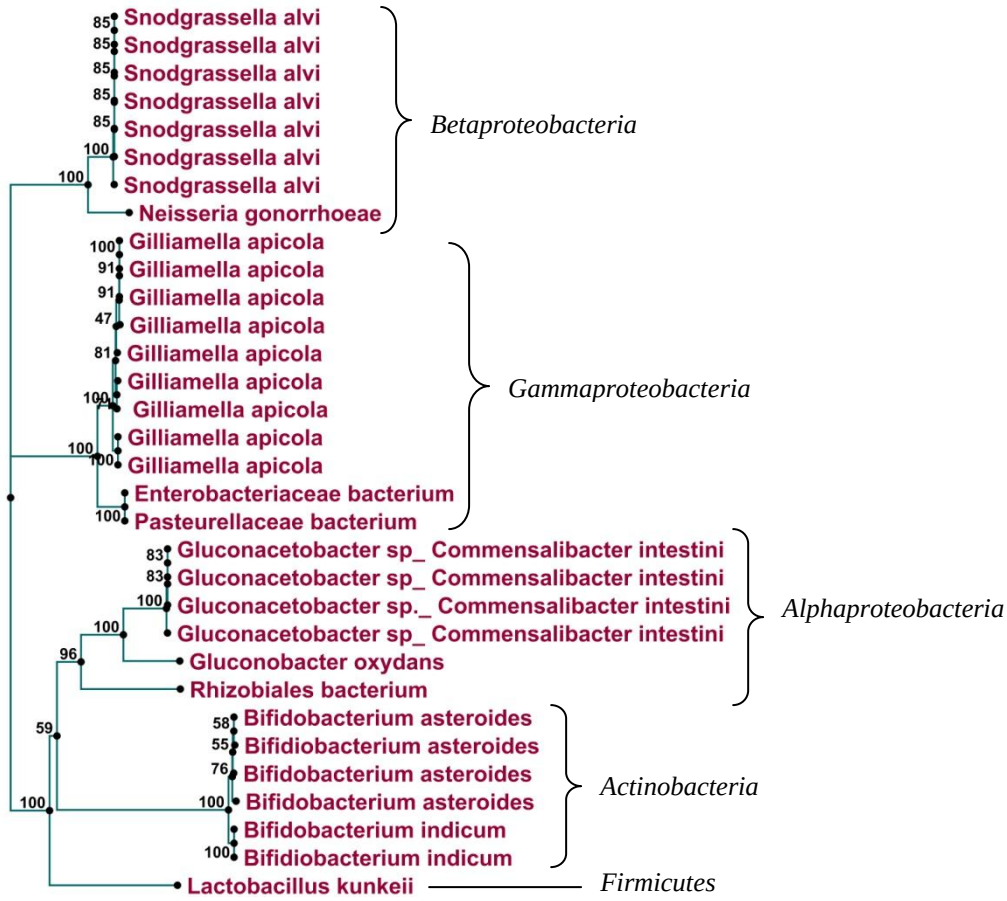

Figure 3.6: Neighbor joining phylogenetic tree made including all 29 bacteria isolates, and the growth control *Neisseria gonnorrhoeae* (for taxonomy comparison), and bootstrap values are shown in %. Brackets include bacteria with taxonomy classification at the class level written in italics, and nine of the isolates had distinct taxonomy assigned at the genus level.

3.2.3 Phylogenetic comparison of the bacteria isolates and the components

Since the components are retrieved from mixed sequencing it is kept in mind that divergence can result from impure sequences, this also applies for the bacteria isolates, regarding pure cultures. A phylogram made out of 180 bp sequences of the components and the nine distinct bacteria isolates are shown in figure 3.7.

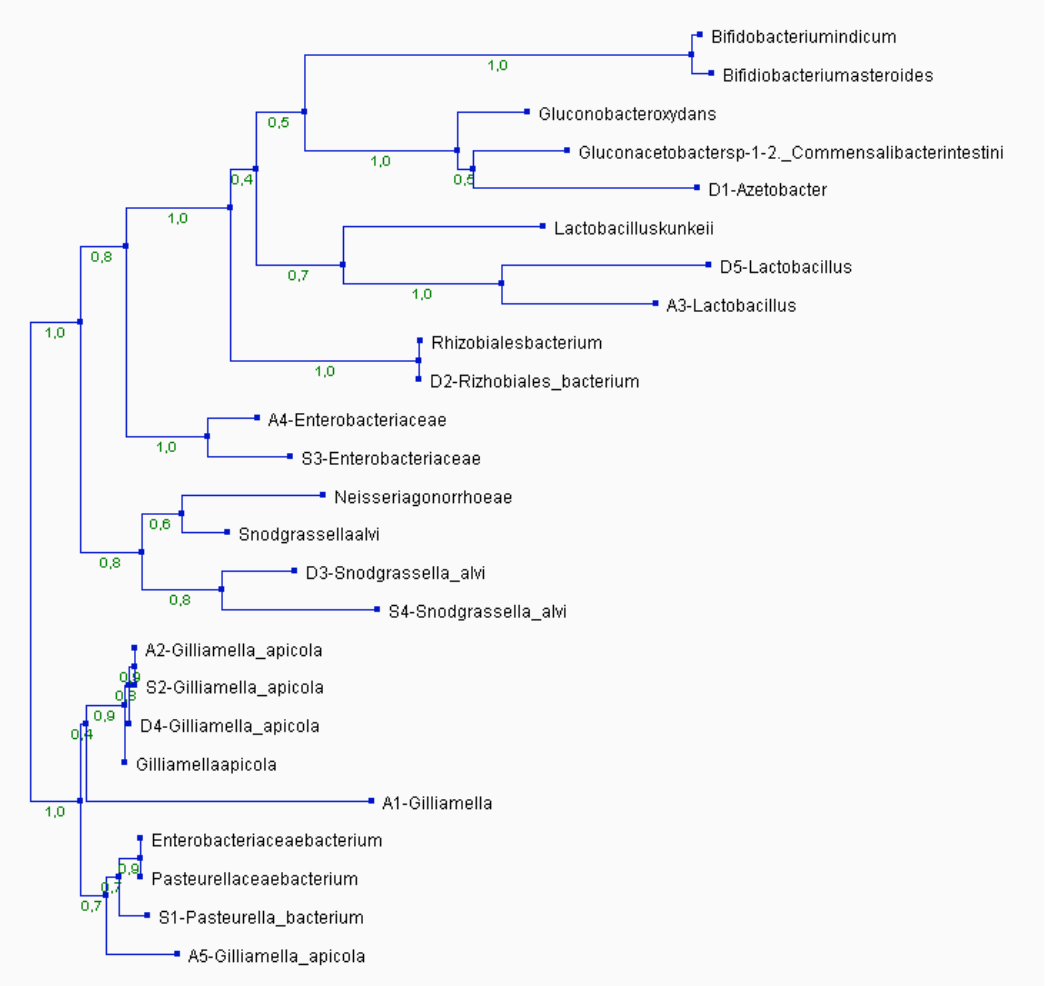

Figure 3.7: Neighbor joining phylogram of the components and the nine bacteria isolates including the growth control strain *Neisseria gonorrhoeae*, bootstrap values in % are shown. The components are annotated with component number and from which set they were retrieved from; S = Seasonal stability, D = Diet, A = Arizona. The bacteria isolates are written without any annotations.

Some of the components are found in more than one set; The *Gilliamella apicola* was found in all three sets, *Snodgrassella alvi* was present in the UMB and Seasonal stability set, but not in the Arizona set as one of the dominating bacteria. The *Lactobacillus* is present in the UMB and Arizona sets, whereas the Seasonal stability dataset and Arizona set, seems to include bacteria in the *Enterobacteriaceae* family as one of the dominating midgut bacteria. This family included genus of *Erwinia*, *Enterobacter* and *Serratia*, which all scored equal in the RDP database, thus could not be distinguished from each other. Interesting is that the UMB set includes two components; *Azetobacter* and *Rhizobiales bacterium*, which were only found in this set.

The components, with taxonomy assignment referring to the same bacteria, clustered together, even though they were not retrieved from the same set, showing genetic similarity both

between different time points and geography. All though similarity can be seen on the genus level for the *Lactobacillus* and *G.apicola*, the *S.alvi* seems to differ even on the spp. level. The *Gilliamella apicola* cluster is split into three groups, where the Arizona *G.apicola* (component 2) seems to be the same as the two *G.apicola* retrieved from the UMB and the Seasonal stability sets, whereas the Arizona *G.apicola* (component 1) seems to be genetic divergent from this group. The third Arizona *G.apicola* (component 5), which cluster together with *Pasteurellaceae bacterium* in the Seasonal stability set, underlining a close genetic resemblance within these two bacteria, something which also was indicated during the taxonomy assignment of the components using RDP.

The midgut bacteria, which were isolated and cultured, clustered with all the components found in the Norwegian bees (Seasonal stability and UMB sets), and this links these two sets together, because the Cultured bacteria were isolated from bees collected in July from the same colonies as for the Seasonal stability set. The *Lactobacillus* components did not cluster with the *Lactobacillus kunkeii* strain isolated, showing that the components were not *L.kunkeii*. The *S.alvi* strain also did not resemble the two *S.alvi* components to such a degree as the other isolates and their components. In addition to the different components, two bacteria isolates of *Bifidobacteria* were identified, which belong in the *Actinobacteria* phyla. The *Gluconacetobacter sp_ Commensalibacter intestini* clustered with the *Azetobacter* component, displaying high genetical resemblance, and where the *Azetobacter* also belongs to the *Alphaproteobacteria* class. The Rhizobiales component clustered with the Rizobiales bacterium isolated. The interesting with the clustering of these bacteria is that is shows that these bacteria are present in the honeybee midgut in July (2012) without being amongst the dominating bacteria, but dominating in October and February (2011 & 2012) during different diets.

3.2.4 Phenotypic characterization of *G. apicola* and *S. alvi*

From the 29 bacteria isolates two of the most abundant once were chosen for additional phenotypic characterization. These were the *G. apicola* and *S.alvi*, which both composed 23,3% of total bacteria isolated.

Pure cultures

The two strains were tested to grow on blood agar and TSA, and they grew well on both plates (fig 3.8a and b). The *G.apicola* colony was medium large (1mm) after 2 days of

incubation and had a yellow to brown color on blood plates. The *S.alvi* colony was larger (2mm) on blood plates and stuck to the agar, and had a pink to grey color, and when the colony was picked up, onto a loop, a more yellow color was apparent. When grown on TSA agar the *S.alvi* grew down into the agar and was difficult to remove. Both bacteria were colorless on TSA agar.

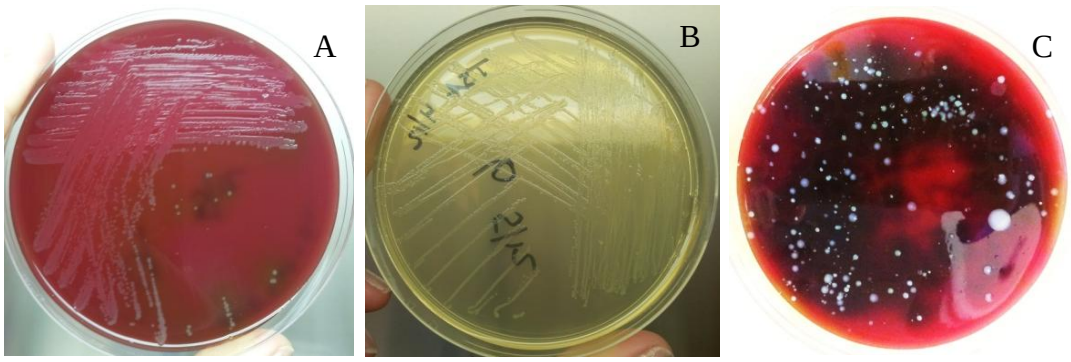

Figure 3.8: *S.alvi* pure culture on blood agar (a) and *G.apicola* pure culture on TSA agar (b). (c) Midgut blood agar spread plate (the10^{-4} dilution) for CFU counts. The calculated bacteria amount was $1,4*10^{6}$ CFU/midgut.

Gram staining

Gram staining of pure cultures of *G.apicola* and *S.alvi* showed two Gram negative rods, where *G.apicola* was medium large and slightly curved (fig 3.9a), and *S.alvi* was relative small and more coccoid in shape and some cells were arranged in pairs (fig 3.9b).

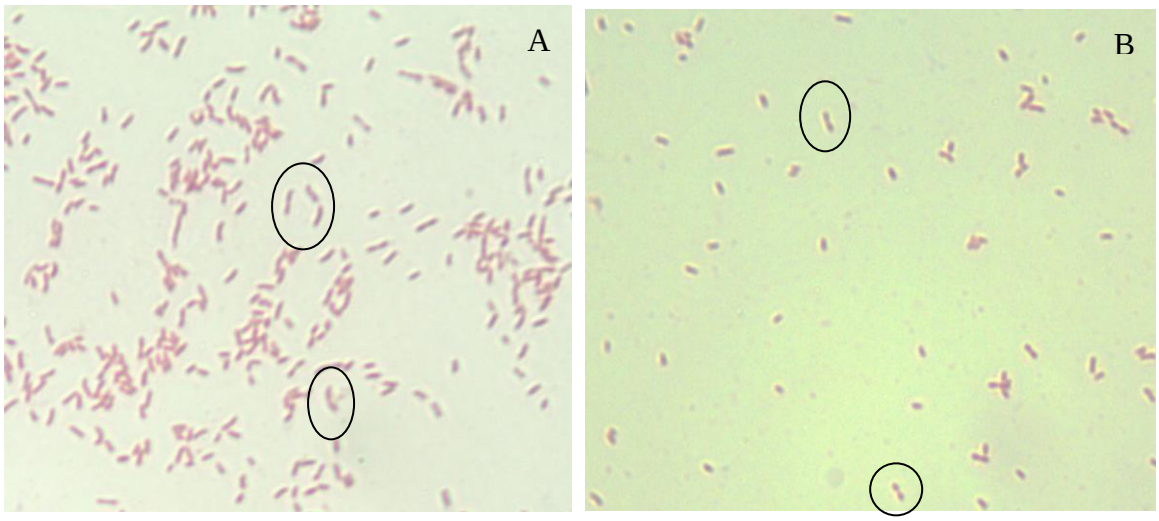

Figure 3.9: Gram staining of *G.apicola* (a) and *S.alvi* (b), both Gram negative rods. Both viewed using 100X oil lens. Circles showing curved rod bacteria (a) and coccoid rod shapes in pairs (b).

Oxydase and catalase ezymes

The two strains were tested in parallel on two different strains of the same bacteria, and the *G.apicola* tested negative for both oxydase and catalase, while the *S.alvi* tested positive for katalase and negative for oxidase.

Furic acid growth experiment

There was no significant reduction detected in growth after grown with formic acid to growth without formic acid, after two days, for the two bacteria.

API profile

The API NH system was chosen to be used for the ***S.alvi*** bacteria because of the genetic similarity to the *Neisseriaceae* family, and the following reaction gave **positive results**; **urease**, ProA, and gamma glutamyl transferase. The *G.apicola* was also tested using the API NH yielding these positive results; **glucose** and fructose (acidification), alcaline phosfatase, and Gamma glutamyl transferase. API 20 E was also tried on both bacteria, with these positive reactions; ***S.alvi*****: urease**, ***G.apicola:*** **glucose** (fermentation- oxidation) and l-Arabinose. The API 20 NE system did not give any results for either bacterium.

3.2.5 Competition experiment

The two isolates were first grown in eppendorf tubes with TSB for a try out before the actually growth experiment was conducted, and this showed that the *S.alvi* only grew in the bottom of the tube as a defined mass of cells, whereas the *G.apicola* filled the entire tube. The results showing the verification, of correct specific bacteria qPCR amplicons, on agarose gel and HRM for both primer pars, are given in Appendix 5. The technical parallel correlation was good (r^2 = 0,967), and a Ct –value mean was calculated both for the sample parallels and the sample triplicates for Universal 16S rRNA gene and specific bacteria. The Universal 16S rRNA qPCR gave no difference in Ct-value regardless of different starting cultures or if bacteria were grown alone or together, and a mean Ct-value of 17.9 was calculated, which show that both bacteria had grown into a steady state at the end of this experiment. In figure 3.10 the specific qPCR Ct-values are shown, and displaying when grown alone, both as 100 µl and 200 µl starting culture, both the *G.apicola* and *S.alvi* grew to a final bacteria load, with mean Ct-values of 15.7 and 17.5 cycles respectively. This correaponds to the result from the Universal 16S qPCR run. When grown together the *G.apicola* growth was not influenced by

the *S.alvi* bacterium, but the *S.alvi* showed less growth, which is displayed by a raised Ct-value of 20.0 cycles when comparing the bacteria growth of the *S.alvi* when grown together.

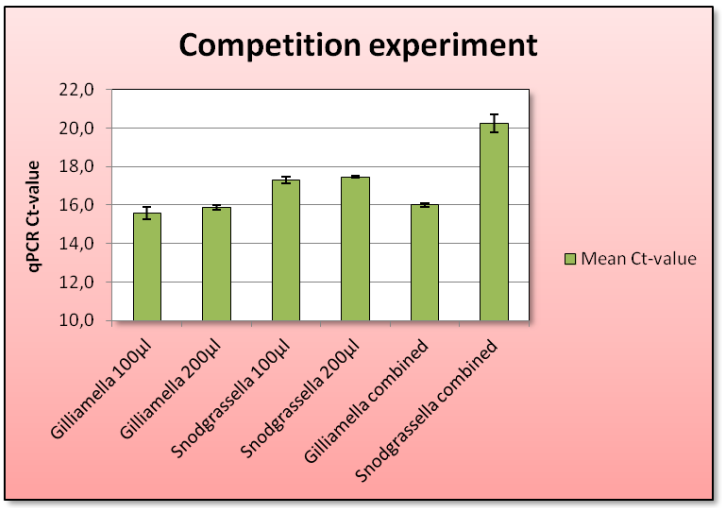

Figure 3.10: Competition experiment with *G.apicola* and *S.alvi* showing the calculated mean Ct- values of biological parallels. The 100µl and 200 µl annotation describe the starting concentration before the final incubation, see fig 2.6, and the combined annotation describes when the two bacteria were grown together. Error bars show calculated SEM.

3.3 Methodology optimization

The methodology optimization involved a stability test on the S.T.A.R buffer, and amplicon selection from different PCR and qPCR primers.

3.3.1 Stability test

Extraction of gDNA of samples incubated for 3 days at RT and then stored at -20 °C and samples stored directly at -20 °C were performed, and was confirmed on agarose gel. The results show high density DNA in the upper part of the gel of both RT and -20 °C samples, but also a band smear around 500 bp for the RT samples (fig 3.11a). The gDNA was run again on agarose gel after treated with10 µg/ml RNase (Sigma Chemicals, Germany) at 37 °C for 30min, using same conditions as before for, and no visible band smear on the RT samples around 500bp could be detected (fig 3.11b), confirming them to be degraded RNA and not DNA.

RT **-20 °C**

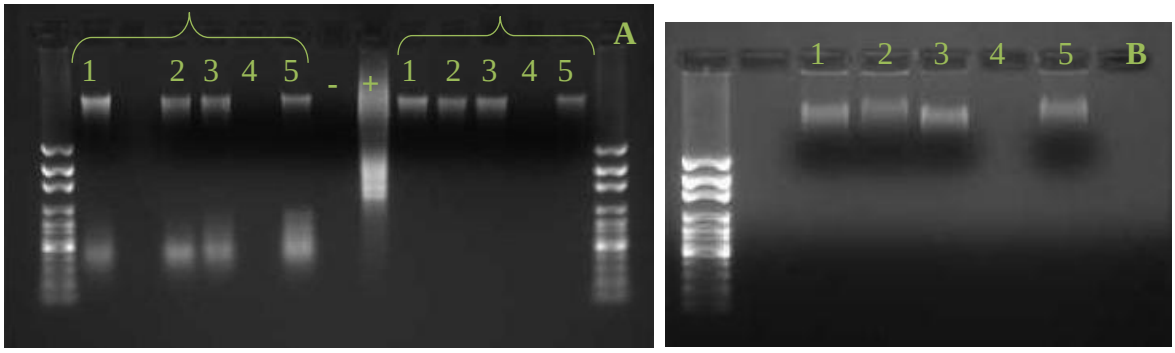

Figure 3.11: Confirmation by RNase treatment of gDNA from the stability test on agarose gels using 100 bp ladder as size marker. 15 µl of gDNA was added on 1% agarose gel, and run for 35 min on 75V. (a) Five samples, both form RT and -20 °C were run, and a neg and pos control (*L.kunkeii*) was added on to the gel. High density DNA and a band smear (around 400 bp) could be seen. Pos control show normal gDNA smear. (b) After RNase treatment the band smear around 400 bp could not be seen, showing it to be degraded RNA, and not DNA.

The stability test sequencing result was analyzed by others and the correlation results are calculated from comparisons between nucleotide peaks, at each position, in sequences derived from mixed sequencing. The results show good correlation between midgut microbiota gDNA stored for 3 days in RT and midgut microbiota gDNA stored at -20 °C after lysis, therby not showing any notable degradation of the sample stored at room temperature for three days (Knut Rudi personal communication). This could then be applied to the Arizona samples.

3.3.2 Universal 16S rRNA amplicon selection

Three universal primer pars were tested; Universal 16S rRNA, PRK, and CoverAll®, and thermo cycling protocols as recommended in literature were used. For the three primers, the working solutions were prepared with the same final concentrations of the constituents. The PCRs were performed on midgut samples with varying gDNA concentrations and a pos control (*L.kunkeii*) and neg control (Nuclease-free water) were added, and the amount of amplicon and its size was detected on agarose gels (fig 3.12). The Universal 16S rRNA primers showed strong amplification of the correct size (466 bp), in all samples, but also amplification of the negative control and bee DNA specificity control, which was verified as bacterium contamination by Sanger sequencing. The CoverAll® primers showed strong amplification in all samples by two bands, one of correct size (1200 bp), consisting of both bacteria 16S and bee DNA 23S, also verified by Sanger sequencing. The PRK primers had amplification of the samples with high gDNA concentration, also with correct amplicon size (466 bp), but failed to amplify some of the samples with low gDNA concentration. No unspecific amplification in negative control or bee DNA control was detected. From these results the Universal 16S rRNA-gene primers were selected to be used in this thesis.

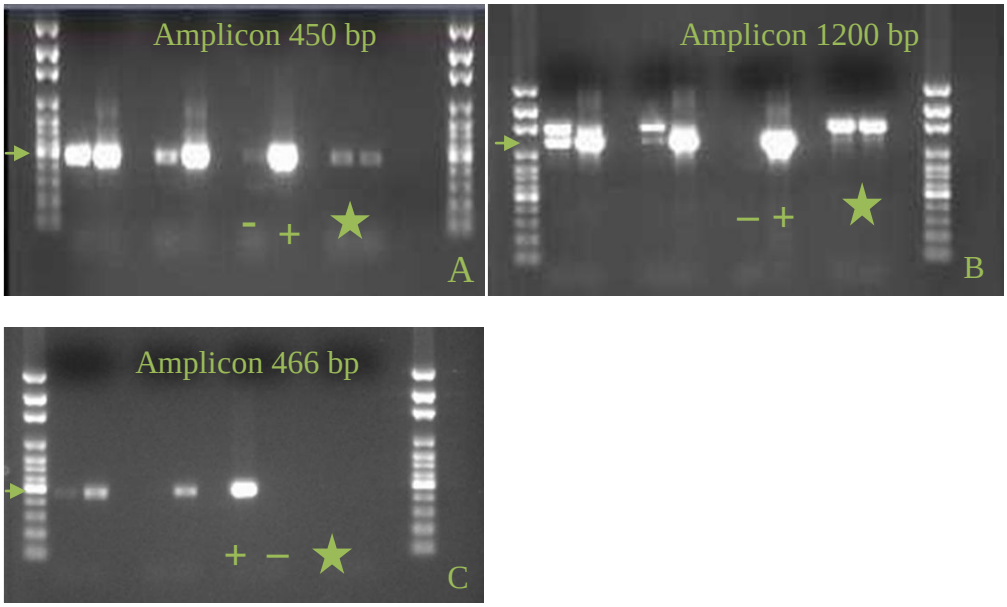

Figure 3.12: Amplicon validation of 5 µl PCR product on 1% agarose gels using 100bp ladder as a size merker. Three primer pars were run; Universal 16S rRNA (a), CoverAll® (b), and PRK (c) on the same samples with slightly varying gDNA concentrations. Arrows mark point at expected amplicon size for the primer pars. Bee brain DNA was added (x2) as specificity control and is marked with a star. Positive (*L.kunkeii*) and negative controls are shown. (a) Strong amplification in all samples and the weak band in the negative control and bee DNA samples was bacterium contamination (b) Strong amplification in all samples, except pos control, of two bands consisting of both bacteria 16S and bee DNA 23S. (c) Amplification of the samples with high gDNA concentration and no amplification of samples with low gDNA concentrations, but no unspecific amplification in negative and bee DNA controls. The unspecific bands were verified by sequencing.

3.3.3 Bee specific primer selection

Two bee specific primer pars were tried out to be used in the relative bacteria/bee ratio experiment; Actin (Amplicon;149 bp) and Vitellogenin (Amplicon;150 bp) (Amdam et al. 2004b). Both PCRs were done with the same working solution and protocol and the gel pictures show strong amplification of correct size for both primers, on all samples including pos control when run on 1,5% agarose gel for 30 min at 75V (fig 3.13). No amplification in the neg control and bacteria specificity control (*L.kunkeii*) was seen for the vitellogenin primers, but for the actin primers a faint band was seen in the specificity control. The vitellogenin primes were selected to be used as the bee specific gene for the qPCR relative bacteria/bee ratio experiment.

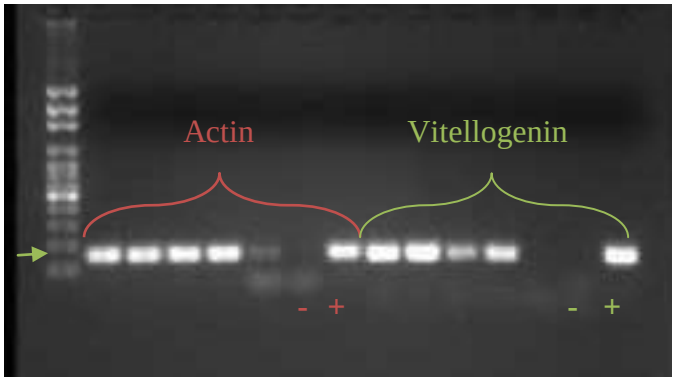

Figure 3.13: Conformation of bee specific primer PCR on agarose gel using 100 bp ladder as size marker. The samples tested (from left to right) were; bee brain1&2, midgut 1&2, and L.kunkeii (specificity control). Arrow marks the correct amplicon size. The actin primers showed strong amplification on all relevant samples, but also a faint band in the specificity control. The vitellogenin primers had strong amplification of all samples and showed high specificity.

3.4 Technical variation

The plate variation was addressed by correlation calculations on the main samples, both for the qPCR analysis and the MCR-ALS analysis, and detailed data are shown in Appendix 5.

3.5.1 Extraction control

The p-values calculated ideally subjects to no other variations in the data analysis then the biological variation, and the overall technical variation for both bacteria composition and relative bacteria/bee ratio was addressed through the extraction control. The extraction control was measured as parallels on each plate and this was repeated three times resulting in six measurements. The parallels gave rise to some plate variation, both for the MCR-ALS analysis (fig 3.14a) and the qPCR analyis (fig 3.14b). The *G. apicola* showed a mean difference of 0,3 units from may until August, which is the biggest difference seen amongst the bacteria in the set. The extraction control showed that the plate to plate (monthly) variance was lower than the significant biological variance, and that opposite trends are apparent compared to the sample results (fig 3.1), underlining that technical variation in the experiment was well within the biological variation.

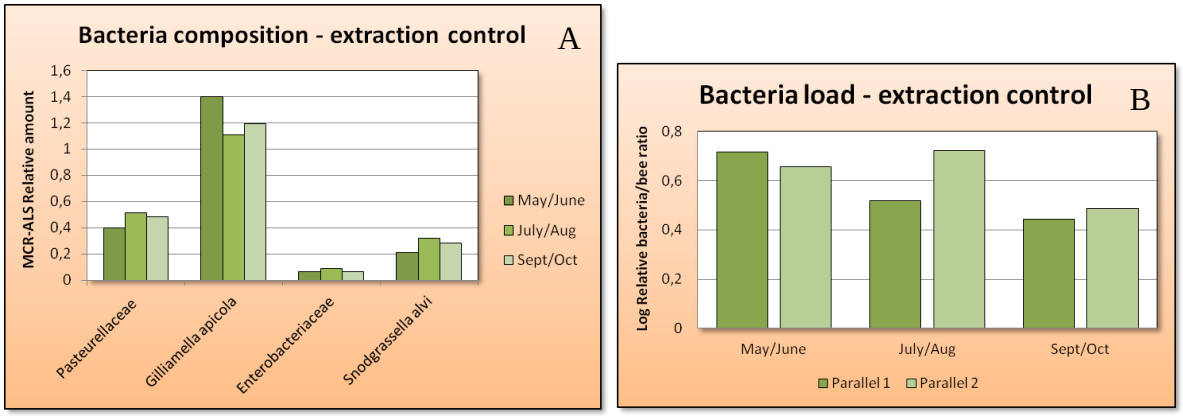

Figure 3.14: Extraction control results from MCR-ALS and qPCR analysis. A) Mean relative ratio from plate parallels are shown on each bacteria. *Gilliamella apicola* showing a mean difference of 0,3 units from may until August, which is the biggest difference seen amongst the bacteria in the set. B) qPCR mean log relative ratio results from plate parallels are shown, where the mean difference spans 0,2 units from May until October.

3.5.2 Illumina data

Illumina sequencing of the extraction control gave a total of 188189 reads, and data processing generated taxonomy classification of the identified bacteria down to class level. A low diversity bacteria composition was detected (eight classes) and only four bacteria classes comprised more than 1% of the total bacteria load. The four most abundant classes or OTUs were *Gamma-*, and *Beta- proteobacteria*, *Bacilli*, and *Alphaproteobacteria* in decending order of relative ratio, which were; 0,79; 0,09, 0,05, and 0,045, respectively. The *Gammaproteobacteria* singled out as the main bacteria taxa with about 80% dominance.

4.0 Discussion

The seasonal stability in midgut microbiota and the diet impact on midgut microbiota in honeybees has never before been investigated, which makes the work in this thesis relevant when gut microbiota is being determined, and when effects of geography or diet on the bee gut microbiota composition are addressed.

4.1 Bacteria composition and relative bacteria load

4.1.1 Seasonal stability

The main question in this thesis, addressed if the midgut microbiota would be stable throughout a season. According to this, it was assumed that comparison between bees with varying geography, like done in other studies (Jeyaprakash et al. 2003; Martinson et al. 2011; Moran et al. 2012), would give answers about factors shaping the midgut microbiota in honey bees.

Surprisingly, the four components in the seasonal stability dataset showed apparent trends in bacteria composition and relative ratio throughout the whole season. The seasonal trends can be explained through a range of factors, some of which are more obvious, like varying seasonal diet and infections, and some of which has not yet been considered. The diet impact will be discusses in accordance to the mean colony variation in the UMB and Arizona sets. Colony variation was investigated through separate sampling of bees from all three colonies in September and October. Colony variations were detected, and the variation stayed the same from September until October, where two of the colonies (number 2 and6) resembled each other in overall composition, but were distinct from the third (number 4). The two colonies with resembled each other were the once responsible for the shift in bacteria composition from September to October, where colony 4 stayed the same, and the reason for this is not clear, but naturally colony variations have been detected (Moran et al. 2012). Colony variation was also addressed in the Arizona set and colony differences among bees of the same type were detected, which could support that colony variation is naturally occurring. Either way only one of the three colonies showed altered composition in the

Enterobacteriaceae component in the Seasonal stability dataset, and by including a mean of all three colonies, the overall bacteria composition and trends would not be significantly altered. This is supported because a mean measurement of the bacteria composition at one location was guaranteed by sampling the bees from all three colonies together in one wooden box, which overcame the bee to bee and colony to colony variation at one site as seen in the study by (Moran et al. 2012).

The significant decrease of relative bacteria load seen from May until July could reflect the *G.apicola* component`s ratio trend in the composition analysis. This is consistent with the study of (Martinson et al. 2012) where *G.apicola* was found to be most abundant in the midgut amongst the dominating bee bacteria and thus will be causing most of the variance. The peak in September most likely reflect the peak of *Enterobacteriaceae*, but the reason for this is inconclusive and could resolve around both start of sugar feeding and foraging. The latter is most likely to be the reason, because the trends start before the sugar is administrated. This increase in nutrition intake could give rise to bacteria proliferation of any of the dominating bacteria, and not just one. Even though a peak in relative abundance is seen this does not always mean that the bacterium in question suddenly becomes much more abundant. This could be shown in the October measurement, where the *S.alvi* component seems to be much more abundant that the other components, but this increase in relative ratio could instead result from the decrease of the other three components. When the qPCR results are analyzed a dramatic decrease in bacteria load was seen in October, and most likely is the *S.alvi* component fairly stable throughout the Seasonal stability, but still dominating in relative ratio to the three other components, which seem to have very low abundance in October. The *S.alvi* bacterium has been found to stick to the midgut wall (Martinson et al. 2012), hence a layer of bacteria is formed and most probably a steady number of bacteria would be present, due to this niche pecificity.

The decrease in bacteria load from September to October was most prominent in colony 6, which also was the least infected with *V.destructor*, and showed highest bacteria load in September. The high number of parasites in colony 2 and 4 could have affected the total bacteria load in September (Amdam et al. 2004a; Dainat et al. 2012; Evans & Schwarz 2011), but this was not possible to address in this study.

4.1.2 Diet impact

With respect to dietary effects, the mean composition was not significant altered between the two colonies, in the UMB set, fed sugar over a period of 4 months, which points to the adjustment towards a stable microbiota when fed the same nutrient. Arguably could the September /October colonies show divergent results, because they too were fed sugar over a period of one month, but to contradict thisobservation, the bees in the Seasonal stability did not solely feed on sugar, because the weather in September was warm and bees were still foraging.

In contrast to this finding, colony A, which was living on honey reserves, displayed a unique composition with significant differences in three of the components compared to colony B and C. Arguably could this colony variation resolve to what Moran and colleagues found involving natural colony variation, but the *Rhizobiales* component was not detected as a dominating bacteria in any of the other sets, thus implying that this bacteria has high preference towards honey. No study has addressed the diet impact on bacteria composition in bees, but in humans studies can show that specific bacteria are stimulated through different diets (Hildebrandt et al. 2009; Martinez et al. 2013). Possibly could this bacterium originate from the honey itself, but due to the antiseptic properties of honey this is not likely. Also the *Azetobacter* component, which was only detected in the UMB, showed significant difference to the honey fed colony, where it displayed more abundance in the sugar fed colonies. This finding suggests more likely of a stimulating event to have occurred, than an origination from honey, which resulted in detection as a dominating component after MCR-ALS analysis in the Diet set. Both the composition and qPCR results specify that the impact on the midgut microbiota by different diets is greater than the impact of same diet for a longer period of time.

Addressing the diet impact on the WT bees harboring difference preferences to varying nutrition in, the Arizona set, it is difficult to be very conclusive, because little variance was detected through the conducted t-test, but one component; the Gilliamella-1 seems to be higher in HP than in LP and WT bees. The absence of significance could be due to large divergence in the parallels used to calculate the p-value, and that the HP samples were few in comparison with the other two sets, due to technical reasons. This component also clustered between the *G.apicola* group and the *Pasterulellaceae/Enterobacteriaceae bacterium*, which singles this bacterium out to be unique and possible specialized for high pollen diet, a

property found within the *Gilliamella* group (Engel et al. 2012). This finding coincides with the theory of niche specificity, which states that bacteria living in the gut of animals and humans have close interactions with each other, and this sharing of the same environment often results in bacteria having specialized tasks or are sustained to certain environmental conditions, due to competition for the same nutrients, utilization of different nutrients, or the ability of bacteria to tolerate different environmental factors. This finding suggests that it is the diet that shapes the microbiota and not the microbioa that controls what the host prefer to eat (Ley et al. 2006a), but the bees in this dataset has been bred to have different nutritional preferences so it could in the end be genetics controlling this microbiota (Tims et al. 2013). With respect to bacteria load, two interesting results are prominent in the Diet dataset. A high difference ($p<0,01$) in bacteria load is seen between colony A, which had the highest bacteria load and B/C in the UMB set. The *Rhizobilaes* component had the highest relative ratio in colony A, which probably represent most of the bacteria load in this colony. The honey diet of colony A contributes to more bacteria in the midgut, either by adding bacteria present in the honey, or by giving bacteria already in the midgut, better opportunities to proliferate then what sugar do. In the Arizona set a lower bacteria load is seen in the HP bees perhaps due to the fact that pollen is more difficult to digest (Davis 2004), which may need specialized bacteria to do the task, but also that bees with high pollen diet experience the toxicity of some pollen carbohydrates (Barker 1977).

Another observation done in this study is obvious when comparing the colonies which were fed on sugar. Both the October and B/C colonies have a low relative bacteria ratio, and this could represent the sterile sugar mix fed to the bees in these colonies, while foraging bees most likely will acquire bacteria from the environment in addition to a normal and variable diet. A probably cause of higher bacteria load in bees with varying diet could be that different nutrients stimulate certain bacteria, but also enhances the bacteria diversity, which are aspects investigated in humans (Martinez et al. 2013).

4.2 Taxonomy assignment

The different components retrieved from the MCR-ALS all corresponds well with dominating phylotypes identified in other studies (Martinson et al. 2011), which are found almost in every bee (Moran et al. 2012), and their identical taxonomic classification emphasize that the

components represent the dominating bee specific gut bacteria. In addition the UMB set which do not include foraging bees, also found the *G.apicola* and *S.alvi* to be two of the most dominating bacteria show that these bacteria do not come from the environment, but is restricted to the bee gut. Even though the methodology used in this thesis is different from all the other studies, which also were different from each other, the same bacteria groups are found. The isolation of the 29 midgut bacteria also results to the same eight phylotypes as described by (Martinson et al. 2011), which was easily detected analyzing the phylogenetic clustering in fig 3.7. These results in addition to previous findings, indicates that the bee microbiota is more or less consistent despite geographic distance, time or diet, and that these dominating bacteria probably are present in the bee gut for a reason, suggesting honeybee symbiosis.

When comparing the phylogeny of components and isolates it is prominent to address the high resemblance between one component of *G.apicola* and *Pasteurella bacterium/ Enterobacteriaceae bacterium*, where the same resemblance within the two isolates; *Enterobacteraceae bacterium* and *Pasteurellaceae bacterium* is also seen. In addition, these isolates resembles the *Pasteurella bacterium* component, indicating the all of these sequences could belong to the same bacteria family and even genus, but are assigned different taxonomy because of inconsistent BLAST nomenclature (Kwong & Moran 2012). The taxonomical assignment of both components and isolates was complicated due to the many almost equal matches both in RDP and BLAST, which can have lead to incorrect classification of the groups mentioned above.

The Illumina data was used as a verification of the component analysis the two methods corresponds well. The three components; *G. apicola, Pasteurellaceae* and *Enterobacteriaceae* belong to the *Gammaproteobacteria* class and *S.alvi* to the *Betaproteobacteria* class, which were found as the most dominant taxa in both analyzes.

The variation among the sequences in the possible *G.apicola* group could be due to the mixed sequencing or working with impure isolates, but the study by Moran et al. (2012) also revealed large strain variation among this group of bacteria due to recombination events, which supports the relevance of this detected variation, and this underlines the sensitivity and accuracy of the sequencing methodology used in this thesis. The same study also found large divergence between *S. alvi* sequences retrieved from bees from separate colonies and geography, something also detected in this thesis, amongst the *S.alvi* components, which were

sampled from colonies spaced in time. The *S.alvi* bacteria was not detected in the Arizona set as a dominating bacteria, probably due to the many variants of *G.apicola* found instead, and the MCR-ALS analysis only finds components above a certain detection limit. In the Seasonal stability set the *S.alvi* was also present in low abundance. The strain variation of these two bacteria could suggest adaptation to changes in the environment and in this thesis this would possibly reflect colony habitat differences because the components and isolates reflect a mean sample, and not a single bee.

The phylogeny of the components and bacteria isolates suggests that for the bacteria found across time and geography, another factor have to be of importance for the transition to become one of the dominating once in the midgut. The *Azetobacter* component shows high resemblance to the isolated bacteria *Cluconacetobacter/ commensalibacter* from July bees (Seasonal stability dataset), as do the *Rhizobiales* component and isolate, and the obvious factor differentiating these sets is diet.

4.3 Phenotypic characterization

4.3.1 API

To be able to get correct results using API systems for biochemical profiling it is important to use the right type of test, which is made for the bacteria of interest. In accordance to a study done by Koch and Schmid-Hempel (2011a) the phylogeny data suggested the *S.alvi* to be closely related to the *Neisseriaceae* family, and therefore the API NH was used to investigate the biochemical traits of the *S.alvi* and *G.apicola*. The *G.apicola* has high similarity to the *Pasteurallaceae* family, which include the genus *Haemophilus*, which the API is capable of characterizing. The study by Kwong and Moran (2012) used the API 20 NE and was not able to get any growth of the two bacteria in the medium supplied by the kit, and made their own growth medium, in which only *S.alvi* could give positive results. Coinciding with their results is the detection of urease in *S.alvi*, which is an interesting trait, found in bacteria like Helicobacter pylori, which is known to adhere to the gut wall in humans, as a pathogen causing ulcers. *S.alvi* has also been found to adhere to the midgut wall (Martinson et al. 2012), resembling the *H. pylori* in these two traits. In contrast to the study by Kwong and Moran (2012) detection of glucose and fructose acidification was found in the *G.apicola*

strain, properties which were also found in a metagenome study of bee gut microbiota by (Engel et al. 2012). In accordance to this study both glucose and arabinose fermentation was detected in *G.apicola* using API 20E in this thesis, which also verifies the API NH result.

4.3.2 Competition experiment

The competition experiment reviled less growth of *S.alvi* when grown together with *G.apicola*, than when grown alone. The results show that when competing for the same nutrients and grown under these conditions *G.apicola* will outgrow *S.alvi*. This competition could restrain *S.alvi* to grow close to the gut wall, and it could also be a result of the different growth properties of the two bacteria. As seen in this thesis, the *G.apicola* can grow throughout the whole broth, due to flagella (Engel et al. 2012), and the *S.alvi* adheres to the bottom of the tube, thus not able to compete for the nutrients.

If these results are seen in accordance to the bacteria composition data there seems to be proof to the competitive nature of these two bacteria, in the Seasonal stability set. The only time when *S.alvi* can dominate is when it is very low relative amount of the other bacteria, like in October. Due to the declining trend of *G.apicola* one could expect *S.alvi* to rise and not decline in relative amount from July to August, but as discussed above, if the *Pasteurellaceae bacterium* is just a variant of the *G.apicola*, the *S.alvi* would suffer from the competition advantage of this component also. The colony variation between October colonies shows strikingly similar results to what found, investigating the trends between time points. Colony 4 have low relative abundance of *S.alvi*, but higher of the other three components, but the other two colonies show the opposite, suggesting that *S.alvi* has less competition from the other bacteria. This competition seems to be profound also in accordance to other bacteria as well, which again the conclusion could be that the *S.alvi* is constrained to its specified niche and has no need for competing for nutrients between the other bacteria in the midgut.

4.4 Confounders

In relatedness to the Seasonal stability dataset several factors can be of influencal importance, as for instance the infection of bees with the parasite *V.destuctor*, which could affect both the composition and the relative bacteria load in the end of the season as the parasite multiplies. The effect of this parasite in the bacteria composition could be a reason for the divergence of

colony 4 in the September/October samples. Colony 4 was found to harbor the most parasites, and colony 6 the least, possible explaining that this colony had a higher relative bacteria load than the other two colonies. To give any conclusive statements on this subject more elaborator studies has to be conducted, but the *V.destructor* has previously been found to weaken bee colony`s health in CCD (Dainat et al. 2012), and infections with various pathogens can alter gut bacteria composition (Cornman et al. 2012). As a future study thematic could it be reasonable to belive that massive infections with *V.destructor* could cause weak colonies unable to swarm late in the season, thus not building up enough nutritional storage to make it through the winter? And could this also be the effect of the colony variation, where colony 4 did not forage one last time in September?

Another factor to include and not yet described through this work, is the natural bee hatching throughout the season, and how this could affect the relative bacteria load, since it is found that newly hatched bees harbor less bacteria in the midgut than do older foraging workers (Martinson et al. 2012).

4.5 Conclusive remarks

The clustering of components and isolates in this study do underline the previously finding that despite big geographic distance and paced in time, these bacteria are found in almost all bees. The relative bacteria ratio throughout a season, can be altered through different diets, hence surroundings, and the impact on the midgut microbiota by different diets is greater than the impact of same diet for a longer period of time. This work`s findings, which also include individual bacteria with nutritional preferences and possible specific niche adaptations, makes it interesting to further investigate the diet impact on midgut microbiota. The bee gut composition is simple compared to the human or other animal guts, and therefore easy to monitor. In addition, the gut is easily accessible, and bacteria isolated from the bee gut microbiota, in combination with molecular techniques, give significant information about the physiological nature of gut bacteria. Hence, all these results combined support the idea of honeybees as a well suited model for gut bacteria studies (Engel & Moran 2013).

References

Alaux, C., Dantec, C., Parrinello, H. & Le Conte, Y. (2011). Nutrigenomics in honey bees: digital gene expression analysis of pollen's nutritive effects on healthy and varroa-parasitized bees. *Bmc Genomics*, 12: 496.

Amdam, G. V., Hartfelder, K., Norberg, K., Hagen, A. & Omholt, S. W. (2004a). Altered physiology in worker honey bees (Hymenoptera: Apidae) infested with the mite Varroa destructor (Acari: Varroidae): a factor in colony loss during overwintering? *J Econ Entomol*, 97 (3): 741-7.

Amdam, G. V., Norberg, K., Fondrk, M. K. & Page, R. E., Jr. (2004b). Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proc Natl Acad Sci U S A*, 101 (31): 11350-5.

Avershina, E., Storro, O., Oien, T., Johnsen, R., Wilson, R., Egeland, T. & Rudi, K. (2013). Bifidobacterial Succession and Correlation Networks in a Large Unselected Cohort of Mothers and Their Children. *Applied and Environmental Microbiology*, 79 (2): 497-507.

Barde, M. P. & Barde, P. J. (2012). What to use to express the variability of data: Standard deviation or standard error of mean? *Perspect Clin Res*, 3 (3): 113-6.

Barker, R. J. (1977). Some carbohydrates found in pollen and pollen substitutes are toxic to honey bees. *J Nutr*, 107 (10): 1859-62.

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M. & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, 28 (3): 495-503.

Burgmann, H., Pesaro, M., Widmer, F. & Zeyer, J. (2001). A strategy for optimizing quality and quantity of DNA extracted from soil. *Journal of Microbiological Methods*, 45 (1): 7-20.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 7 (5): 335-6.

Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. (2013). Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J*, 7 (3): 592-602.

Chaimanee, V., Chantawannakul, P., Chen, Y., Evans, J. D. & Pettis, J. S. (2012). Differential expression of immune genes of adult honey bee (Apis mellifera) after inoculated by Nosema ceranae. *J Insect Physiol*, 58 (8): 1090-5.

Chen, J. & Weimer, P. J. (2001). Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. *Microbiology*, 147 (1): 21-30.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*, 37 (Database issue): D141-5.

Cornman, R. S., Tarpy, D. R., Chen, Y., Jeffreys, L., Lopez, D., Pettis, J. S., vanEngelsdorp, D. & Evans, J. D. (2012). Pathogen webs in collapsing honey bee colonies. *PLoS One*, 7 (8): e43562.

Crailsheim, K. (1988). Intestinal transport of sugars in the honeybee (apis mellifera l.). *J.Insect Physiol.*, 34 (9): 839-845.

Crailsheim, K. (1990). The protein balance of the honey bee worker. *Apidologie*, 21: 417-429.

Dainat, B., Evans, J. D., Chen, Y. P., Gauthier, L. & Neumann, P. (2012). Predictive markers of honey bee colony collapse. *PLoS One*, 7 (2): e32151.

Davis, F. C. (2004). *The honey bee inside out*. Great Britain: Bee Craft Limited.

De Filippo, C., Cavalieri, D., Di Paola, M., 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 U S A*, 107 (33): 14691-6.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res*, 36 (Web Server issue): W465-9.

Disayathanoowat, T., Young, J. P., Helgason, T. & Chantawannakul, P. (2012). T-RFLP analysis of bacterial communities in the midguts of Apis mellifera and Apis cerana honey bees in Thailand. *FEMS Microbiol Ecol*, 79 (2): 273-81.

Engel, P., Martinson, V. G. & Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci U S A*, 109 (27): 11002-7.

Engel, P. & Moran, N. A. (2013). Functional and evolutionary insights into the simple yet specific gut microbiota of the honey bee from metagenomic analysis. *Gut Microbes*, 4 (1): 60-5.

Espy, M. J., Uhl, J. R., Sloan, L. M., Buckwalter, S. P., Jones, M. F., Vetter, E. A., Yao, J. D., Wengenack, N. L., Rosenblatt, J. E., Cockerill, F. R., 3rd, et al. (2006). Real-time PCR in clinical microbiology: applications for routine laboratory testing. *Clin Microbiol Rev*, 19 (1): 165-256.

Evans, J. D. (2004). Transcriptional immune responses by honey bee larvae during invasion by the bacterial pathogen, Paenibacillus larvae. *J Invertebr Pathol*, 85 (2): 105-11.

Evans, J. D. & Armstrong, T. N. (2006). Antagonistic interactions between honey bee bacterial symbionts and implications for disease. *BMC Ecol*, 6: 4.

Evans, J. D. & Schwarz, R. S. (2011). Bees brought to their knees: microbes affecting honey bee health. *Trends Microbiol*, 19 (12): 614-20.

Fabbrini, E., Higgins, P. B., Magkos, F., Bastarrachea, R. A., Voruganti, V. S., Comuzzie, A. G., Shade, R. E., Gastaldelli, A., Horton, J. D., Omodei, D., et al. (2013). Metabolic response to high-carbohydrate and low-carbohydrate meals in a nonhuman primate model. *Am J Physiol Endocrinol Metab*, 304 (4): E444-51.

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. (2009). The bacterial species Challenge: making sense of Genetic and ecological Diversity. *Science*, 323 (6. February): 741-746.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*, 14 (7): 685-695.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Mol Ecol Resour*, 11 (5): 759-69.

Hall, T. (1999). BioEdit:a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Res*, 41: 95-98.

Hamdi, C., Balloi, A., Essanaa, J., Crotti, E., Gonella, E., Raddadi, N., Ricci, I., Boudabous, A., Borin, S., Manino, A., et al. (2011). Gut microbiome dysbiosis and honeybee health. *Journal of Applied Entomology*, 135 (7): 524-533.

Haugland, R. A., Vesper, S. J. & Wymer, L. J. (1999). Quantitative measurement of Stachybotrys chartarum conidia using real time detection of PCR products with the TaqMan(TM)fluorogenic probe system. *Mol Cell Probes*, 13 (5): 329-40.

Hildebrandt, M. A., Hoffmann, C., Sherrill-Mix, S. A., Keilbaugh, S. A., Hamady, M., Chen, Y. Y., Knight, R., Ahima, R. S., Bushman, F. & Wu, G. D. (2009). High-fat diet

determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology*, 137 (5): 1716-24 e1-2.

Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A*, 88 (16): 7276-7280.

Ihrig, J., Lill, R. & Mühlenhoff, U. (2006). Application of the DNA-specific dye EvaGreen for the routine quantification of DNA in microplates. *Analytical Biochemistry*, 359 (2): 265-267.

Jasnos, L., Sliwa, P. & Korona, R. (2005). Resolution and repeatability of phenotypic assays by automated growth curve analysis in yeast and bacteria. *Analytical Biochemistry*, 344 (1): 138-140.

Jeyaprakash, A., Hoy, M. A. & Allsopp, M. H. (2003). Bacterial diversity in worker adults of Apis mellifera capensis and Apis mellifera scutellata (Insecta: Hymenoptera) assessed using 16S rRNA sequences. *J Invertebr Pathol*, 84 (2): 96-103.

Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I. & Khorana, H. G. (1971). Studies on polynucleotides: XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. *Journal of Molecular Biology*, 56 (2): 341-361.

Koch, H. & Schmid-Hempel, P. (2011a). Bacterial communities in central European bumblebees: low diversity and high specificity. *Microb Ecol*, 62 (1): 121-33.

Koch, H. & Schmid-Hempel, P. (2011b). Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc Natl Acad Sci U S A*, 108 (48): 19288-92.

Koch, H., Abrol, D. P., Li, J. & Schmid-Hempel, P. (2013). Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. *Mol Ecol*.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoback, R., Sjogreen, B., Strombom, L., et al. (2006). The real-time polymerase chain reaction. *Mol Aspects Med*, 27 (2-3): 95-125.

Kwong, W. K. & Moran, N. A. (2012). Cultivation and characterization of the gut symbionts of honey bees and bumble bees: Snodgrassella alvi gen. nov., sp. nov., a member of the Neisseriaceae family of the Betaproteobacteria; and Gilliamella apicola gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the Enterobacteriales order of the Gammaproteobacteria. *Int J Syst Evol Microbiol*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23041637 (accessed: Oct 5).

Ley, R. E., Peterson, D. A. & Gordon, J. I. (2006a). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, 124 (4): 837-48.

Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. (2006b). Microbial ecology: human gut microbes associated with obesity. *Nature*, 444 (7122): 1022-3.

Lu, J.-J., Perng, C.-L., Lee, S.-Y. & Wan, C.-C. (2000). Use of universal primers and restriction endonuclease digestion for detection and identification of common bactrial pathogens in cerebrospinal fluid.

Mao, F., Leung, W. Y. & Xin, X. (2007). Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol*, 7: 76.

Martinez, I., Lattimer, J. M., Hubach, K. L., Case, J. A., Yang, J., Weber, C. G., Louk, J. A., Rose, D. J., Kyureghian, G., Peterson, D. A., et al. (2013). Gut microbiome composition is linked to whole grain-induced immunological improvements. *ISME J*, 7 (2): 269-80.

Martinson, V. G., Danforth, B. N., Minckley, R. L., Rueppell, O., Tingek, S. & Moran, N. A. (2011). A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol Ecol*, 20 (3): 619-28.

Martinson, V. G., Moy, J. & Moran, N. A. (2012). Establishment of characteristic gut bacteria during development of the honeybee worker. *Appl Environ Microbiol*, 78 (8): 2830-40.

Moran, N. A. (2006). Symbiosis. *Current Biology*, 16 (20): R866-R871.

Moran, N. A., Hansen, A. K., Powell, J. E. & Sabree, Z. L. (2012). Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS One*, 7 (4): e36393.

Morrison, T. B., Weis, J. J. & Wittwer, C. T. (1998). Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques*, 24 (6): 954-8, 960, 962.

Mullis, K. B. (1990). The Unusual Origin of the Polymerase Chain Reaction. *Scientific American* 262: 56-65.

Nadkarni, M. A., Martin, F. E., Jacques, N. A. & Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology*, 148 (Pt 1): 257-66.

Nyholm, S. V. & Graf, J. (2012). Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. *Nat Rev Microbiol*, 10 (12): 815-27.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol*, 5 (7): e177.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *Bmc Genomics*, 13: 341.

Rangberg, A., Diep, D. B., Rudi, K. & Amdam, G. V. (2012). Paratransgenesis: an approach to improve colony health and molecular insight in honey bees (Apis mellifera)? *Integr Comp Biol*, 52 (1): 89-99.

Ririe, K. M., Rasmussen, R. P. & Wittwer, C. T. (1997). Product Differentiation by Analysis of DNA Melting Curves during the Polymerase Chain Reaction. *Analytical Biochemistry*, 245 (2): 154-160.

Rudi, K., Zimonja, M., Trosvik, P. & Naes, T. (2007). Use of multivariate statistics for 16S rRNA gene analysis of microbial communities. *Int J Food Microbiol*, 120 (1-2): 95-9.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4 (4): 406-425.

Salonen, A., Nikkilä, J., Jalanka-Tuovinen, J., Immonen, O., Rajilić-Stojanović, M., Kekkonen, R. A., Palva, A. & de Vos, W. M. (2010). Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *Journal of Microbiological Methods*, 81 (2): 127-134.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* , 74 (12): 5463–5467.

Schwartz, S., Oren, R. & Ast, G. (2011). Detection and Removal of Biases in the Analysis of Next-Generation Sequencing Reads. *PLoS One*, 6 (1): e16685.

Sekelja, M., Rud, I., Knutsen, S. H., Denstadli, V., Westereng, B., Naes, T. & Rudi, K. (2012). Abrupt temporal fluctuations in the chicken fecal microbiota are explained by its gastrointestinal origin. *Appl Environ Microbiol*, 78 (8): 2941-8.

Shade, A., Hogan, C. S., Klimowicz, A. K., Linske, M., McManus, P. S. & Handelsman, J. (2012). Culturing captures members of the soil rare biosphere. *Environ Microbiol*, 14 (9): 2247-52.

Singer, V. L., Jones, L. J., Yue, S. T. & Haugland, R. P. (1997). Characterization of PicoGreen Reagent and Development of a Fluorescence-Based Solution Assay for Double-Stranded DNA Quantitation. *Analytical Biochemistry*, 249 (2): 228-238.

Skanseng, B., Kaldhusdal, M. & Rudi, K. (2006). Comparison of chicken gut colonisation by the pathogens Campylobacter jejuni and Clostridium perfringens by real-time quantitative PCR. *Mol Cell Probes*, 20 (5): 269-79.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H. & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, 321 (6071): 674-679.

Ståhlberg, A., Zoric, N., Åman, P. & Kubista, M. (2005). Quantitative real-time PCR for cancer detection: the lymphoma case. *Expert Review of Molecular Diagnostics*, 5 (2): 221-230.

Takenaka, Y., Takeda, K., Yoshii, T., Hashimoto, M. & Inohara, H. (2012). Gram staining for the treatment of peritonsillar abscess. *Int J Otolaryngol*, 2012: 464973.

Tims, S., Derom, C., Jonkers, D. M., Vlietinck, R., Saris, W. H., Kleerebezem, M., de Vos, W. M. & Zoetendal, E. G. (2013). Microbiota conservation and BMI signatures in adult monozygotic twins. *ISME J*, 7 (4): 707-17.

Trosvik, P., Skånseng, B., Jakobsen, K. S., Stenseth, N. C., Næs, T. & Rudi, K. (2007). Multivariate Analysis of Complex DNA Sequence Electropherograms for High-Throughput Quantitative Analysis of Mixed Microbial Populations. *Applied and Environmental Microbiology*, 73 (15): 4975-4983.

Van de Peer, Y., Chapelle, S. & R, D. W. (1996). A quantitative map of nucleotide subtitution rates in bacterial rRNA. *Nucleic Acids Res*, 24 (17): 3381-3391.

Vanengelsdorp, D. & Meixner, M. D. (2010). A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *J Invertebr Pathol*, 103 Suppl 1: S80-95.

Vasquez, A., Forsgren, E., Fries, I., Paxton, R. J., Flaberg, E., Szekely, L. & Olofsson, T. C. (2012). Symbionts as major modulators of insect health: lactic acid bacteria and honeybees. *PLoS One*, 7 (3): e33188.

Vebø, H. C., Sekelja, M., Nestestog, R., Storro, O., Johnsen, R., Øien, T. & Rudi, K. (2011). Temporal development of the infant gut microbiota in immunoglobulin E-sensitized and nonsensitized children determined by the GA-map infant array. *Clin Vaccine Immunol*, 18 (8): 1326-35.

Webster, C. S. & Merry, A. F. (1997). The standard deviation and the standard error of the mean. *Anaesthesia*, 52 (2): 183.

Willey, J. M., Sherwood, L. M. & Woolverton, C. J. (2009). *Prescott`s principles of microbiology*. New York: McGraw-Hill.

Wilson-Rich, N., Dres, S. T. & Starks, P. T. (2008). The ontogeny of immunity: Development of innate immune strength in the honey bee (Apis mellifera). *J Insect Physiol*, 54 (10–11): 1392-1399.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev*, 51 (2): 221-71.

Wu, M., Sugimura, Y., Takaya, N., Takamatsu, D., Kobayashi, M., Taylor, D. & Yoshiyama, M. (2013). Characterization of bifidobacteria in the digestive tract of the Japanese honeybee, Apis cerana japonica. *J Invertebr Pathol*, 112 (1): 88-93.

Yoshiyama, M. & Kimura, K. (2009). Bacteria in the gut of Japanese honeybee, Apis cerana japonica, and their antagonistic effect against Paenibacillus larvae, the causal agent of American foulbrood. *J Invertebr Pathol*, 102 (2): 91-6.

Young, S. L., Simon, M. A., Baird, M. A., Tannock, G. W., Bibiloni, R., Spencely, K., Lane, J. M., Fitzharris, P., Crane, J., Town, I., et al. (2004). Bifidobacterial species

differentially affect expression of cell surface markers and cytokines of dendritic cells harvested. *Clinical and Diagnostic Laboratory Immunology*, 11 (4): 686-690.

Yu, Y., Lee, C., Kim, J. & Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnology and Bioengineering*, 89 (6): 670-679.

Zhang, T. & Fang, H. H. (2006). Applications of real-time polymerase chain reaction for quantification of microorganisms in environmental samples. *Appl Microbiol Biotechnol*, 70 (3): 281-9.

Zimonja, M., Rudi, K., Trosvik, P. & Næs, T. (2008). Multivariate curve resolution of mixed bacterial DNA sequence spectra: identification and quantification of bacteria in undefined mixture samples. *Journal of Chemometrics*, 22 (5): 309-322.

Appendix 1: Instruments and reagents

Name	Manufacture
S.T.A.R buffer	Roche , Germany
Micro tubes	Sarstedt, Germany
Glass beads – <106 µm acid washed	Sigma-Aldrich, Germany
MagNA Lyzer	Roche, Germany
Magnetic sillica particles	Chemicell GmbH, Germany
epMotion 5070 pipetting robot	Eppendorf, Germany
KingFisher® Flex DNA extraction robot	ThermoScientific, USA
Quant-iT PicoGreen® dsDNA assay	Life Technologies™, USA
DNeasy Blood and Tissue Kit	Qiagen, USA
Black 96 well nunc® plate	ThermoFisher, USA
MilliQ water	Millipore Corporation, USA
96 well PCR plate	VWR, USA
2720 ThermalCycler	Applied Biosystems, USA
HOT FIREpol® DNA polymerase	Solis BioDyne, Estonia
Nuclease Free water	Ambion, USA
Ethidium bromide	Electran®, VWR International, England
100 bp ladder	Solis BioDyne, Estonia
Molecular Imager® Gel Doc™ XR Imaging System	Bio-Rad laboratories, USA
LightCycler 480	Roche, Germany
5X HOT FIREPol® EvaGreen qPCR Mix Plus	Solis BioDyne, Estonia
5X HOT FIREPol® Probe qPCR Mix Plus	Solis BioDyne, Estonia
BigDye® Terminator v1.1 Cycle Sequencing Kit	Applied Biosystems, USA
Agencourt® CleanSEQ® Dye-terminator Removal	Beckman Coulter, USA
96 well Super Magnet plate	Alpaqua®, USA
EDTA	Sigma –Aldrich, Sweden
AB-1100 Thermo Fast 96 PCR plate	ThermoScientific, USA
3130xl genetic analyzer	ABI, USA
MiSec® sequencer	Illumina, USA
TSB	Merck KGaA, Germany
500 ml Horse blood	ThermoFisher Scientific, USA
GasPack™ EZ CO_2 container system	Becton Dickinson, Ireland
Hart Infusion Broth	DIFCO Laboratories, USA
API-systems	Bio merieux® sa, France
ID colorcatalase (ID-ASE)	Bio merieux® sa, France
Bactident® Oxidase	Merck KGaA, Germany
RNase	Sigma Chemicals, Germany
All NTPs and $MgCl_2$	Solis BioDyne, Estonia
All primers and probe	Invitrogen™ Life Technologies™, USA

Appendix 2: Solutions

DNA extraction buffers

Table A2-1: Preparation of buffers used for DNA isolation from bee midgut

	Chemical	*Final concentration*	*pH*
Bindingbuffer	Guanidinium thiocyanat	5M	
	1M Tris HCl	150mM	Adjust to 6,8
	0,5M EDTA	7,5mM	
	Sterile H_2O	(add to final volume)	
TE buffer*	Tris HCl	10mM	8,0
	0,5M EDTA	1mM	
	Sterile H_2O	(add to final volume)	
Washbuffer1	Guanidine hydrochloride	60% (w/v)	
	Triton-X100	1% (w/v)	
	TE buffer*	10%	8,0
	Sterile H_2O	(add to final volume)	
Washingbuffer 2	EtOH	70%	
	Sterile H_2O	(add to final volume)	
Washingbuffer 3	Sterile H_2O	N/A	
Elution buffer	TE buffer*	N/A	8,0

DNA quantification

Table A2-2: PicoGreen® working solution for one reaction.

Master mix	**1 reaction**
PicoGreen®	0.25 µl
20X TE buffer	4.74 µl
MilliQ	90.01 µl
Total volume	**95 µl**

Table A2-3: The dilutions of λ- DNA standard for calibration curve, followed the listing below and the final concentration for the λ- DNA standard in each dilution is showed.

λ- DNA + MilliQ	Final conc. **ng/µl**
10 **µl** + 0	100
2 **µl** (100 ng/µl**)** + 8 **µl**	20
1 **µl (**100 ng/µl**)** + 9 **µl**	10
4 **µl (**10 ng/µl) + 6 **µl**	4
1 **µl (**10 ng/µl) + 9 **µl**	1
1 **µl (**1 ng/µl) + 9 **µl**	0,1

Working solutions for PCR and qPCR

Table A2-4: Master mix for PCR using HOT FIREpol® DNA polymerase. Different primers can be used in this mastermix. Total volume for one reaction is shown, as are initial and final concentrations of the constituents.

Solution/Reagent	Concentration	Volume 1x	Final
HOT FIREpol® DNA polymerase	5 U/µl	0.25 µl	1,25U
HOT FIREpol ® buffer B2	10x	2,5 µl	1x
Magnesium-dichloride, $MgCl_2$	25 mM	2,5 µl	2,5mM
dNTP	10mM	0,5 µl	0,2mM
Forward	10uM	0,5 µl	0,2uM
Reverse	10uM	0,5 µl	0,2uM
Nuclease free water	N/A	17,25 µl	N/A
Total volume		**24 µl**	

Table A2-5: Mastermix for qPCR of the vitellogenin and bee specific bacteria, using the 5X HOT FIREPol® EvaGreen qPCR Mix Plus. This mastermix can be used with different primers. Total volume for one reaction is shown, as are initial and final concentrations of the constituents.

Solution/Reagent	Concentration	Volume, 1x	Final Concentration
5X HOT FIREPol® EvaGreen qPCR Mix Plus	N/A	4 µl	N/A
Forward primer	10 µM	0.4 µl	0.2 µM
Reverse primer	10 µM	0.4 µl	0.2 µM
PCR water	N/A	14.2 µl	N/A
Total volume		**19 µl**	

Table A2-6: Mastermix for qPCR of Mangala universal bacteria 16S rRNA, by the use of 5X HOT FIREPol® Probe qPCR Mix Plus. This mastermix can be used with different primers. Total volume for one reaction is shown, as are initial and final concentrations of the constituents.

Solution/Reagent	Concentration	Volume, 1x	Final Concentration
5X HOT FIREPol® Probe qPCR Mix Plus	N/A	4 µl	N/A
Forward primer	10 µM	0.5 µl	0.2 µM
Reverse primer	10 µM	0.5 µl	0.2 µM
Probe	5 µM	1 µl	250 nM
PCR water	N/A	13 µl	N/A
Total volume		**19 µl**	

PCR purification with Exo1

Table A2-7: Exo1 working solution for one reaction. Exo1 supplied in the BigDye® Terminator v 1.1 kit.

Solution/Reagents	1x
5X Sequencing buffer	1.0 µl
ExoI	0.1 µl
DNase/RNase free water	2.9 µl
1:2 PCR product	1.0 µl
Total volume	**5,0 µl**

Sanger sequencing ddNTP labeling

Table A2-8: Working solution for the Sanger sequencing ddNTP labeling reaction using the BigDye® Terminator v 1.1, for one reaction.

Solution/Reagents	1x
BigDye® Terminator v 1.1	1 ul
BigDye® sequencing buffer 5x	2 ul
Sequencing primer (3,2uM)	1 ul
DNase/RNase free water	5 ul
Exo1 treated PCR product	1 ul
Total volume	**10 ul**

ddNTP removal

The following solutions were prepared for the CleanSEQ® Dye-terminator Removal protocol:

85% EtOH = 85 ml of distilled etanol
15 ml of Sterile MilliQ
(Final volume of 100 ml)

0,1M EDTA = Molecular weight of EDTA – 372,2 g/mol
3,7 gram in 100 ml Sterile MilliQ

0,1 mM EDTA= 100 µl 0,1M EDTA
99,9 ml Sterile MilliQ
(Dilluted 1:1000)

Bacteria cultivation

The following agar plates were made

TSA plates = 30 gram TSB } TSB
1 liter MilliQ }
15 gram agar

TSA with 5% horse blood = Add 50 ml sterile horse blood to the liquid agar (50 °C)
(final volume of 1,05 l)

HIB broth with 11% glycerol = 2,5 gram Heart Infusion Broth
11 ml 85% glycerol
100 ml MilliQ

PBS = 8,5 gram NaCl
0,85 gram $Na_2HPO_4*2H_2O$
0,25 gram KH_2PO_4
1 liter MilliQ

Appendix 3: Primer sequences and PCR protocols

Universal bacteria primers

Universal bacteria 16S rRNA: Amplicon 450 bp

Forward- F11: TCCTACGGGAGGCAGCAGT

Reverse- A01: GGACTACCAGGGTATCTAATCCTGTT

Cover All™: Amplicon 1200 bp

Forward: TCCTACGGGAGGCAGCAG

Reverse: CGGTTACCTTGTTACGACTT

PRK universal 16S: Amplicon 466 bp

Forward: CCTACGGGRBGCASCAG

Reverse: GGACTACYVGGGTATCTAAT

Universal bee primer

Vitellogenin: Amplicon 150 bp

Forward: GTTGGAGAGCAACATGCAGA

Reverse: TCGATCCATTCCTTGATGGT

Bacteria specific primers

***Gilliamella apicola*: Amplicon 210bp**

Forward: GTATCTAATAGGTGCATCAATT

Reverse: TCCTCTACAATACTCTAGTT

***Snodgrassella alvi*: Amplicon 128 bp**

Forward: CTTAGAGATAGGAGAGTG

Reverse: TAATGATGGCAACTAATGACAA

Tacman probe: (6-FAM)-5«-CGTATTACCGCGGCTGCTGGCAC-3«-(TAMRA) (*T*m, 69±9 °C)

PCR protocols

Table A3-1: PCR protocols. All protocols were run with 30 cycles, and cycled steps are marked green.

Target	**Activation**	**Denaturation**	**Anealing**	**Elongation**	**Final elongation**
Universal bacteria 16S rRNA	95 °C 15 min	95 °C 30 sec	60 °C 30 sec	72 °C 60 sec	72 °C 7 min
Cover All™	95 °C 15 min	95 °C 30sec	55 °C 30 sec	72 °C 1 min and 20 sec	-
PRK	95°C 15 min	95 °C 30 sec	57 °C 30 sec	72 °C 45 sec	-
Vitellogenin	95 °C 15 min	95°C 30 sec	54 °C 45 sec	72 °C 30 sec	72 °C 10 min
G.apicola	95 °C 15 min	95 °C 30 sec	55 °C 30 sec	72 °C 30 sec	72 °C 10 min
S.alvi	95 °C 15 min	95 °C 30 sec	55 °C 30 sec	72 °C 30 sec	72 °C 10 min

Table A3-2: QPCR protocols. All protocols were run with 40 cycles. Cycled steps are marked green.

Target	**Activation**	**Denaturation**	**Anealing**	**Elongation**
Universal bacteria 16S rRNA	95 °C 15 min	95 °C 30 sec	60 °C 1 min	Included in the annealing
Vitellogenin	95 °C 15 min	95°C 30 sec	54 °C 45 sec	72 °C 30 sec
G.apicola	95 °C 15 min	95 °C 30 sec	55 °C 30 sec	72 °C 30 sec
S.alvi	95 °C 15 min	95 °C 30 sec	55 °C 30 sec	72 °C 30 sec

Sequencing protocols

Table A3-3: Exo1 protocol. Only one cycle was used

	Exo1 activity	Inactivation
Exo1	37 °C 60 min	85 °C 15 min

Table A3-4: The Sanger sequencing ddNTP labeling reaction protocol. 25 cycles were run, and cycled steps are marked green.

Target	**Activation**	**Denaturation**	**Anealing**	**Elongation**
Universal bacteria 16S rRNA	95 °C 1 min	96 °C 15 sec	60 °C 4 min	Included in the annealing

Appendix 4: Methodology work flow

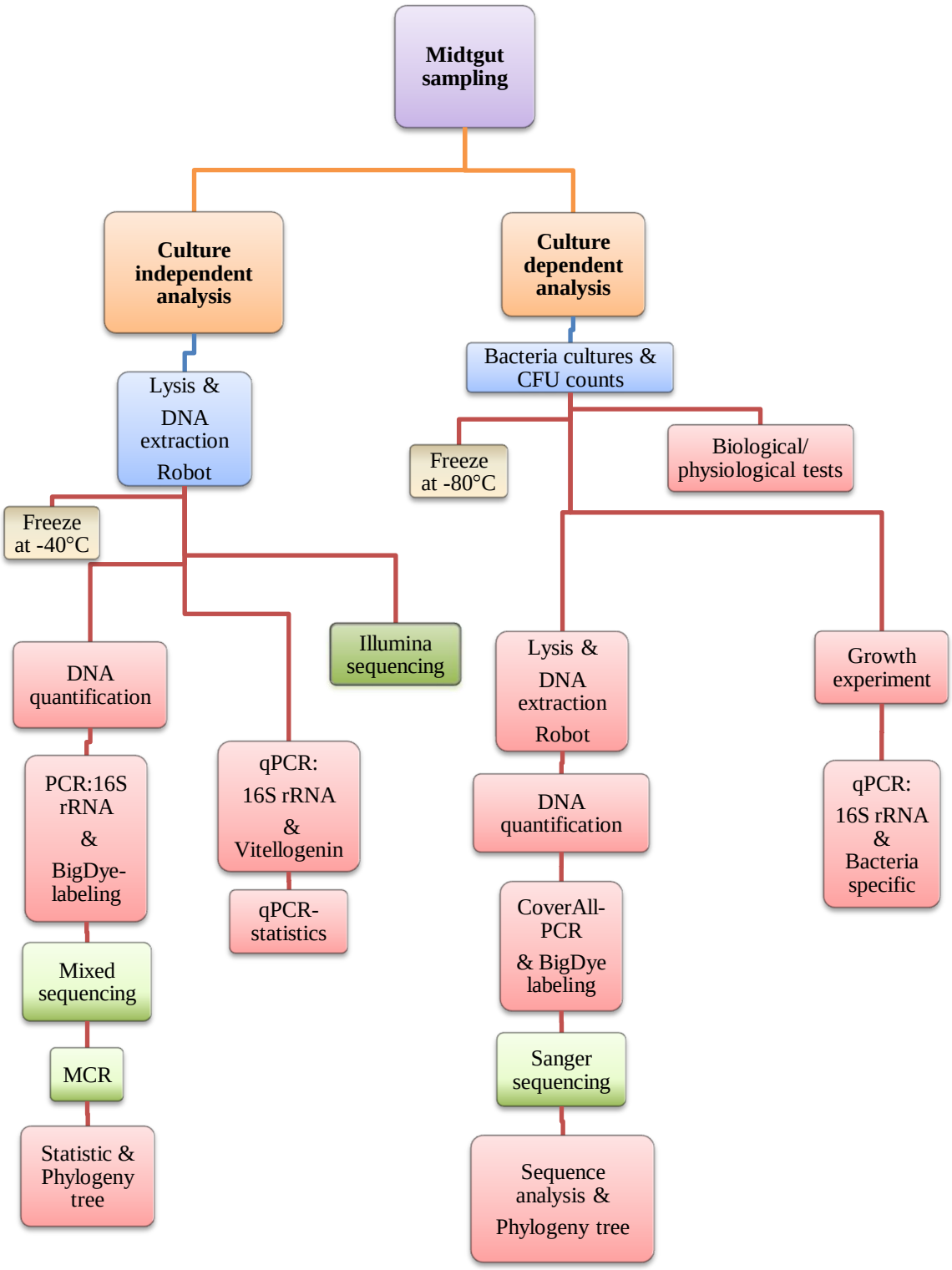

Figure A4-1: Flow chart of the different methods used in this thesis. Fading green methods = final analysis done by others. Green method = the hole method done by others.

Appendix 5: Validations and technical results

DNA quantifications

The midgut DNA was extracted and the gDNA yield is listed in table A5-1. The colony A in the UMB set yielded the highest amount of gDNA. For the bacteria isolates, the DNA was extracted by sweeping the agar plate and thereby varying amount of DNA from each bacterium were extracted, yielding a greater concentration rage.

Table A5-1: Concentration range for gDNA extractions calculated by PicoGreen® measurments. Sample sets used in the MCR-ALS analyzes are in black writing, and bacteria isolates in red.

Set	Concentration range gDNA ng/µl
Seasonal stability	0,3 – 2,0
Colony A	2,0 – 7,0
Colony B & C	0,9 – 2,0
Arizona	1,0 – 3,3
Bacteria isolates	1,6 – 9,2

The quantification of PCR product using PicoGreen® measurements generated a concentration range for each set as listed in table A5-2.

Table A5-2: Concentration range for PCR products calculated by PicoGreen® measurments. Sample sets used in the MCR-ALS analyzes are in black writing, and bacteria isolates in red.

Set	Concentration range PCR ng/µl
Seasonal stability	6,8 – 17,3
Colony A, B & C	5,4 – 11,6
Arizona	4,3 – 12,1
Bacteria isolates	6,1 – 10,8

Mixed sequencing and MCR-ALS analysis

Mixed sequencing produced electropherograms of different quality, and figure A1a show an example of one elctropherogram with adequate signals, which satisfied the MCR-ALS computer program criteria. Figure A5-1b and c show electropherograms which did not fulfill the same criteria.

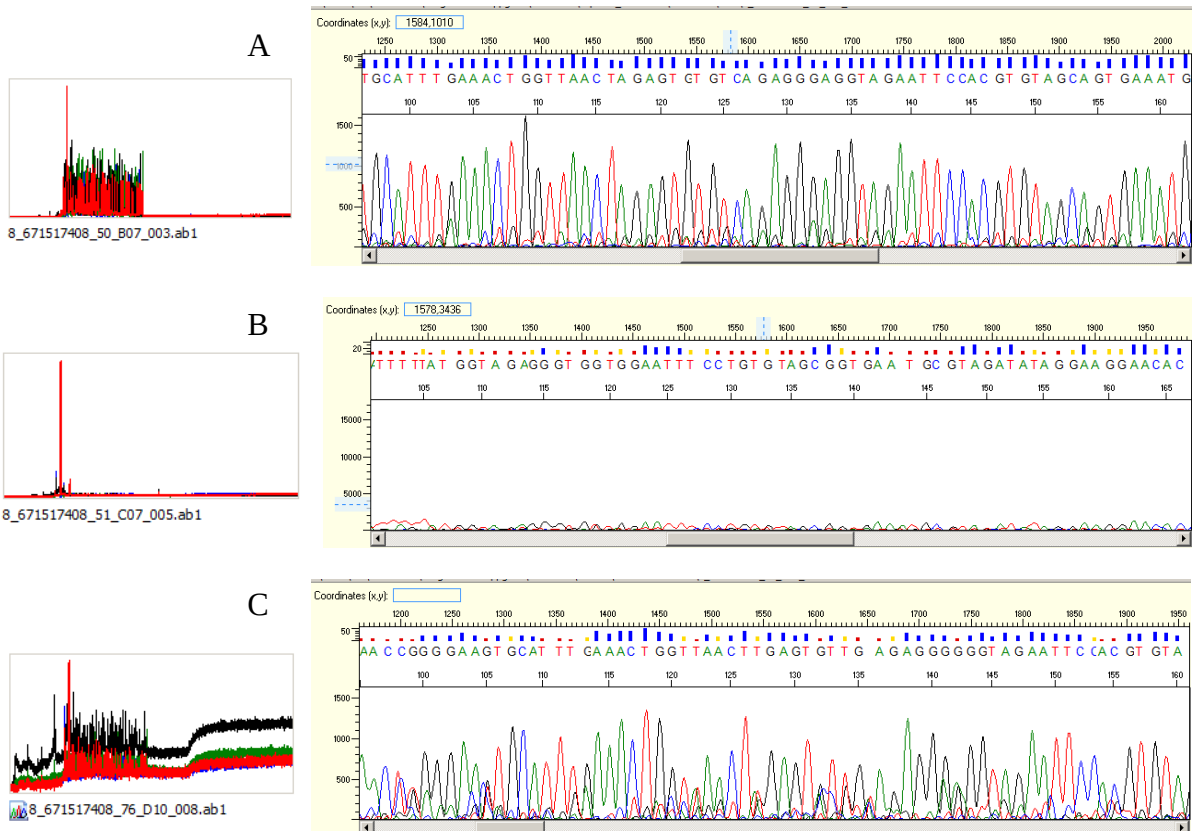

Figure A5-1: Electropherograms from mixed sequencing processed in Sequence Scanner v. 1.0 (Applied Biosystems). Sequences were investigated in thumbnail mode (left side pictures) for potential artifacts or errors. Right side pictures show a part of the sequence from 100 – 160 bp. A) Electropherogram from a midgut sample yielding adequate signals after mixed sequencing to be included in the datasets after MCR-ALS analysis. B) Electropherogram from a midgut sample yielding inadequate signals after mixed sequencing, thus not included in the dataset after MCR-ALS analysis. C) Electropherogram from a midgut sample yielding lower quality signals after mixed sequencing, which include more noise due to technical reasons, thus not included in the dataset after MCR-ALS analysis.

Table A5-3: Number of samples included in each set after MCR-ALS analysis, and calculated % coverage in reference to included midgut samples in table 2.1, 2.2, 2.3. Color marking of the different sets; Green = Seasonal stability, Red = UMB, Blue = Arizona.

Set	Samples included	% coverage
Mai	30	100
June	30	100
July	29	97
August	25	83
September	25	83
October	10	33
Colony A	22	73
Colony B	21	70
Colony C	25	83
WT	18	60
LP	29	97
HP	13	43

The components and their taxonomy

MCR-ALS computer analysis conducted in three separate turns, gave different dominating components for all three sets. The components with their assigned names (either family, genus or species), and S_ab score and unique common oligomers (from RDP database) are listed in table A5-4, as is the selected reference sequence`s accession number, for the components that could be determined on species level.

The midgut components can be classified in two phyla; Firmicutes and Proteobacteria, and four classes; *Bacili, Alpha-*, *Beta-*, and *Gamma -proteobacteria*. The *Lactobacillus* belongs in the Firmicutes phyla, *Lactobacilliales* order, *Bacilli* class. The *Azetobacter* and *Rhizobiales bacterium* are members of the *Alphaprotepbacteria* class, the *Snodgrassella alvi* belongs in the *Betaproteobacteria* class and the *Gilliamella apicola*, *Pasteurellaceae* and *Enterobacteriaceae* are members of the *Gammaproteobacteria* class.

Table A5-4: Components retrieved from MCR-ALS analysis for Seasonal stability and Diet datasets. Color markings; Seasonal stability – green, UMB – red, Arizona – blue. The taxonomy classification was made with S_ab score and unique oligomers criterias using RDP database and accession number are shown for isolates assigned at strain level.

Components	Assigned name	S_ab score/ unique common oligomers	Accession number GenBank
Component 1	Pasteurella bacterium	0.743 / 1320	EF187250
Component 2	Gilliamella apicola	0.971 / 1414	AY370191
Component 3	Enterobacteriaceae		
Component 4	Snodgrassella alvi	0.782 / 1422	AY370189
Component 1	Azetobacter		
Component 2	Rhizobiales bacterium	0.935 / 1388	HM108393
Component 3	Snodgrassella alvi	0,524 / 1422	AY370189
Component 4	Gilliamella apicola	1,000 / 1414	AY370191
Component 5	Lactobacillus		
Component 1	Gilliamella		
Component 2	Gilliamella apicola	0,971 / 1414	AY370191
Component 3	Lactobacillus		
Component 4	Enterobacteriaceae		
Component 5	Gilliamella apicola	0,707 / 1442	JQ93667

Regression analysis

The technical variation analysis on the qPCR relative ratio analysis, were first conducted by correlation analyzes of the sample parallels, which gave poor correlation in some of the samples, but by excluding a few outliers in each plate run, the correlation was given by r^2= 0,610; 0,573; 0,785, but despite this, the data was showing a correct trend. Fig A5-2a show an example using the May and June parallels.

For the bacteria composition r^2 calculated for the sample parallels was $r^2 \approx 0{,}98$ for all components at time points, and an example using May samples is shown in fig A2b. Internal plate variation in the sequencing methodology seems to be minor.

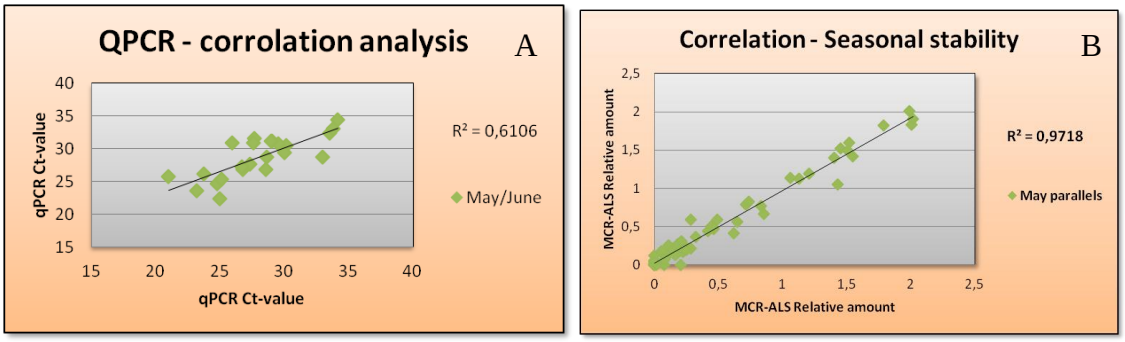

Figure A5-2: R^2 analysis of plate parallels. A) Correlation analysis of Ct-values from qPCR analysis on plate parallels of May and June samples. B) Correlation analysis of sample parallels of the four components, retrieved from MCR-ALS analysis of the Seasonal stability set, calculated from May samples.

HRM-analysis and verification of qPCR products

HRM analysis was run on all qPCRs that used the EVAGreen Hot Fire® Pol mix, and the different HRMs are shown in figure A5-3 a,b, and c. The Vitellogenin aplicon was very specific and showed little primer dimer amplification. For the *Gilliamella apicola* some unspecific signal could be detected, and for the *Snodgarssela alvi* a steady low background signal was seen, but no primer-dimers were detected for either of the specific qPCRs.

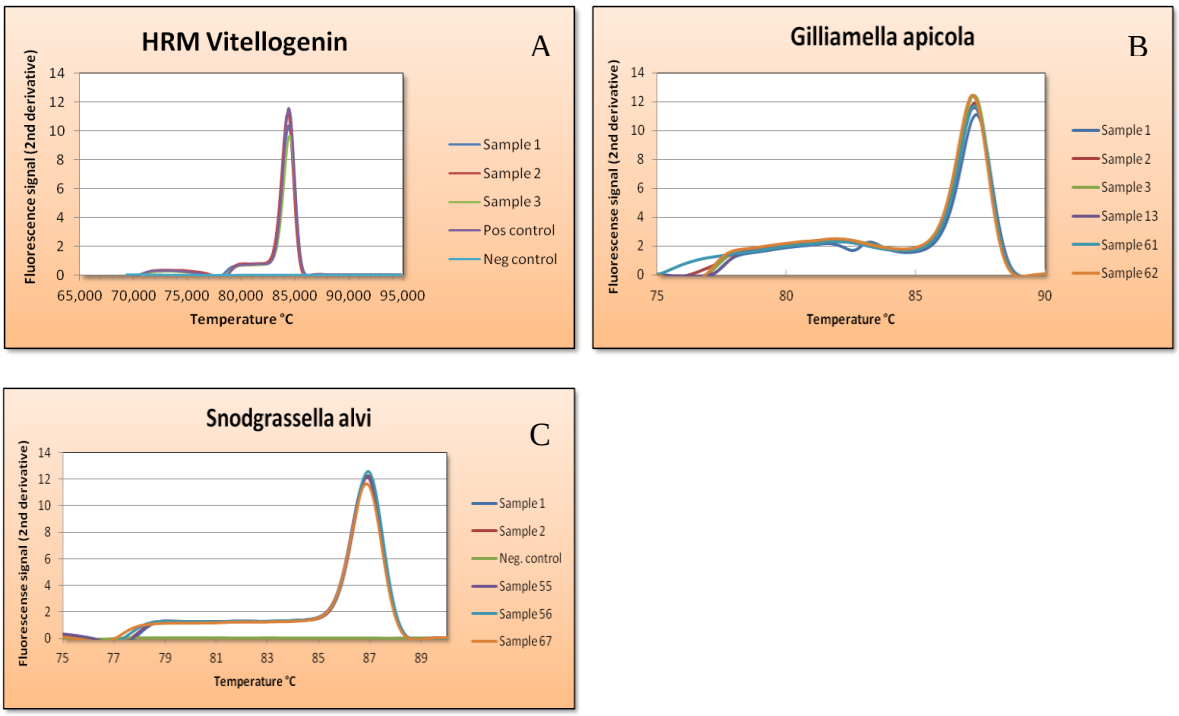

Fig A5-3: HRM analysis where fluorescense signal in the 2nd derivative is plotted against temperature to show melting peak of the specific amplicon. A) HRM analysis of the vitellogenin gene. B) HRM analysis of the *Gilliamella apicola* specific qPCR. C) HRM of the *Snodgrassella alvi* specific qPCR.

Amplicon verification of *G.apicola* and *S.alvi* qPCR

The specific qPCR amplicons were verified on 1,5 %agarose gels, 70V for 50 min (fig A4). Desired amplicon size; *G. apicola* = 210 bp, *S. alvi* = 128 bp, both bacteria showed strong and correct amplicon size.

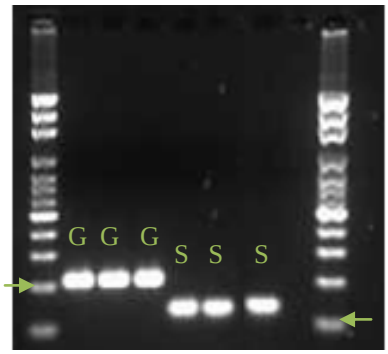

Figure A4: *Gilliamella* and *Snodgrassella* qPCR from the competition experiment verified on agarose gel, with 100 bp ladder. Three samples was selected from each specific qPCR; G = *Gilliamella*, S = *Snodgrassela*. The arrows mark 200bp (left side) and 100 bp (right side).

Bacteria isolates

Table A5-5: Taxonomy assignment of all bacteria isolates showing their accession number according to 95% BLAST matches and E-value < $1.0*10^{-5}$ for all matches. The number column describes the order in which the bacteria were isolated and the color marking show similar bacteria, and the bacteria written in black on white background were only isolated once.

Number	Assigned bacteria name	Accession Number
1	Bifidobacterium asteroides	AB437355
2	Gilliamella apicola strain wkB1	pAJ204; AY370191
3	Snodgrassella alvi strain wkB2/ wkb12	JQ746651
4	Lactobacillus kunkeei	JQ009336/AB559821
5	Gluconobacter oxydans	X73820
6	Candidatus Gilliamella apicola clone pAJ206	AY370192.1
7	Snodgrassella alvi strain wkB12/2	JQ746651
8	Bifidobacterium asteroides	AB437355
9	Snodgrassella alvi strain wkB2	JQ746651
10	Gluconacetobacter sp. clone pAJ205/Commensalibacter intestini	AY370188.1/A911; EU409601
11	Gluconacetobacter sp. clone pAJ205 /Commensalibacter intestini	AY370188.1/A911; EU409601
12	Bifidobacterium indicum	JCM 1302; D86188
13	Bifidobacterium asteroides	AB437355
14	Candidatus Gilliamella apicola clone pAJ204	AY370191
15	Bifidobacterium asteroides	AB437355
16	Enterobacteriaceae bacterium Acj 122	AB480765
17	Snodgrassella alvi strain wkB2	JQ746651
18	Gluconacetobacter sp. clone pAJ205 /Commensalibacter intestini	AY370188/A911; EU409601
19	Candidatus Gilliamella apicola	pAJ204; AY370191
20	Rhizobiales bacterium **Uncultered -martinson**	SHAI005; HM108364
21	Snodgrassella alvi strain wkB2	JQ746650
22	Candidatus Gilliamella apicola clone pAJ206	JQ673255
23	Snodgrassella alvi strain wkB2 /12	JQ746651.1
24	Candidatus Gilliamella apicola clone pAJ206	AY370192.1
25	Bifidobacterium indicum/ coryneforme	JQ673255
26	Pasteurellaceae bacterium LvLi2 /Enterobacteriaceae bacterium	EF187250.1/Acj 122; AB480765
27	Gluconacetobacter sp. clone pAJ205 /Commensalibacter intestini	AY370188/A911; EU409601
28	Candidatus Gilliamella apicola	AY370191/92
29	Snodgrassella alvi strain wkB2	JQ746650